Enabled to Work: The Impact of Government Housing on Slum Dwellers in South Africa

Simon Franklin*

May 15, 2020

Abstract

Do informal housing conditions constrain labour supply? I estimate the effect of receiving a free house under the South African government's housing program, which has given away over 3 million housing units since 1994. Using four waves of household panel data from Cape Town and geographical data on the location of large housing projects, I exploit a natural experiment whereby households living close to projects were first in line to get them to instrument for selection into the programme. I use projects that were planned and approved, but never actually built, to deal with non-random placement of housing projects. Government housing has a significant positive effect on household earnings. This is driven primarily by increases in earnings for women. I present evidence consistent with a mechanism whereby formal housing frees up time by alleviating the demands of work in the home.

*Queen Mary University of London (s.franklin@qmul.ac.uk). I acknowledge funding from the World Bank and OxCarre. This paper is a part of a Global Research Program on Spatial Development of Cities, funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development. For their assistance in giving me access to specific household GIS data and an early release of the 5th wave of CAPS data, I'd like to thank Jeremy Seekings, Brendan Maughan-Brown and David Lam. Thank you to Magdalen College for providing a great deal of the funding that made possible my research in Cape Town. Thank you Marcel Fafchamps for input, oversight and guidance of the project. Thank you to helpful seminar participants at the CSAE Conference, Oxford OXIGED housing workshop, as well various internal seminars in Oxford. In particular I thank Tony Venables, Stefan Dercon, Paul Collier, Somik Lall, Taryn Dinkelman, James Fenske, Simon Quinn, Francis Teal, Julien Labonne, & Kate Vyborny. I also thank two anonymous reviewers for their detailed and useful comments.

1 Introduction

Substandard housing conditions are one of the main deprivations suffered by the poor around the world. It is estimated that over 860 million people live in slums in developing countries, and this number has been growing rapidly with increasing rates of urbanisation (UN Habitat, 2003, 2010). Poor housing is associated with a lack of access to running water, electricity, heating, ventilation, and security of tenure. Policy interventions to improve housing conditions through slum upgrading, land titling and public housing are widespread (Buckley et al., 2016; Michaels et al., 2017). Such policies may be partly motivated by the negative externalities caused by slums, but also by the idea that living in slums can have detrimental impacts of the economic lives on residents, and therefore constitute "poverty traps" for those living in them (Marx et al., 2013).

A large literature shows that improvements in home technologies can increase labour supply, especially for women, by freeing up time previously spent in the home (Duflo, 2012; Dinkelman, 2011; Greenwood et al., 2005). In a related way, poor housing could place burdens on the time of residents, limiting their ability to actively participate in labour markets.[1] It is plausible to think, therefore, that providing serviced formal housing could improve labour market outcomes, especially for women. However, relatively little research directly links housing conditions and employment outcomes.

The South African government's large-scale housing program has delivered over 3 million stand-alone housing units free of charge to people living in informal housing conditions over the last 26 years. In this paper, I use longitudinal household data from the city of Cape Town over four waves from 2002 to 2009 to assess the impact of government housing on household labour supply.[2] In order to isolate the effect of moving from informal to formal housing I focus on households that live in informal settlements at baseline.

Evidence on the effects of improved housing quality is hard to identify, in part because government housing programs usually combine improved housing conditions with relocation to different (often worse) neighbourhoods. The South African case is no exception. As the program was scaled up and cities grew, government housing has been built increasingly far from the city centre. However, due to the placement of housing projects in this city during the early years of the program when my study is conducted, households in my study moved, on average, only slightly further from their original place of living when they won housing. Furthermore, my identification strategy leverages the proximity of potential beneficiaries to new housing projects. This allows me to isolate the impact of improved formal housing from possibly confounding impacts on of neighbourhood changes or distance from employment, which have been shown to have negative impacts on access to jobs and existing social networks

[1] Field (2007) argues that improved tenure security can allow households to access labour market opportunities without fear of their homes being expropriated.

[2] It is estimated that between one quarter and one fifth Cape Town's entire population benefited from government housing since 1994 (Seekings et al., 2010).

for relocated households (Jacob, 2004; Barnhardt et al., 2016; Picarelli, 2019). This is the main contribution of my paper. My results should be seen less as representative of the average effects of government housing programs, including the South African program, but rather as the isolated effect of improved housing alone. Indeed, the South Afircan program has generally moved beneficiary households great distances from the city centre, which has been shown to have deleterious effects (Picarelli, 2019).

Households that receive a government house experience large increases in income, relative to households that do not receive one. My fixed effects OLS estimates show an increase in household earnings of 18% due to improved housing. This increase is driven primarily by increased earnings from employment for female members of treated households.[3] Turning to data on earnings for individual earnings, I find modest (and statistically insignificant impact) on total employment among women, but significant increases in salaries for women who were already working before they won a house. These effects on earnings are further concentrated among women who were working part-time at baseline. The results suggest that improved housing lead women to increase their labour supply at the intensive margin. Consistent with this, I find a significant effect on hours worked for a sub-sample of young adults (nearly 50% of my sample) for whom I have data on working hours.[4] I do not find the same effects when I replicate my analysis on a sample of households who did not live in informal housing before they received government housing, suggesting that it is housing itself rather than some other effect of winning housing (such as wealth effects) that drive the results.

To improve causal identification, I use the allocation procedure used by local government as a natural experiment: program recipients were selected on the basis of their proximity to government housing developments. I proceed in two steps: first, I use the distance between households' original place of living and the location of newly built housing projects to instrument for selection into treatment at the household level. Although these IV estimates are less precise, they are larger than the main OLS estimates, suggesting negative selection at the household level, leading to downward bias in the main OLS estimates. Second, I deal with possibly non-random placement of housing developments across different areas by using a set of housing projects that were approved and planned at the same time as completed projects, but cancelled or delayed for reasons unrelated to the communities they were intended to benefit. I then compare only households that were near to completed projects to those near incomplete projects and find quantatively similar effects on income and female earnings.

I discuss a number of mechanisms that could be driving the increase in earnings. I provide suggestive evidence for one key causal mechanism: that poor housing conditions place particular burdens on the time use of women.[5] I use evidence from the baseline data, as

[3] I cannot rule out that housing also had very small effects on male household members.

[4] My data does not include information on hours worked for all household members.

[5] My data does not record time use for women beyond the first wave (baseline) of the panel, which prevents me from estimating the effects of housing on labour in the home.

well as representative time-use surveys from South Africa, to show that women allocate a remarkably large amount of time to housework, preparing food, and care for children and the elderly (Budlender et al., 2001). Consistent with evidence from other developing countries (Berniell and Sánchez-Páramo, 2012), time allocations to work in the home are significantly larger for women living in informal housing. I estimate significant positive treatment effects of government housing on electrification, direct access to running water, and modern home appliances, all of which could be saving significant amounts of time for women. Finally, I conduct exploratory sub-group analysis, which shows that the effects are concentrated among households with a larger number of children, and among individual women who are older than 30 and female heads of households. I argue that the combination of three effects (domestic work in the home, home maintenance, and tenure security) could be jointly responsible for a large proportion of the increases in earnings. Ultimately, however, my data does not allow me to rule out other possible mechanisms.[6]

These results contribute to the literature on the relationship between physical living conditions and female labour supply. Labour saving improvements to the lives of the poor can free up time to work in the labour market (Duflo, 2012; Greenwood et al., 2005; Devoto et al., 2011). Dinkelman (2011) shows that electrification of homes increases female labour supply and earnings by freeing up time from work at home in South Africa. Field (2007) shows that improved tenure security, without improving housing quality, frees up time that otherwise would have been spent at home defending the home from expropriation. Few papers have looked at the impacts of slum upgrading or housing programs on labour supply. Keare and Parris (1982) find that provision of tenure and basic services in four countries had positive impacts on employment and income generation.

Second, I contribute to the broader literature on slums as poverty traps. Marx et al. (2013) discuss three channels through which living in a slum could constrain household income and investment: human capital and health effects, policy neglect, and under-investment due to weak property rights. I add a fourth channel to this list by showing that poor housing conditions can constrain labour market participation. While a growing literature looks at the impact of housing on health and well-being (Cattaneo et al., 2009; Galiani et al., 2017), there are relatively few studies showing a link between poor housing and ability to participate in labour markets.

Third, this paper contributes to the evidence on the effects of large scale housing programs. Public housing is often associated with negative economic and social consequences, largely because the evidence on the impacts of housing estates in developed countries has largely been negative (Olsen and Zabel, 2014). Yet large-scale housing projects of this kind are in-

[6]For example, I cannot investigate whether or not the effects are partly driven by improvements in the health of recipients (for which there is some evidence (Zwane and Kremer, 2007; Pitt et al., 2006; Jalan and Ravallion, 2003; Duflo et al., 2012; Muyeba, 2014)), improved educational attainment (Kumar, 2019), or that households used new homes to conduct small business activities. Howeer, I find no evidence of changes in occupational choice as a result of getting housing.

creasingly popular with governments and electorates in the developing world (Gilbert, 2004; Buckley et al., 2016; Franklin, 2019). My study does not estimate the average treatment effect of government housing in South Africa. Due to my estimation strategy, I estimate the effects of improved government housing on households that were previously in slums, while keeping geographic isolation (distance from the city) constant. On the other hand, Picarelli (2019) evaluates the same government program at the national level using a different identification strategy. This study includes households that were already living in formal houses before they received government housing, and so may not experience the same housing-quality shock that households in my study experience. In the post-2009 period, when well-located land was increasingly scarce, she finds that government housing does indeed relocate households further from economic opportunity, which in turn negatively affects female labour supply. The results from Picarelli (2019) can be better thought of the total effect of the policy, for the contemporary period, while the results in this paper isolate the effect of government housing only for households that previously lived in informal settlements and in the absence of the effects of relocation. Recent evidence from India suggests positive human capital and employment effects of government housing, also from a program that did not move households too far from their original locations (Kumar, 2019). Taking these results together underscores the economic benefits of improving housing conditions for the urban poor in developing countries, but suggests that these benefits can be undermined by requiring households to move further from the city, as urban housing programs often do.

2 Setting and Framework

2.1 Informal housing and policy in South Africa

The South African government's housing policy was part of the Reconstruction and Development Program (RDP), introduced after the end of apartheid in 1994 to address racial inequalities in both wealth and service delivery. During the dismantling of apartheid-era restrictions on the movement of people before and after the first democratic elections in 1994, high rates of migration to cities put increasing pressure on the already insufficient housing stock, leading to a housing crisis. When the first democratic government was elected, there were an estimated 12.5 million people without adequate housing. Only 65% of the total population was housed in formal (cement and brick) dwellings nation-wide. In the city of Cape Town, where this study is conducted, it is estimated that the number of informal dwellings grew from 28 000 in 1993 to roughly 100 000 in 2005 under the pressure of migration and urban population growth (Rodriques et al., 2006).

The new government promised to deliver 1 million free houses from 1994 to 1999, partly in compliance with the newly enshrined constitutional right to adequate housing for all citizens. This ambitious target was more or less met, and in May 2013 the government announced that it had passed 3 million houses delivered.[7] Beneficiaries get a house in large-scale developments. These developments are coordinated by the government and built by private developers. The housing units are uniformly single-storey, stand-alone houses on distinct plots, usually with one or two bedrooms, one bathroom and a communal kitchen and living area, connected to electricity and running water in the home. Beneficiaries own the houses outright with no mortgage or other financial contributions.

Housing developments have usually been built in vacant areas among or adjacent to existing informal settlements and townships, or on new land on the edge of cities. Beneficiary households move directly from informal settlements into new sites. In other cases, existing informal settlements were demolished and re-built: existing residents were required to relocate to temporary camps before moving back, into the completed housing units.[8] Resale rates of housing are around 20%, mostly on an informal market.[9] Renting of these houses does not appear to be at all common, while the practice of building a small "backyard" shack or shelter is.[10] More than one-third of households had a backyard structure within a few years of receiving the house.[11]

To be eligible to receive housing, an applicant needs to be married or otherwise supporting dependents in a household with a total income of less than R3500 per month, cannot own a registered property, and must be a South African citizen (Department of Human Settlements, 2009). However these eligibility requirements were often unverified in practice, and weren't for most slum dwellers (Tissington et al., 2013).[12] Demand for houses has always been high, and waiting lists long. In Cape Town, local government groups and committees allocated units to housing applicants whenever new housing projects were developed in their local area. I discuss issues to do with the allocation of housing to applicants when I discuss the identification strategy used in this paper in Section 4.

The policy has a mixed legacy. Housing delivery has been a cornerstone of the African National Congress's electoral campaigns since 1994. More than 10 million people are estimated to have benefited directly from the program. Unsurprisingly, free housing is very popular, and demand for the housing still far outstrips delivery. In a period of increasing unemploy-

[7] According to the census of 2011 there are approximately 14.5 million households in total in South Africa.

[8] This summary of the program, of course, omits many important details and changes to the policy over time, which are described in earlier versions of this paper, or for a comprehensive summary, see Tissington (2011).

[9] Among households who move out of government housing after winning, just more than half move back into informal settlements, the remainder move to other formal housing. I find no significant correlation between changes in income and the decision to move out of government housing.

[10] These figures are from the Western Cape Occupancy Study (Vorster and Tolken, 2008).

[11] These structures are sometimes used to accommodate other members of the household who could not fit in the original structure. In the cases when the structures were occupied by non-household members, only about half of pay rent. Some households still owned their previous (informal) dwelling and were renting it out, but in most cases, their informal dwellings were demolished when they left, or they gave them over to a friend or family member.

[12] There are relatively few individuals (only 12%) in my sample of slum-dwellers who fall above the income cutoff of R3500: the eligibility constraint did not bind for them.

ment and urbanization, the number of households living in informal housing has increased, especially among the black population. The policy has been criticised for not doing enough to deal with the housing backlog, and for providing low-quality housing without accompanying infrastructure and public investments (Tomlinson, 1998; Huchzermeyer, 2003).[13] The most pervasive criticism of the policy has been the location of housing which has often been determined by private construction companies that choose to build on the cheapest possible land. In most cities housing has been built far away from the city centres in a way that has reinforced the spatial segregation of South African cities (Huchzermeyer, 2006; Bundy, 2014; Charlton and Kihato, 2006).

2.2 Home production and labour supply

In this section, I provide evidence that slum dwellers' time may be constrained by their physical living environment, particularly due to the time requirements of home production. I argue that the alleviation of these time constraints could explain the effect of improved housing on labour supply that I document in the paper. I conceive of home production as a variety of time-consuming activities related to the production of goods and services consumed at home. These can be divided into three categories: 1) domestic activities such cooking, cleaning and childcare, 2) the maintaining of the physical structure of homes and rebuilding of structures after damage from fires or flooding, and 3) time spent in the home ensuring it's safety across home invasion and theft, including protection of dependents residing in the home.

Home production is likely to be less efficient in informal housing. Cooking and bathing are likely to be easier in a home with running water and electricity, as opposed to an informal dwelling where other carbon fuel sources are often collected, and water has to be fetched from communal taps. Maintaining a sanitary home environment is also likely to be far easier in a cement-floored home without leaking roofs or permeable walls.[14] In Cape Town most slum dwellers have to use badly maintained communal toilets located some distance from their homes, or buckets which must be emptied regularly. Paraffin is a common use of fuel for cooking and heating and is known to be a cause of fires and respiratory disease (Schwebel et al., 2009). As is the case in slums around the world, formal electricity connections are rare for shack dwellers; more than 50% having fire-prone illegal connections, or no electricity at all (City of Cape Town, 2005). Lack of access to electricity may, in turn, limit the use of labour-saving appliances such as fridges, stoves and microwaves. These appliances may harder to secure and maintain in informal housing.

In 2000, only 28% of informal dwellings in urban areas used electricity for cooking, versus 77% of formal urban homes. Similarly, they were far more likely to use gas or paraffin

[13] Housing programs have generated considerable political conflict: they are said to have contributed to forced evictions (Chance, 2008), particularly to areas further away (Centre on Housing Rights and Evictions, 2009) and to have been biased towards particular racial or political groups (Seekings et al., 2010).

[14] Cattaneo et al. (2009) have looked at how cement floors improve health from improved sanitary conditions.

stoves for cooking and heating and lighting. Only 46% of shack dwellers have a refrigerator, compared to 90% of families in brick houses. Only 10% of shack dwellers own a microwave.

Time-use surveys of poor South Africans confirm that a considerable amount of time is consumed by domestic activities, particularly for female members of households, who are primarily responsible for chores at home. South African women spend on average 3.5 hours a day on unpaid work, three times as long as men (Budlender et al., 2001).[15] Crucially, the evidence suggests that these activities take far longer in informal housing. In the national data, individuals living in informal housing in urban areas spent 25% more time on non-labour market work than other urban households (Budlender et al., 2001). In my data, shack dwellers in Cape Town report 17.1 hours per week spent on housework, more than twice as much as their formally housed counterparts (7.5 hours per week).[16]

Risks from accidents and weather could also create time burdens. Households living in the slums in Cape Town are vulnerable to township fires and, during winter months, storms and flooding.[17] Fire hazards are due, in part, to the types of appliances used for cooking and heating outlined above. Township fires are common and lead to widespread destruction of housing infrastructure, which takes time and money to rebuild. Pharoah (2012) provides an overview of some of the risks facing informal dwellers in Cape Town. The greatest impacts come from the health problems and losses of days worked and at school because of the disruption caused by fires.[18] In that study, 83% of shack dwellers had experienced some kind of flooding while living in Cape Town.

Finally, slum dwellers have considerably less security, since their homes can easily be broken into. This could impose limits on tenants' ability to commute into the city to look for work for fear of theft.[19] This could also limit the ability of families who work in the home to leave school-age children unattended at home for periods of time after school, as it would be impossible to properly lock an informal home against intruders.

Standard economic models link home production, labour supply to wage-paying jobs, and leisure by specifying household utility as a function of a home production input which is produced by combining labour in the home and the home technology (Becker, 1965; Gronau,

[15]These patterns are consistent estimates for other developing countries (Berniell and Sánchez-Páramo, 2012).

[16]These were calculations based on the CAPS datasets used for this paper, outlined in Section 3.1. Unfortunately, this data was not collected for periods of the survey beyond the first wave, which makes it impossible to estimate the impact of housing on time-use in this setting.

[17]In 2005 a particularly damaging fire razed over 3000 shacks in Joe Slovo informal settlement just outside of Cape Town ("Shack-dwellers have nothing left after blaze" (iolnews, January 17, 2005)) The victims of the fire were promised government housing after being displaced, but many remain in temporary relocation camps years later (Centre on Housing Rights and Evictions, 2009).

[18]Some 20% of households live in high flood risk areas, and roughly 40 000 people were directly affected by townships fires in Cape Town between 1995 and 2004.

[19]The threat of invasion seems more urgent than that of expropriation risk documented inField (2007). While tenure security is an issue for informal dwellers in South Africa (Royston, 2002), this risk is related to eviction by the state to make way for new housing or urban development projects more than contestation of property right by other private agents.

1977). In the Online Appendix, I formally outline a very simple model that expresses this intuition. In my setting, households in informal housing with these poor home technologies may choose to dedicate an enormous amount of time to home production. It is plausible that improving the quality of the home environment through the upgrading of homes would free up this time to shift towards work in the labour market.

3 Data

3.1 Housing Project Data

I generate a georeferenced panel of the number of households built per project for each year. First, I construct a dataset of government housing projects built in the period 1999 to 2009, using administrative data gathered from the Provincial Department of Human Settlements and Local Government Planning departments in Cape Town. Second, for each housing project constructed during this time I matched project maps with meta-data on the construction dates, administrative approvals and number of houses built and delivered in each year for each project. This data is presented in Figure 1 showing the expansion of housing projects over the years for the part of Cape Town where most free housing was built.

3.2 Household Panel

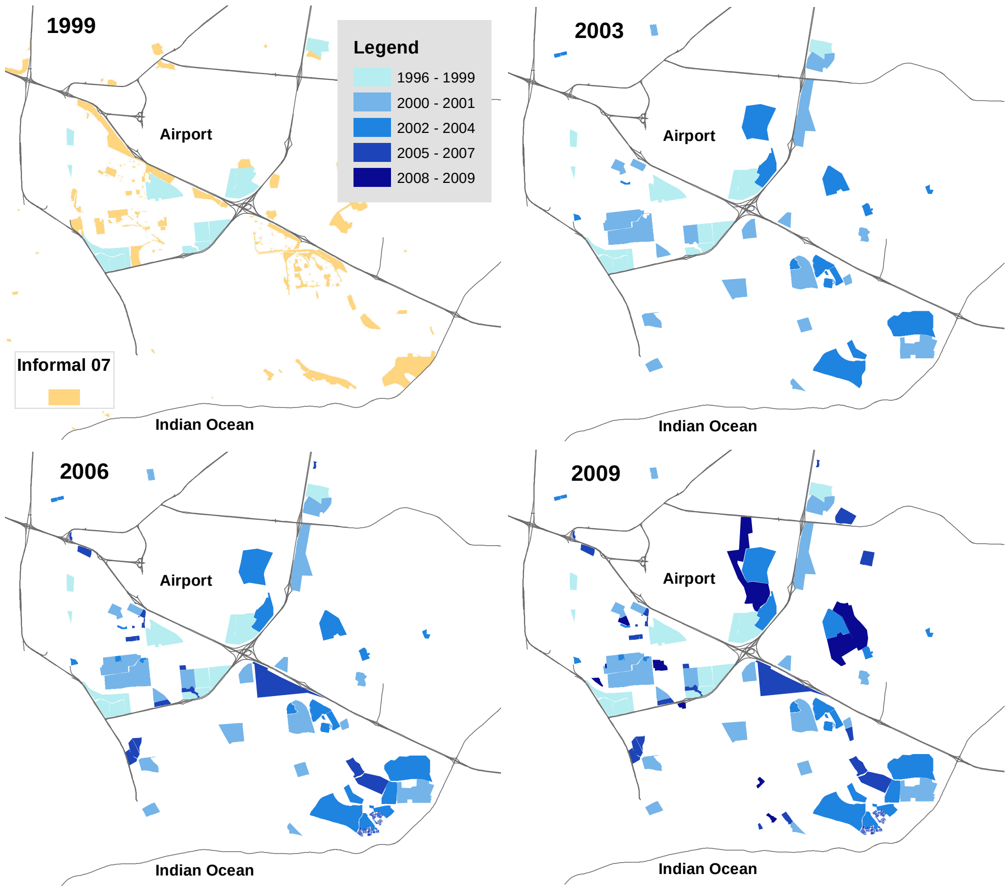
I use the CAPS panel survey of Cape Town metropolitan area, with four waves collected in 2002, 2005, 2006, and 2009.[20] The sample was randomly selected using probability proportional to size sampling and stratification by racial group using 1996 local area census data.[21] The primary sampling unit is the enumeration area, area the size of a large city block (roughly 0.1 square kilometres in my data) and I work with 159 enumeration areas in my main analysis.

I restrict my analysis to a sample of households that began the study living in informal housing. Roughly one-third of the sample are living in such "shacks" at baseline and thus were eligible for the housing program. In addition, I drop all households that had already won government housing and thus were ineligible for government housing even if they had returned to live in informal housing. This leaves me with a sample of 1350 eligible households, of which 1097 (or 81%) of these are found at least once in subsequent waves. In this way, I begin by comparing only similar households that are eligible for the program.

[20]The Cape Area Panel Study Waves 1-2-3 were collected between 2002 and 2005 by the University of Cape Town and the University of Michigan, with funding provided by the US National Institute for Child Health and Human Development and the Andrew W. Mellon Foundation. Wave 4 was collected in 2006 by the University of Cape Town, University of Michigan and Princeton University. Major funding for Wave 4 was provided by the National Institute on Aging through a grant to Princeton University, in addition to funding provided by NICHD through the University of Michigan (Lam et al., 2006). Further information can be found on the CAPS website at http://www.caps.uct.ac.za

[21]This survey was conducted with the primary motivation of tracking young adult's behaviour, sexual attitudes, labour force participation and health. However household data were also collected, with an extensive household roster questionnaire which surveyed the entire household.

Figure 1: Housing roll out in Cape Town's townships.



Housing projects boundaries are shown in blue. The official map of informal settlements in 2007 is shown in the top left panel in yellow.

Table 1 provides relevant descriptive statistics for the eligible sample. Over just seven years, nearly 40% of the sample has received a government house. More households are treated between the first and second periods (19.7%) than any other. I will use the cumulative treatment status of households (whether they have received housing ever before) as the treatment variable. Not all households remain in government housing projects after they win the lottery. I find that 23 percent of households who received a government house are not living in one at the time of the final round of data collection in 2009.

The study takes place during a time of improving living conditions even outside of the government housing projects. Households living in informal housing saw improvements in the quality of their living and housing conditions rapidly over time, at a rate that cannot be attributed to the construction of free housing alone. This is due to a combination of public service delivery, private investments in housing improvements, and some households managing to afford to move to formal areas.

Individuals in my sample face tough labour market conditions. Roughly 50% of adults in my sample (ages 18 to 60) are working at all in my data, throughout the eight years of the panel. Fifteen percent of households have not one working adult at baseline. The strict unemployment rate (requiring respondents to be searching for work to counted as unemployed) is 24 percent, but this rate is widely thought to hide high rates of discouraged job seekers in the South African context. Occupational choices for women in my setting are limited. Nearly 50% of women in my sample work as cleaners, usually in private households. The majority of other jobs are in the low-skilled service industry. See Table A.2 for a breakdown of occupation choices and corresponding wages in my data. Wages are also low. Women working as a cleaner in my sample earn R1,238 per month on average in 2006, which is equivalent to roughly $160 per month in 2019 nominal terms.[22]

Table 1: Evolution of sample household characteristics

Wave	1	2	3	4
Year	2002	2005	2006	2009
Gov housing this year	0.0%	19.7%	8.87%	12.09%
Gov housing (cumulative)	0.0%	19.7%	25.9%	38.6%
Shack	100.0%	65.6%	62.4%	45.6%
Flush Toilet	70.8%	79.2%	85.6%	90.7%
Piped Water	12.3%	25.3%	28.4%	42.0%
HH Size	5.26	5.31	5.17	5.74
% Female	54.0%	55.1%	54.0%	53.3%

3.3 Location and Proximity to Housing

I match household data to housing and planning data, using original enumeration areas maps to locate the original living location of households in the first wave of the sample.[23] I calculate a set of pairwise distances for each household, and each public housing site, for each time period in the data. Here, I use the data on whether a housing site was constructed in the years between the survey waves to determine whether it should be used in the calculation of project-distances for that wave.

Table 2 shows these measures, the average distance from the closest housing project, then the second and third closest. I then rank city blocks (or enumeration areas) by their distance to each housing project, to calculate the extent to which a community was, geographically speaking, "first in line" for a given project. Each household-project distance pair has a corresponding rank assigned to it, from which I generate a dummy variable indicating that a household was in one of the three closest communities to a public housing site. The data show that "treated" households lived, on average, closer to housing projects at baseline, a finding that will be used to construct a first stage prediction of whether households received

[22] South Africa's nominal GDP per capita is 2018 was $6,331.

[23] This location data was accessed with the help of Jeremy Seekings of the Centre for Social Science Research, University of Cape Town, and David Lam of Population Studies Center, University of Michigan in January 2011.

housing.

Table 2: Proximity data (using household baseline location, and promixity to public housing projects built before wave 3)

	Mean	Min	Max	N	Control	Treat	Diff
Distance from housing projects in kilometers							
Distance Proj 1	0.88	0	16.7	970	1.13	0.41	-0.73***
Distance Proj 2	2.41	0	28.6	970	2.82	1.69	-1.13***
Distance Proj 3	3.24	0.046	31.2	970	3.60	2.57	-1.04***
Household block is among closest 3 blocks in the dataset to nearby project							
Closest3 Proj 1	0.39	0	1	970	0.35	0.48	0.13***
Closest3 Proj 2	0.11	0	1	970	0.089	0.16	0.068***
Closest3 Proj 3	0.10	0	1	970	0.067	0.17	0.101***
Relocation distances in kilometers							
Distance from CBD	25.8	4.05	53.3	968	24.7	27.8	3.07***
Moved home (dummy)	0.36	0	1	970	0.34	0.40	0.060
Distance moved	1.53	0	36.8	968	1.37	1.83	0.46

I use household location at the baseline (wave 1 of the panel) to generate instrumental distance variables at all times, regardless of whether households moved home between waves of the panel. However, I was able to clean the household address data, which was updated with each survey wave, and therefore track households movements over time, and calculate changes in distances from the city, and services, between waves. I use these as outcome variables, which are summarized in Table 2. Households in this study live very far from the city centre at baseline. Roughly 30% of the sample move at some point during the study.[24] The average distance of a move within the city is relatively small.

4 Empirical Strategy

This paper proceeds in three main steps to identify the causal effects of government housing on households. First, I look at Panel OLS regressions with household fixed-effects to estimate the impact of receiving housing. Second, using a natural experiment, I instrument for individual selection into treatment (receiving a house) by using proximity to government housing projects. Third, I use a set of housing projects that were planned but not built in order to control for selection at the geographic level, by dropping from the control group those areas that never had projects planned nearby.

4.1 Household fixed-effects

Equation 1 is my main fixed effects OLS estimator of the effect of winning household T_{it} in period t. Note that my use of the fixed-effects estimator assumes a constant effect of T on

[24]Most of these moves were within the boundaries of the City of Cape Town, but there were a few households that moved back to rural areas in the Eastern Cape or KwaZulu Nata, some hundreds of kilometres away.

y regardless of how long each household has been living in government housing.[25] I show the demeaned version of the fixed-effects estimator in Equation 1 in Equation 2, where $\widetilde{y_{it}} = y_{it} - \bar{y}_i$. I later use this to introduce instrumental variables into the fixed-effects framework.

$$y_{it} = \lambda_t + X_{it}\beta + T_{it}\tau + \gamma_i + \epsilon_{it} \tag{1}$$
$$\widetilde{y_{it}} = \widetilde{\lambda}_t + \widetilde{X_{it}}\beta + \widetilde{T_{it}}\tau + \widetilde{\epsilon_{it}} \tag{2}$$

The fixed-effects estimator identifies the effect of government housing under the assumption that treatment is uncorrelated with individual time shocks. There are reasons to think this assumption may be violated in this context. There have been widespread reports of manipulation of the housing allocation lists, with certain individuals receiving preferential treatment based on political connections or other means to access housing (even paying bribes) (Seekings et al., 2010; Tissington et al., 2013). If households who received windfalls or good new jobs were able to leverage their increased incomes to access housing, this could bias the estimates upwards.[26]

I find that housing is more likely to go to households that were poorer at baseline. See Table A.3 in the online appendix. Households who are given housing are more likely to be black, more likely to be headed by a woman, having lower employment rates among household heads, and earn less than households who do not get housing. Furthermore, households that suffer negative income shocks might be more likely to be awarded housing. This would bias the estimates of the impact of housing downwards, as households are less likely to experience increases in their incomes are most likely to get housing.

The long waiting lists for houses could also also lead to selection bias. Many households that get treated are likely to be the ones who have remained in informal dwellings the longest, making them high up the community waiting lists. Those who were able to get out of poverty and upgrade dwellings on their own are, by definition, off the waiting lists (or at least out of my sample of eligible individuals). Finally, measurement error could be a source of downward bias: the extent of measurement error in the sample could be substantial, especially in the measurement of incomes, and even in the treatment variable.

[25]In other words, if a household without housing at time t receives a government housing at time t, my model imposes that assumption that the household will experience an average increase in y equal to β in period t but also $t+1$ and every period after that. I do not have sufficient data to estimate the trajectory of impacts over time, but provide some suggestive evidence on this later in the paper.

[26]Given the roll-out of numerous government programs at the same time as the housing project, it is possible that households that managed to get government housing, also received other benefits simultaneously, which might improve their economic outcomes

4.2 Natural Experiment: allocation by proximity

I deal with non-random selection into housing at the individual level through use of an instrumental variables (IV) estimator, using distance from housing projects as instrument for selection into housing projects.[27] I use the location-based allocation policy used by the government as a natural experiment, whereby individuals who lived close by to newly built housing were more likely to be allocated to live there.

I gathered detailed information on the housing allocation procedures used by the local city government, based on numerous meetings and discussions with officials in the local government in early 2011. Policy documents, reports and research papers corroborate my findings (Tshangana, 2009; Seekings et al., 2010; Tissington et al., 2013). Households that get the housing need to be eligible and registered on a national waiting list. But as a result of the project-by-project nature of the rollout, households were selected into projects according to catchment areas around the projects. From these areas, some communities were selected as potential stakeholders. These stakeholders were allocated a certain quota of housing units from the project (Tshangana, 2009).[28] Communities establish project committees responsible for allocating housing to their members, with the restriction (not always enforced) that all selected candidates much be on the housing waiting lists.

In this way, households that were living close to housing projects that were built between 2002 and 2009 were more likely to be treated than those living further away. This identification strategy identifies the effects of the program under the assumption that geographic proximity to new projects is uncorrelated with changes in household outcomes, except through the channel of improved housing.

4.2.1 IV estimation

I use an instrumental variable Z_{it}, summarizing a households proximity to nearby housing projects, to estimate a fixed-effects-two stage least squares (FE-2SLS) estimator, given by

$$\widetilde{y_{it}} = \widetilde{\lambda}_t + \widetilde{X_{it}}\beta + \widetilde{T_{it}}\tau + \widetilde{\epsilon_{it}} \tag{3}$$

$$\widetilde{T_{it}} = \widetilde{\lambda}_t + \widetilde{X_{it}}\pi_1 + \widetilde{Z_{it}}\pi_2 + \widetilde{\epsilon_{it}} \tag{4}$$

[27] In this way I follow McKenzie and Seynabou Sakho (2010), Attanasio and Vera-Hernandez (2004) and Ravallion and Wodon (2000), who use distance from tax registration offices, community centres and schooling project, respectively, to control for selection into social programs. This fits with a larger literature of using geographic instruments. Dinkelman (2011) and Klonner and Nolen (2010) use terrain data to instrument for the placement of electrification programs and mobile phone antennas, respectively. These papers follow a methodology pioneered in Duflo and Pande (2007) to evaluate the growth impact of dams. Similarly, Banerjee et al. (2012) uses distances from major roads built across China to evaluate the impact of these roads on local growth.

[28] In a few cases a certain number of units would be reserved for households outside of the catchment area, usually communities that had been waiting for houses for a particularly long time, or had been recently relocated. An example is the Joe Slovo informal settlement near Langa, which was allocated housing in the N2 Gateway Project due to a fire that affected that community.

where Equation 4 gives the first stage prediction of the probability of switching to be treated (receiving a house) from non-treated in time period t. The fitted values for $\widetilde{T_{it}}$ are then used as regressors in Equation 3.

I aim to use an aggregate measure of exposure to treatment as a function of distance to multiple housing projects. The distance from a single (closest) housing project is not entirely informative about the probability of treatment by itself, nor is the average distance to multiple projects. I have time-varying data on the pair-wise distances d_{iat} for each household i and each housing project a, but including all of these in as instruments Z_{it} would lead to a problem of too many, likely weak, instruments. Instead, I want to capture the combined contribution of all nearby housing projects on the probability that a household is treated.

To do so, I use maximum likelihood estimator across the three waves of post-baseline data, and distances to newly built projects before each respective wave. I estimate the probability of treatment by a non-parametric function $G(D;\rho) = P(T = 1|D)$, which uses multiple distance instruments D and a common parameter ρ determining the impact of distance on the probability of treatment. I calculate time-varying instruments to use in my panel setup. Since the number of completed projects increases monotonically with time, these distance instruments are weakly monotonically decreasing from year to year, which allows the model to account for the increased probability of receiving a house over time. The probability that a household is treated by time period t is the joint probability that the household is treated by *any* of the nearby housing projects, where the probability of being treated by a given project a is given by a logistic function of distances to projects completed up to that point in time d_{iat}:

$$P(T_{it} = 1|D) = 1 - \prod_a^A (P(T_{iat} = 0)) \tag{5}$$

$$= 1 - \prod_a^A \Lambda(-d_{iat}\rho) \tag{6}$$

This estimator has several advantages over a linear specification. First, multiple distances to multiple housing projects influence the probability that a household will be treated. Second, the effect of nearby projects decays with distance in a non-linear form, but this decay parameter is the same regardless of whether a project is closest to a housing project or not. Third, controlling for distance, the marginal effects of a new project are decreasing with each additional nearby project, because households cannot be treated more than once. And finally, controlling for distance, I can allow for the marginal effect of a nearby project to be greater if a household is ranked among the closest areas to the project, by expanding the list of instruments to include rank variables R in the function form in Equation 5: $P(T_{iat} = 0) = \Lambda(-d_{iat}\rho - r_{iat}\eta)$.

I estimate equation 5 using a maximum likelihood estimation in a panel with all three post-baseline time periods, with time-varying distance variables, and with time fixed effects including to account for possible changes in the rate of delivery across years. I then generate fitted probabilities of treatment $\widehat{G_{it}}$, for each in each period using the full set of proximity variables. I then use those predicted values as an instrument for treatment status T_{it} in the FE-SLS estimator. In other words I estimate Equation 4, with $Z_{it} = \widehat{G_{it}}$. Crucially, I do not use the fitted probabilities of the probability of treatment $\widehat{G}$ as regressors in the second stage of the usual 2SLS estimator (Equation 3).[29] In Section A.2.1 in the Appendix I discuss this functional form in more detail. Taken together, I will estimate a system of equations taking the following form:

$$\widetilde{y_{it}} = \widetilde{\lambda}_t + \widetilde{X_{it}}\beta + \widetilde{T_{it}}\tau + \widetilde{\epsilon_{it}} \tag{7}$$

$$\widetilde{T_{it}} = \widetilde{\delta}_t + \widetilde{X_{it}}\delta_1 + \widetilde{\widehat{G_{it}}}\pi + \widetilde{v_{it}} \tag{8}$$

$$\widehat{G_{it}} = G(X_{it}, D_{it}; \widehat{\rho}) \tag{9}$$

4.2.2 First Stage Results

Table 3 shows the results of the estimatino of Equation 5 by maximum likelihood. There are large and significant impacts of distance from housing on the probability of receiving housing. Living in an area that was used a site for a housing site is also a significant predictor of treatment, because communities that were evicted to make way for housing were given priority access to the new housing.The combined effect of the distance has enormous predictive power on the probability of treatment in the data. I use the coefficients from the estimates in Table 3, Column (4) to predict treatment in each time period ($\widehat{G_{it}}$).[30] I then perform linear 2SLS using $\widehat{G}$ as an instrument in the first stage in Equation of 8. The first stage estimation of T on $\widehat{G}$ is shown in the Online Appendix, Table A.4.

4.3 Cancelled or Delayed Projects

The third and final stage of the estimation strategy deals with the possibility of non-random placement of housing projects. The identifying assumption of the instrumental variables strategy is that the chosen location of projects is not correlated with changes in households outcomes over time, except through the channel of government housing. Anecdotal reports suggest that communities had very little power in initiating new housing projects, which were mostly driven by the land demand needs of private construction companies. Land availability and affordability were the main forces determining construction locations. This meant that housing projects were usually built in areas where land was relatively abundant or cheap, or

[29] This method is adapted from Wooldridge (2002), who refers to this as using 'generated instruments', as opposed to 'generated regressors'. If the instruments Z are informative and valid, then $G(x, z; \hat{\rho})$ will be too. Wooldridge (2002) shows that in the IV framework, we can ignore the method of estimation of ρ in the first stage. Inference in the 2SLS with $\widehat{G_{it}}$ as instruments is consistent, and no standard error corrections are required.

[30] I plot the kernel density of predicted treatment by those that actually received housing, and those that did not, by the final wave of data from 2009, in Figure A.3 in the appendix. The results show a very clear right shift in the distribution for treated households.

Table 3: Maximum likelihood estimation of treatment status

	(1)	(2)	(3)	(4)
Project distance (km)	-0.845***	-0.934***	-0.704***	-0.499***
	(0.109)	(0.117)	(0.107)	(0.100)
Project distance squared		0.0118***	0.00884***	0.00622***
		(0.00163)	(0.00151)	(0.00146)
Rank among closest areas to projects			0.794***	0.687***
				(0.127)
Area cleared for project				1.113***
				(0.176)
Obs	2,694	2,694	2,694	2,694
Time FE	Yes	Yes	Yes	Yes

Notes: Clustered standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1. Dependent var is a dummy for treated in current period. Each coefficient on the project-household pair variables estimates the common parameter specifying the effect of that variable on the probability of being selected for a project. Project dis: the coefficient of the distance from a housing project to the household. Rank: dummy variable = 1 if the enumeration area of the household is among the three closest EAs to the project. Project dis sq: project distance squared. *Area cleared for project*: dummy variable = 1 if the enumeration area of the household was contained within a housing project that was build within the last year, so that the household had to move to make way for the development.

in parcels of undeveloped within the city. Still, it is possible that certain connected individuals were able to lobby effectively for housing on the part of their communities. These individuals have been successful at lobbying for other services or employment projects. Further, the government may have prioritized the development of certain neighbourhoods or areas for political reasons and simultaneously awarded those areas other social programs. If this is the case, the results could be biased by the placement of new projects.

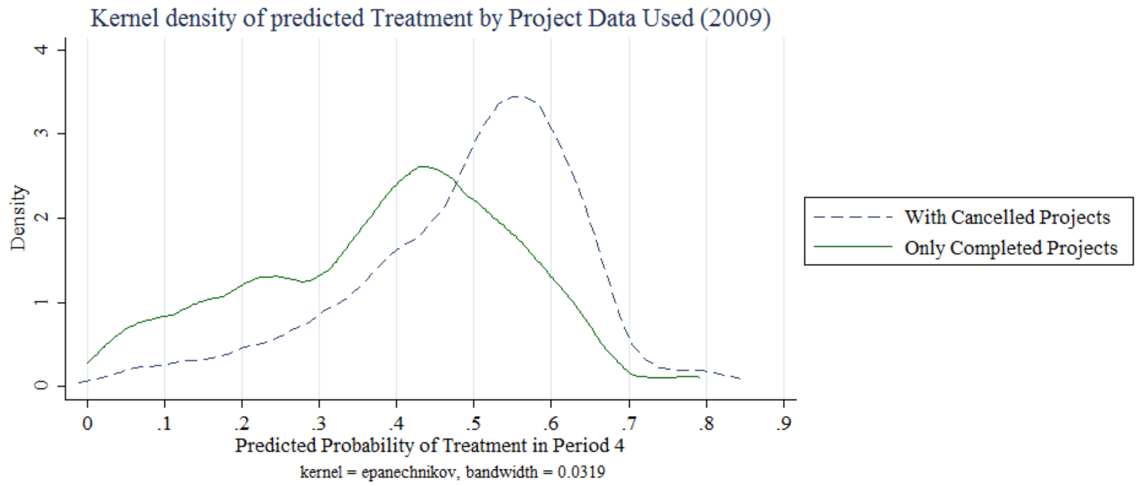
I use a natural experiment using housing projects that were planned and approved during my study period but cancelled or delayed in their implementation. I compare households that lived near planned but incomplete projects to those that lived near housing projects that were completed while excluding from the sample those households that lived in areas where projects were not even planned.

To do this, I generate a new dataset of distances from housing projects, this time *including* projects that were cancelled or delayed until after the end of the panel data.[31] This new dataset of distances, combined with coefficient estimates from Table 3, is then used to predict the probability of having been treated had the incomplete projects been completed. These predicted probabilities are the counterfactual probabilities of treatment had all housing been built. Mechanically, the predicted probability of treatment ('placebo probabilities') is considerably higher once the incomplete projects are included in the prediction. Figure 2 shows the predicted probabilities of treatment with and without incomplete projects. Clearly, in-

[31] I verified that incomplete projects had indeed been not been built, and that completed projects had indeed been completed, by using satellite imagery from the time.

cluding the cancelled projects induces a large right-shift in the distribution, suggesting that many households would have had a considerably higher chance of getting a house, had the cancellations and delays not happened.

Figure 2: Predicted probability: comparison with and without incomplete projects



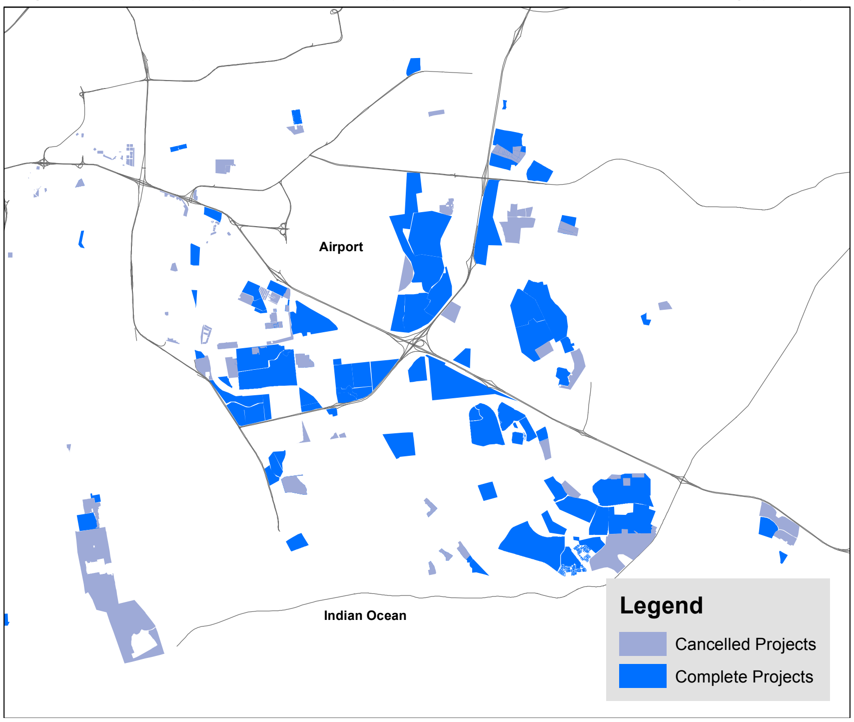
I then trim the sample by dropping households that never had planned projects. I drop individuals with a low probability of treatment with incomplete projects included.[32] I then re-estimate the impacts of receiving housing using trimmed samples, using the fixed-effects estimator, and iteratively dropping quintiles of the "placebo probabilities'. If the results are driven by non-random placement of projects, one would expect the results to disappear as I drop these areas with no planned projects. As a complimentary robustness check, I re-estimate my fixed-effects and IV results while restricting the sample only to areas that received completed projects, to see whether the results are similar when comparing individuals who got housing that those that didn't, only in areas where housing was built.

I conducted detailed discussions with City officials about the reasons for cancellations or delays in the implementation of projects. I was told that when projects were held up after being approved by politicians the problems were usually related to problems with the construction contractors in charge of delivery the housing including financial difficulties, legal problems or disputes between the contractors and the city government, as opposed to something specific local communities in the project area. Unfortunately, there is no systematic database of projects and their reasons for cancellation. I selected a random set of roughly 12 cancelled or delayed projects from my database and investigated their histories through press reports and government documents. In the City of Cape Town's annual reports on housing

[32]Since many individuals with a relatively high probability of treatment when cancelled projects are included have a relatively low probability of treatment with only completed projects, the distribution of the predicted treatment once the sample has been trimmed still has support over the full range of predicted probabilities. This is illustrated in Figure A.4 where all households with a probability of treatment less than 40% are dropped from the sample.

delivery I often saw projects that were listed in the set of planned projects at by 2003 or before showing up repeatedly in government planning documents in later years with no explanation of why they were not completed earlier.[33] Based on this small sample, I found that about a third of incomplete projects were cancelled outright, about a third were at least partially completed, while another third are still discussed in planning documents, although have evolved in design over the years in line with changing national guidelines. For one example, a housing site called "Pelikan Park" appears in my planning data as early as 2000, and on the planning maps as an enormous housing project in the south of the city. News reports online reported a relaunch of the project in 2012. Today it appears that some of the housing as been built, but much of the planned housing site still remains as vacant land.

Figure 3: Comparison of complete and cancelled housing projects



Delays to housing projects are a common problem nationally. The National Housing Development Agency produced a report on how to minimize these delays "Guidelines for Blocked Housing Projects".[34] This report mentions a few main reasons for delays after the project approval stage. "Emerging contractors unable to provide bridging finance or carry operational costs, leading to delays in implementation" and "Employment challenges – wage disputes, causing delays". The report does also mention "Political instability, resulting in dysfunctional councils and municipalities – impacting on decision making and delivery / implementation"

[33]See for example The City of Cape Town Integrated Human Settlements Five-Year Strategic Plan July 2012 – June 2017 2016/17 Review available online at: https://tdacontenthubstore.blob.core.windows.net/resources/3bb1cd9a-233b-47b2-9719-21ef9a16ac83.pdf

[34]Available online at http://thehda.co.za/pdf/uploads/multimedia/HDA_Blocked_housing_projects_guidelines.pdf

as a reason for delay or cancellation. This suggests that local municipality decisions would be driving cancellations. For this reason, I will check for the robustness of my results to including local area and ward specific time trends. I cannot rule out that there were cases where projects were delayed or cancelled for reasons related to local communities. Certainly, there are news reports about community objections or negotiations delaying projects. However, there is no evidence to suggest that this was particularly likely to happen in certain types of neighbourhoods.

Figure 3 shows a map of the incomplete projects along with completed projects, which gives an idea of the variation in treatment probabilities induced by project cancellations. In total, the paper uses data on 78 cancelled projects, to the 109 completed projects used in the main analysis. As of 2011 some of the projects in the cancelled lists were still delayed in their implementation. Following up in 2019, I found the majority of these projects had never been completed. To test for systematic differences in the projects that were completed and those were not, I compare the size and location of the two groups of projects. The differences are not meaningful or statistically significant. Completed projects are 0.33 square kilometres in size on average, compared to 0.28 sq km for cancelled projects. Completed projects are 23.1 km from the city centre on average, compared to 22.8 km for cancelled projects.

5 Main Results

5.1 Household income

To estimate the effects of government housing on income, I proceed in the three steps outlined in Section 4. I estimate regular OLS models with household fixed effects, then use instruments to deal with selection on individual unobservables, and finally re-estimate both the FE and IV result with a trimmed sample that excludes areas that were far away from both cancelled and completed projects.

Columns 1 to 3 of Table shows the fixed effects estimates of the impact of receiving a house on the log of household income.[35] The estimated effect on income is 15.5 log points without controls and remains stable and significant when I control for changes in household size (Column 2) and then for changes in household composition that could be driving the effects (Column 3). With these controls included, I estimate an effect of 0.165 log points or 18%. In Table A.5, I show that these results are robust to restricting my sample only to households that appear in all four waves of my panel.

I then turn to the IV results. I use the predicted probabilities from Table 3 (Column 1) in each period as instruments in a two-stage least squares estimator. The IV shows large and significant effects of housing on household income. These estimates are considerably noiser

[35]This is the most comprehensive income variable which includes data from a one-shot total household income question, which include a number of sources of non-labour related income, but excludes income from rent.

Table 4: Effects of government housing on log per capita total household income

	(1)	(2)	(3)	(4)	(5)	(6)
	FE	FE	FE	IV1	IV1	IV1
	lgincpc	lgincpc	lgincpc	lgincpc	lgincpc	lgincpc
house	0.155***	0.169***	0.165***	0.570*	0.620*	0.547*
	(0.0509)	(0.0517)	(0.0510)	(0.315)	(0.329)	(0.316)
Observations	3,626	3,595	3,587	3,572	3,564	3,552
HH Size	No	Yes	Yes	No	Yes	Yes
HH Controls	No	No	Yes	No	No	Yes
R-squared	0.200	0.222	0.250	0.194	0.198	0.232
Number of personid	1,097	1,082	1,082	1,059	1,056	1,052
Weak IV F				97.26	84.77	86.17

Notes: Standard errors clustered at the enumeration area level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. house=1 if household reported getting a subsidized house at any point in the past. Dependent variable is the log of per capita household income.

than the OLS results, although the first stage is strong.[36] The IV coefficients are generally larger than the OLS results, which indicates poassible correlation between the probability of treatment and *negative* income shocks.[37] As discussed in the conceptual framework, this could be due to households in worse circumstances (like the aged or recently unemployed), or victims of recent shocks, being assigned houses by their communities. They might have expected them to have done considerably worse without the treatment. This would lead to downward bias of OLS estimates. In addition, measurement error in the housing subsidy status of households could lead to bias in the OLS estimates. On the other hand, as I discuss in more detail in Section 6.5, these IV estimates represent the local average treatment effect for households that won housing without having to relocate very far, who may have benefited more than the average household that had to relocate further. Taken together, the OLS estimates can be interpreted as a lower bound for the average effect of government housing, while the IV estimates are a (noiser) upper bound for the effect.

5.2 Earnings for men and women

Next, I show that these results are due to higher wage earnings for members of the household. Table 5 shows the OLS fixed-effects results for impacts on raw total household salaries. My estimate of the effect on total household salaries is very much in line with the increase in log total income: an increase of R465 on a mean of R2443 or 19 percent (or 18 percent when winsorised to deal with outliers in column 2). I then break down this average effect into impacts by gender. Here I look at total earnings at the household level by men and women

[36] I report the Kleibergen-Paap Wald F statistic for weak instruments in the last row. I am always able to reject the null hypothesis of weak instruments.

[37] I report post-estimation tests of endogeneity, which, in the fixed effect IV setting is the Davidson-MacKinnon F statistic. This test rejects the null that the IV estimates are the same as the OLS model. See Davidson and MacKinnon (1993).

Table 5: Effects of government housing on total household earnings (OLS with fixed effects)

	(1) total earnings	(2) total earnings winsorised	(3) total female earnings	(4) total male earnings	(5) total female earnings recoded	(6) total male earnings recoded
house	465.8***	401.9***	264.6*	191.7	266.6**	201.3
	(167.3)	(133.0)	(148.1)	(157.0)	(133.7)	(132.9)
Observations	3,699	3,699	3,438	2,902	3,699	3,699
R-squared	0.063	0.084	0.023	0.042	0.024	0.066
Households	1,064	1,064	1,045	985	1,064	1,064
Mean dep var	2443	2208	1173	1652	1093	1335

Standard errors clustered at the enumeration area level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. house=1 if household reported getting a subsidized house at any point in the past. Dependent variables are total household earnings from wage income only. All regressions include controls for household demographics. Column (2) reports a version of total househod earnings winsorised at the 5th and 9th percentiles. Column (3) reports total earnings for all female members of the household, while Column (4) reports total earnings for all male members of the household. Fewer observations in Columns (3) and (4) reflect households without any adult female or male members , respectively. In Column (4) the lower sample size reflects the high number of single-mother or female-headed households in South Africa. For robustness, Columns (5) and (6) report the measures in Columns (3) and (4), respectively, but with missing values recoded to zeros. For example, in Column (6) a household with no male members would be coded as having zero male earnings rather than having this coded as missing, as is the case in Column (4).

separately. Some households do not have either any adult women or adult men. Columns 4 and 5 estimate the effect on household totals where such cases are coded as missing for the respective genders, Columns 5 and 6 recode these to zeros: for example household with no adult men take on values of zero for total household male earnings. The results are similar. Women's earnings increase by a significant R264, while for men the increase is R191.

5.3 Individual level regressions

To better understand the source of the increase in household earnings, I turn to the individual level data used to construct my aggregate earnings used above, from the household roster in my data. This allows me to confirm the existence of individual-level earnings increases, and to look more closely at heterogeneity at the individual level. Also, by studying panel of individuals, restricted only to individuals who observed at least twice in the panel, I rule out that my household-level results are driven by the arrival of new household members. Table 6 shows the main findings, estimating the effect on earnings for men and women separately. In this regressions, I include only working-age individuals, between the ages of 18 and 60, and I include individual fixed effects in all regressions.

First, the results confirm the finding that average earnings for women increase significantly, while there is no significant impact for men. The results are robust to winsorizing my raw individual earnings measure to remove outliers.[38] Columns 3 and 4 then provide a sense of where the increase in average earnings comes from. Receiving a government house leads

[38] I winsorize at the 99th percentile.

Table 6: Impacts on individual income among men and women (OLS with fixed effects)

	(1) Earnings	(2) Earnings winsor	(3) Earning (conditional)	(4) Employment
house*woman	142.6**	116.2*	219.6*	0.013
	(68.5)	(63.0)	(123.9)	(0.032)
house*man	-0.3	3.3	66.9	0.013
	(106.4)	(92.2)	(152.5)	(0.040)
Observations	5,896	5,896	2,527	6,187
R-squared	0.057	0.066	0.162	0.007
Individuals	2,194	2,194	1,395	2,194
Men=Women (p)	0.258	0.311	0.436	0.997
Mean (women)	618.6	605.8	1667	0.418
Mean (men)	1088	1056	2141	0.564

Notes: Standard errors clustered at the enumeration area level in parentheses. Sample is restricted to individuals of working age: ages 18 to 60. All regressions include individual fixed effects. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Column (2) replicates Column (1), restricting the sample only to individuals who appear in the panel at least three times in the data.

to an increase in employment, at the extensive margin, of 1.3 percent; this is a small effect and not significant. Rather I find a significant impact of government housing on earnings among women who are working. In column 3, I look at the impact of government housing on non-zero earnings (earnings conditional on having a job) for individuals who appear with non-zero earnings at least once in the panel.

To investigate further, I conduct subgroup analysis at the individual level among women, in Table 7. Column headers indicate the *baseline* covariate by which the sample is split: row 1) of each header provides the first group, row 2) the second. This is mirrored in the reporting of the differential treatment coefficients below. For example, column 1 splits the sample by whether an individual woman was working at baseline. In row one, I report the coefficient for women who were working (cov=1) while row 2 reports the coefficient for those who did not have work (cov=2). Column 1 shows that the effects on individual earnings are concentrated among women who were already working at baseline. How were women who were already working at baseline manage to increase their earnings? When I restrict my sample to working women, I find that the coefficients are larger in magnitude among women who were working part-time at baseline, as opposed to those working full time.[39] Power is somewhat limited when cutting my sample in this way, so results should be viewed with some caution. I am unable to conclusively reject that the coefficients are significantly different across subgroups. The effects are significantly larger for women who the heads of their respective households.

Unfortunately, I do not have data on hours worked for all individuals in my data. And

[39]Table A.6 replicates this analysis for employment as the outcome variable and shows no evidence of impacts on employment in any particular subgroup.

Table 7: Heterogeneity in effects on individual female earnings

	(1)	(2)	(3)	(4)	(5)
Baseline covariates	1) work 2) no work	1) part time 2) full time	1) over 30 2) under 30	1) Not single 2) Single	1) HH Head 2) Not head
house*cov=1	226.0**	250.9	194.8**	190.5*	371.9***
	(110.8)	(215.0)	(88.5)	(105.0)	(117.0)
house*cov=2	81.4	169.2	64.1	76.3	56.3
	(74.5)	(181.4)	(82.1)	(76.8)	(72.7)
Observations	3,739	1,056	3,739	3,739	3,739
R-squared	0.0708	0.0173	0.0891	0.0668	0.0748
Households	1,329	363	1,329	1,329	1,329
Cov1=Cov2 (p)	0.279	0.772	0.271	0.380	0.0220
Mean (cov=1)	1027	836.7	703.7	633.9	781
Mean (cov=2)	475.7	1146	557.7	611.9	572.9

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors clustered at the enumeration area level in parentheses. Sample is restricted to women of working age: ages 18 to 60. All regressions include individual fixed effects. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. Each Column shows heterogeneous treatment effects by a particular coviarate. Column subtitles specify first a covariate category 1 and then category 2 below that. Results below are then divided by that category. For example, in Column (1) I show heterogeneous treatment effects by whether women *i* was working at baseline. Category 1 looks at women who worked. Column 2 restricts the sample to only women who were working at baseline, and the looks at differential treatment effects according to whether they had part time or full time work. "Cov1=Cov2 (p) " shows the p-values from t-tests of equivalence of the category 1 and category 2 coefficients.

the measurement of part-time and full-time work was only conducted at the baseline for this survey, and so I cannot formally test whether the effects were driven by women switching from part-time to full-time work. However, my data does include more detailed labour market information on a subsample of young adults in the households in my sample. I am able to estimate the effects of government housing on employment outcomes for these individuals. Results are presented in Table 8. The effects on working hours are sizeable. Again, I look at hours *conditional* on working and find that these increase by 0.7 hours per day on average, and increase equivalent to just less than 10%. Unfortunately, I am unable to show similar results among older members of the family, but the young adult population makes up nearly 50% of the total individual sample size used in Table 6. The effects on employment at the extensive margin are consistent with those in the household as a whole; they are positive, but not quite significant. While I cannot confirm the existence of analogous effects among the older population, the young adult sample makes up nearly 50% of my working-age adult sample (the sample used in Table 6), and so provides at least some evidence that my main results are driven by increases in workings hours.[40]

[40] Recall that the effects on earnings are smaller in magnitude for young adults. If the effects on older women do indeed come through the channel of increased working hours, one would expect the effects to be bigger than those for young adults too.

Table 8: Impacts on Young Adult Emplomen t (including hours worked)

	(1) FE Employed	(2) FE Employed	(3) FE Hours work conditional	(4) FE Hours work conditional	(5) FE Hours	(6) FE Hours
house	0.026	0.030	0.711***	0.680***	0.650*	0.622*
	(0.041)	(0.038)	(0.261)	(0.257)	(0.365)	(0.372)
Observations	2,695	2,737	1,590	1,554	2,790	2,722
R-squared	0.118	0.122	0.018	0.031	0.146	0.153
Number of personid	1,061	1,047	827	808	1,069	1,047
Av Group	2.540	2.614	1.923	1.923	2.610	2.600
Mean dep var	0.422	0.422	8.330	8.330	4.998	4.998

Notes: Standard errors clustered at the enumeration area level in parentheses. Sample is of young adults in treated households over the age of 18. Mean age of the sample is 22. The ninety-fifth percentile of age is 28. All regressions include individual fixed effects. house=1 if household reported getting a subsidized house at any point in the past. *** p<0.01, ** p<0.05, * p<0.1.

5.4 Robustness of the main results

Before turning to a discussion of the mechanisms that could be driving my main effects on earnings, I prodvide a series of robustness checks for my headline findings.

5.4.1 Robustness using cancelled projects

I use the data on cancelled projects to deal with non-random selection of project sites, to rule out that my results are driven by the placement of housing sites in locations on different trajectories. I use the predicted probability of receiving housing including projects that were cancelled and then drop areas with a low predicted probability of treatment. I start by dropping the 20% of households with the lowest placebo-probability of treatment, then 40%, and so on. The remaining variation in the probability of receiving housing under the true predicted probability of treatment (excluded cancelled projects) is then driven by the projects that were built versus those that weren't built. The more of the sample that is trimmed, the more of the remaining variation is due to project cancellations alone.

I find that the results are robust to the trimming the sample in this way. Table 9 shows the impacts on total household earnings (Panel A) and for female earnings (Panel B) outcomes, both using OLS with fixed effects. The coefficients are similar to those in the full sample and significant as I gradually drop more of the sample. Only when I have trimmed a whole 80% of the sample (Column 4) are the impacts no longer significant: the sample size becomes too small to intepret these results meaningfully. The estimated coefficients are slightly smaller, are very similar across the distribution. In the Online Appendix, Table A.7, I show that the main IV results are also robust to this trimming proceedure.

Table 9: Impacts on household income with trimming using cancelled projects

	(1)	(2)	(3)	(4)
% Sample Trimmed	**20** percent	**40** percent	**60** percent	**80** percent
Panel A: OLS with Fixed Effects Impacts on Log Total Income				
house	0.163**	0.203***	0.163**	0.116
	(0.0650)	(0.0750)	(0.0793)	(0.0799)
Observations	2,874	2,219	1,507	743
Households	831	625	417	206
R-squared	0.272	0.271	0.328	0.486
Panel B: OLS with Fixed Effects Impacts on Total Household Female Earnings				
house	229.9*	367.3**	413.4**	81.0
	(129.6)	(154.1)	(186.8)	(167.6)
Observations	2,943	2,284	1,565	803
R-squared	0.021	0.018	0.017	0.105
Households	865	652	432	219

Notes: Standard errors clustered at the enumeration area level in parentheses. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is indicated in Panel titles. *** p<0.01, ** p<0.05, * p<0.1. Columns (1) - (4) replicate the main estimates in the paper but with the sample restricted to households with increasingly higher *ex ante* probability of treatment according to proximity to *planned* housing projects. For example, Column 1 I drop the 20% of the sample with the lowest placebo probability of treatment and reestimate my main fixed effects regressions. See Section 4.3 for more details.

5.4.2 Other robustness checks

I perform several robustness checks on the main results. First, I check whether the results could be driven by mobile and opportunist households moving to areas where projects were going to be built, to benefit from the housing. Such behaviour would violate the exclusion restriction as the location of households would not be uncorrelated to the placement of projects: more motivated households would be living closer to housing projects. In practice, this seems unlikely as these households would join the back of the queue for new housing, behind long-term residents. Still, I want to rule out this possibility. Because I use household location in 2002 as instruments, I am not worried about households that move to a new area *after* 2002. Instead, I am worried about households that might have moved to their neighbourhood just before 2002 in order to get access to housing shortly after. I drop from the sample all households that received a house between 2002 and 2005, then replicate the main results presented in the paper with only households that were treated after 2006. Table A.13 shows that almost all of the results presented above remain robust to this check[41].[42] Thus if the results are driven by successful households relocation decisions, it would have to have been that they moved to

[41]Panel A shows the fixed effects, Panel B the IV estimates

[42]Although standard errors are larger since so many households were treated between 2002 and 2005.

their 2002 location in order to pursue a house that they would only get after 2006.[43]

Second, the results may be driven by neighbourhoods that received the housing projects and were also on better earnings trajectories. For instance, if a certain community was able to lobby to get a project, they have been successful in receiving other communities benefits like jobs programs. While comparing areas with cancelled projects to uncancelled projects should account for most of these channels, it could also be that well-organised communities were able to fight project cancellation. To deal with this, I re-estimate my main results with cluster-specific time trends included in my specification. If my results are driven by the fact that local areas that received housing projects were on observably better labour market trajectories the addition of these controls should account for this. Instead, I find that my main results are robust to the inclusion of these trends (Table A.9). My results are robust to including area-time trends at a higher geographic level. Table A.10 shows that my results are robust to including time trend terms for the 12 largest townships in Cape Town.

As another robustness check to deal with the possible placement of projects in higher-growth areas. I restrict my sample only to areas that received projects and compare treated to untreated households within those areas. I look only at clusters (primary enumeration areas) where more than 20% of households received a government house over 4 years (this is exactly half of all clusters, 45% of clusters had no households receiving housing). The results are presented in the Appendix, Table A.12. I find that the results are very similar to the results for total household incomes and salaries. While these results do not account for selection at the individual level, the similarity between the within-group and between-group (IV) estimates is reassuring. I show that the results are not driven by differences between particular ethnic groups or by particular neighbourhoods that were more likely to get housing projects. Housing could have been targeted towards entire townships, racial groups, or areas a certain distance from the city, that were exhibiting other changes at the time.[44] I estimate the effects (FE and IV) *within* specific areas, the two largest townships (Gugulethu and Khayelitsha), and all other areas.[45] I find that the results do not hold within the township of Gugulethu (which is relatively close to the city centre, and relatively wealthy), but the results do hold within the larger area of Khayelitsha, and within other areas scatter around the city. I also show that the results hold when restricting the sample to only black households, to poor households, and areas that were upgraded *in situ*. Sample sizes are small in some of the specifications, but the coefficients remain similar in magnitude to the original results for both FE and IV results for the impacts on household income, suggesting that the results are not driven by different

[43]While housing projects were often subject to delays, it would be unusual if they took longer than five years to build, and would take remarkable foresight for a household to move to a new neighbourhood to wait that long for a house.

[44]Indeed, more housing was built in areas further away from the city, in black areas, and in more deprived areas. See Table A.8.

[45]See Figure A.2. Khayelitsha is further from the centre, while Gugulethu is relatively close. The mean household characteristics for the different areas are presented in Table A.8.

trends *across* communities. These results are in A.14 in the Online Appendix.

Third, I show that the effects are not driven by changes in household composition. All regressions are robust to the inclusion or exclusion of controls for household size and composition, and results are robust to estimating earnings at a per capita or per working-age adult. I do find that receiving government housing is correlated with a higher number of very young children (under the age of three) but there is no impact on other measures of household composition (see Table A.15).

Finally, I show that my results are robust to clustering my standard errors at a level higher than the enumeration area level used as the default for the main results thus far (there are 168 enumeration areas in my data). In Table A.11 in the appendix, I show results where I cluster my data at the voting ward level, of which I have 68 in my dataset, and the findings are qualitatively unchanged. The standard errors are slightly smaller when I cluster at this higher level of geography (Table A.11).

6 Mechanisms: constraints to labour supply

My headline finding is that government housing increases household per capita income by 18 percent. My estimate of the effect on total household salaries is very much in line with that: an increase of R465 per month on a mean of R2,443, or 19 percent (or 18 percent when winsorised to deal with outliers). This average effect can be broken down into impacts by gender of R264 from women's earnings and R191 for men's earnings. The individual-level regressions on raw individual adult income are consistent with the household level results. Earnings among women increase by R142 per adult women. Given that households have, on average 1.6 adult women per household, this plausibly translates into increases in household-level female earnings of close to R264.[46]

The effect on women's average earnings of R142 per month translates to an increase of roughly 23%. What could explain this effect? In this section, I present evidence which is consistent with the theory through which government housing frees up time, particularly for women, that allows them to incrase their employment and earnings.

At baseline, 40% of women work part-time. If all of these women moved to working full-time, thereby doubling their salaries, this alone would account for most of the effect size.

[46]These estimated effect sizes are very much in line with the existing literature in this area. Dinkelman (2011) finds a 16 percent increase in male wages as a result of the average increase in electrification rates of 0.15 percentage points at the district level. Field (2007) estimates the effect of a property title in Peru to be an increase in employment of 23.3 hours per work, or a 25% increase in household labour supply, using her instrumental variables estimates. Given that my study looks at the direct effect at the household level of receiving a house, which provides a combination of improved property rights, physical security, increased space, access to electricity and running water, my estimates do not seem particularly large by comparison. At a macro-scale Greenwood et al. (2005) argue that electricity and modern appliances lead to a reduction in work in the home of 58 hours per week per household in 1900 to 18 hours in 1975. Assuming that 75% of this work falls to female wives or partners, this constitutes a reduction from just over 6 hours to 2 hours a day. In that paper, they partly attribute this reduction in work in the home to a nearly ten-fold increase in female labour force participation.

Second, it is plausible that working hours increased for women, conditional working full-time and part-time. Although I do not have data on all household members' working hours, results looking at young adults for whom I do have this data show that working hours increase by more than 10%, larger than the increase in employment rates of only 3%. It is plausible to think that older adult women also increased their working hours by a similar magnitude. It may also be the case that there are non-linearities in wages-hours profile. That is, part-time jobs in local neighbourhoods may pay considerably less per hour than fulltime jobs that require commutes to the city centre, which would also help to explain the size of the wage effects.

In Section 2.2, I propose three main ways in which housing could free up time. First, government housing comes with improved access to labour-saving technology in the home that could free up time from domestic tasks, the burdens of which fall disproportionately on women. Second, government housing improves home security. This could reduce the need for at least one adult to remain in the home to protect against theft or appropriation. It could also reduce the time mothers have to spend at home looking after school-age children, who can be safely left at home in the hours after school while parents are at work, which may not be possible in a shack. Finally, government housing reduces the need for households to rebuild home after natural disasters and accidents such flooding and fires, which are common in informal settlements in South Africa.

I do not have evidence to show conclusively that the effects on earnings are driven entirely by time-saving technology in the home. That the effects on male earnings that are generally smaller than the effects for women lend credibly to the argument that women's labour in the home is a key mechanism, but I cannot rule out relatively small impacts on male earnings. In Section 6.1 I present evidence that suggests that the effects are indeed driven by improvements in housing quality specifically, rather than other effects of winning the housing like wealth effects. In the subsections that follow, I turn to look for evidence for the channel of time-saving technologies.

6.1 Placebo test: households already in good-quality housing

I re-estimate my specifications looking only at households that were not living in informal housing to begin with. Recall that my results thus far have focussed solely on households that were living in informal (slum) houses at the start of my panel, as I wanted to isolate the effect of moving from informal housing to formal housing. While in theory government housing was supposed to be targeted to households in shacks, in reality, many opportunities went to households that were already in living in brick and mortar homes. Table 10 shows the impacts among these households, who did not experience a marked improvement in home amenities but did experience a similar size wealth shock, at least in absolute terms.

My argument is that the experience of moving from informal housing to formal housing improves labour market outcomes at least in some large part by freeing up time for work in the home associated with living in informal housing. The results in Table 10 help to support this case for two reasons. One, if my main results were the result of omitted variable bias, such as housing being targeted only to improving neighbourhoods or housing coming with additional benefits or access jobs, one would expect these mechanisms to be at play for households regardless of whether households live in slums. Two, if the mechanism at play is not about improved housing quality but rather a wealth effect or a credit market effect, this too would apply to beneficiaries regardless of their original housing conditions.

I find that no evidence of the same positive effects on labour market outcomes among households already living in formal housing. The proportion of households living in formal housing who received government housing is only slightly lower than among households living in informal settlements. This suggests that it is the change in housing conditions that drive my results, and that heterogeneous treatment effects by baseline housing conditions could be quantitatively important for explaining the difference between my main results and the average effects of government housing estimated in Picarelli (2019).

Table 10: Impacts on household income and earnings among households already living in formal housing in wave 1 (OLS fixed-effects regressions)

	(1) Log hh income	(2) HH earnings	(3) Female earnings	(4) Male Earnings
Government house	-0.00726	-163.2	-69.73	-416.2
	(0.0531)	(336.3)	(226.3)	(341.7)
Observations	4,630	4,654	4,554	4,179
R-squared	0.281	0.071	0.035	0.039
Number of personid	1,801	1,800	1,786	1,722
Mean dep var	6.898	3964	1887	2792

Notes: Standard errors clustered at the enumeration area level in parentheses. Sample is all houseoholds in the panel regardless of initial housing conditions. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. *** p<0.01, ** p<0.05, * p<0.1.

6.2 Time-saving technology for domestic tasks

Perhaps unsurprisingly, government housing reduces the probability that households are living in a shack. I find no significant impact on the probability that the household has *any* access to electricity. But housing does significantly increase the access to piped water; there are particularly large effects on having piped water in the home. All of these are technologies that could provide significant time savings for women working in the home. I show these results in Table A.16 in the appendix.

Crucially, I find evidence that households invest in time-saving applications after receiving government housing. These were appliances like fridges and microwave that did not come with the housing automatically but would be far more practical to use, maintain and safe care in a solid dwelling, and could account for the time freed up for women in the home. That households invested in these technologies after getting housing is at least consistent with my claim that housing freed up time from domestic work. Housing reduces the probability that households use paraffin- a form a fuel commonly used in informal settlements in South Africa. Many of the devastating fires that occur in townships in South Africa are attributed to the use of paraffin. There is a negative effect on the occurrence of fires in the home (although this coefficient is not significant, it was only measured once in the follow survey rounds) which could be driven by the reduction in paraffin use.

6.3 Safety and security

One mechanism that could be driving the impacts on increased female labour participation and earnings could be the improved security that good quality housing gives to households. This could allow them to leave the home to take up employment or go in search of work without fear of burglary in their absence. This is related to the hypothesis of Field (2005) who argues households increase their labour supply when security-of-tenure is improved and they less likely to fear expropriation when they are out of the home.

While expropriation risk may be a factor in this setting, de facto tenure security is already thought to be very good in South African informal settlements. On the other hand, security threats from burglary and other crime around the home are far more salient threats in this environment. If this is part of the mechanism driving these results, we should see some evidence that adequate housing has positive impacts on feelings of safety around the home. I provide evidence of this in this in Table A.17. I find that households that received government housing were far less likely to report that they felt unsafe in their homes at night.

Also, home security could play a role in reducing childcare costs, especially for slightly older primary school-aged children who could be left at home in a secure and locked house. Government housing may provide this security in the way that a shack may not.

6.4 Heterogeneity at the household level

Next, I look for heterogeneity in treatment effects by whether households owned specific household assets at baseline, including fridges, stoves and microwaves. When I do this for my main sample of households who lived informally at baseline, the variation in these outcomes is relatively small, since so few households in this group had good quality housing or modern appliances to begin with. Table 11 shows heterogeneity by asset ownership all households regardless of their housing type at baseline. Although the estimates vary across different assets, Column 1 confirms that households with a lower asset indices at baseline experience significantly large impacts on household income.[47]

Next, I want to understand which types of households benefit from informal housing. I

[47] This index is constructed using the sum of the four asset types in Columns 2 to 4 of Table 11, weighted by inverse covariance matrix of these same assets.

Table 11: Heterogeneous impacts on log per capita income by household (all households including those in formal housing in wave 1)

	(1)	(2)	(3)	(4)	(5)
Covariate	Asset Index	Washing Mach.	Elec Stove	Fridge	Microwave
Cov =1	High	Yes	Yes	Yes	Yes
Cov =2	Low	No	No	No	No
house*cov=1	-0.021	-0.025	0.052	0.061*	0.012
	(0.044)	(0.045)	(0.035)	(0.035)	(0.049)
house*cov=2	0.108**	0.092*	0.088	0.071	0.070
	(0.049)	(0.047)	(0.069)	(0.079)	(0.043)
Observations	10,899	10,895	10,895	10,895	10,895
R-squared	0.203	0.203	0.196	0.196	0.203
Households	3,236	3,234	3,234	3,234	3,234
Cov1=Cov2 (p)	0.0591	0.0843	0.641	0.912	0.379
Mean (cov=1)	7.212	7.260	6.941	6.951	7.302
Mean (cov=2)	6.356	6.388	6.280	6.258	6.464

Notes: Standard errors clustered at the enumeration area level in parentheses. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. *** p<0.01, ** p<0.05, * p<0.1 Each Column shows heterogeneous treatment effects by a particular coviarate. Column subtitles specify first a covariate category 1 and then category 2 below that. Results below are then divided by that category.

study heterogeneous effects of government housing by baseline housing characteristics.[48] The results show a clear role of the number of children in the household. Households with three or more children experience significantly larger effects on household income households with fewer children. Similarly households with at least one school-age child benefit more. Other results are inconclusive but point towards larger effects in households where women were already working at baseline, which is consistent with the individual results on female labour supply. The last two columns look at heterogeneity by baseline assets and housing quality, but since households live in almost uniformly bad conditions at baseline, I am careful not to overinterpret these results, which show little evidence of heterogeneity.

6.5 Distance from the city

Government housing in South Africa has been criticised for reinforcing the spatial patterns of segregated living within South Africa cities (Bundy, 2014). Segregation leaves households living far away from jobs and employment opportunities, which is argued to play a causal role in the high rates of urban unemployment for black South Africans (Banerjee et al., 2007). Government housing is often thought to have moved household *further* away from the original place of living, as new housing projects are built increasingly far away.

I confirm that government housing has moved households slightly further away from the

[48] the results are in Table A.18 in the appendix. For these results I return to my restricted sample of households who started out living in shacks.

city, on average, in Table 12, but the magnitude of the effect (roughly half a kilometer) is small. The average distance from the city is about 22 kilometers in this sample. This is in contrast to the results in Picarelli (2019), who finds that government housing causes moves of more than 13 kilometers further from the city centre on average.[49] This likely accounts for the difference between the findings in this paper and Picarelli (2019). It also seems unlikely that changes in distance from the city are driving the effects on labour market outcomes in my study.

Indeed, the IV results discussed in this paper deserve one important caveat, in light of the results on distance above. The impact of housing estimated by instrumental variables should be interpreted as a local average treatment effect (LATE) (Angrist and Imbens, 1994). The instruments identify the effect of treatment on those households that received housing because of their proximity to housing. These are households that were selected to receive housing because of their proximity to housing, and therefore may respond differently to treatment than those who receive housing for different reasons, because they don't have to move very far. They are less likely to have moved a significant distance from their original place of living when receiving housing, precisely because the housing to which they were assigned was close by to their original place of living. Indeed the IV results on the effect of housing on distance are (Column (3) of Table 12) considerably smaller than the FE: suggesting that the IV indeed identifies the LATE for households that did not have to relocate.

Table 12: Effect of government housing on distance from the city center (in kms)

	(1)	(2)	(3)	(4)	(5)
				Movers Only	
	FE	FE	IV	FE	IV
house	0.508**	0.537**	0.0416	1.441**	0.917
	(0.221)	(0.229)	(0.525)	(0.587)	(2.205)
HH controls	No	Yes	Yes	Yes	Yes
Observations	3,765	3,717	3,708	1,243	1,243
R^2	0.008	0.011	0.005	0.028	0.026
Households	1,077	1,077	1,068	362	362

Notes: Clustered standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1. house=1 if household reported getting a subsidized house at any point in the past. Dependent variable citydis is the distance of the household from the city center in kms.

6.6 Alternative channels

I show that sources of income other than labour earnings are not driving the results. In Table A.21 I show that receiving government housing was not correlated with access to other forms of government grants or welfare. Nor is housing correlated with increases in receipts

[49]When I restrict the sample to households for which a move is recorded, the point estimates indicate that treated households moved nearly 1.5 kilometers further away from the city than untreated households who moved.

of remittances or other forms of financial support from other family members or friends.

Finally, I argue that housing projects did not stimulate local employment by creating construction jobs. Construction was always done by external construction companies that brought in their own permanent labour force to the sites. In addition, my results show the biggest impact on the labour supply of women, who are unlikely to have got jobs on construction sites.

7 Conclusion

This paper provides new evidence on the effect of improved housing conditions on labour supply. I study households that were living in slum housing before receiving government and only had to move a short distance further from the city to take up this housing. In this way, I contribute to the literature on the effects of government housing by isolating the direct effects of improved housing from the costs of relocation.

I find that government housing has a large and significant impact on household income and that this effect is driven by increases in earnings from wage labour for female household members. This finding is robust to my instrumental variables estimation, which uses proximity from housing projects to predict selection into housing. It is also robust to the use of cancelled projects to control for non-random location choice for housing projects. I do not find similar effects when I replicate my analysis on a sample of households that were not living in slum housing before receiving government, which suggests that it is the indeed the improved housing conditions that are driving the results. I cannot conclusively ascribe the impact on labour to any one particular mechanism. I show several results that are consistent with a story whereby improved housing freed up time that was otherwise spent on labour in the home, especially for women.

The findings bolster the case for government programs to improve the quality of housing for the poor. Government supply of free housing is a particularly expensive intervention of this kind. However, were this the treatment effects on household earnings estimated in this paper to persistent in their entirety, these increased earnings would account for a substantial share of the total program costs, which are reported to be between R55,000 ($7500 at 2009 exchange rates) and R65,000 around the time of construction.[50] However, evidence from Picarelli (2019)

[50] Accurate estimates of the total cost per unit for specific sites, locations or cities, are surprisingly hard to come by government housing in South Africa. Studies online and government reporting put the cost per unit at closer to R65,000 at the time of the study. I can then compare the effects on income by estimating the net present value (NPV) of the increase in salaries. I start by assuming that the main FE treatment effect on earnings of R465 per household per month is entirely persistent. Discounting future earnings at an annual rate of 5% per year would suggest that government housing pays for itself after twelve to fifteen years. However, if I recalculate the NPV of increased earnings with a housing upkeep cost per year of 2% of initial capital outlay, I find that the benefits of housing would still fall just short of the R65,000 mark after 25 years. I do not have a clear sense of whether the treatment effects decline or grow over time. In Table A.22 I restrict my sample only to the first and last periods (seven years) apart and estimate the effect of winning housing before 2005. I find a slightly higher coefficient than the average treatment effect estimated in main results, but am not powerful to test whether this is significantly larger.

suggests that the effect of the policy post-2009 when housing was generally built further away from beneficiaries' original place of residence was to reduce labour supply. Taken together, and in line with recent work on cities in developing countries (Duranton and Venables, 2018; Michaels et al., 2017; Harari et al., 2018), these findings underscore the importance of delivering improved housing early in the stages of development of a city and finding innovative ways to develop low-cost housing in better locations as cities continue to sprawl outwards.

References

Angrist, J. and G. W. Imbens (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Attanasio, O. and M. Vera-Hernandez (2004). Medium- and long run effects of nutrition and child care: evaluation of a community nursery programme in rural Colombia.

Banerjee, A., E. Duflo, and N. Qian (2012). On the road: Access to transportation infrastructure and economic growth in china. *National Bureau of Economic Research*, 1–22.

Banerjee, A., S. Galiani, and J. Levinsohn (2007). Why has unemployment risen in the new South Africa. *IPC Work Paper Series Number 35*.

Barnhardt, S., E. Field, and R. Pande (2016). Moving to opportunity or isolation? network effects of a randomized housing lottery in urban india. *American Economic Journal: Applied Economics Forthcoming*.

Becker, G. (1965). A theory of the allocation of time. *The Economic Journal 75*(299), 493–517.

Berniell, M. and C. Sánchez-Páramo (2012). Overview of time use data used for the analysis of gender differences in time use patterns. *Background paper for the World Development Report*.

Buckley, R. M., A. Kallergis, and L. Wainer (2016). Addressing the housing challenge: avoiding the ozymandias syndrome. *Environment and Urbanization*, 0956247815627523.

Budlender, D., N. Chobokoane, and Y. Mpetsheni (2001). *A survey of time use: How South African women and men spend their time*. Statistics South Africa.

Bundy, C. (2014). *Short-changed: South Africa since Apartheid*. Jacana.

Cattaneo, M., S. Galiani, P. J. Gertler, S. Martinez, and R. Titiunik (2009). Housing, health, and happiness. *American Economic Journal: Economic Policy 1*(1), 75–105.

Centre on Housing Rights and Evictions (2009). N2 Gateway Project: Housing rights violations as 'development' in South Africa.

Chance, K. (2008). Housing and evictions at the N2 Gateway Project in Delft. *a report for Abahlali baseMjondolo, available online at: [http://www. abahlali. org/]*.

Charlton, S. and C. Kihato (2006). Reaching the poor? An analysis of the influences on the evolution of South Africa's housing programme. *Democracy and Delivery: Urban policy in South Africa*, 252–282.

City of Cape Town (2005). Study on the social profile of residents of three selection informal settlements in Cape Town.

Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics*. Oxford, UK: Oxford University Press.

Department of Human Settlements (2009). National Housing Code.

Devoto, F., E. Duflo, P. Dupas, W. Pariente, and V. Pons (2011). Happiness on tap: piped water adoption in urban Morocco. Technical report, National Bureau of Economic Research.

Dinkelman, T. (2011). The Effects of Rural Electrification on Employment : New Evidence from South Africa valuable. *American Economic Review 101*(7), 3078–3108.

Duflo, E. (2012). Women's empowerment and economic development. *Journal of Economic Literature 50*(4), 1051–1079.

Duflo, E., S. Galiani, and A. M. Mobarak (2012). Improving access to urban services for the poor: open issues and a framework for a future research agenda. *J-PAL Urban Services Review Paper. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab. http://www. povertyactionlab. org/publication/improving-access-urban-services-poor*.

Duflo, E. and R. Pande (2007). Dams. *Quarterly Journal of Economics 122*(2), 601–646.

Duranton, G. and A. J. Venables (2018). *Place-based policies for development*. The World Bank.

Field, E. (2005). Property rights and investment in urban slums. *Journal of the European Economic Association 3*(2-3), 279–290.

Field, E. (2007). Entitled to work: Urban property rights and labor supply in Peru. *Quarterly Journal of Economics 122*(4), 1561–1602.

Franklin, S. (2019). The demand for government housing: Evidence from lotteries for 200,000 homes in ethiopia. Technical report, Working Paper.

Galiani, S., P. J. Gertler, R. Undurraga, R. Cooper, S. Martínez, and A. Ross (2017). Shelter from the storm: Upgrading housing infrastructure in latin american slums. *Journal of Urban Economics 98*, 187–213.

Gilbert, A. G. (2004). Helping the poor through housing subsidies: Lessons from Chile, Columbia and South Africa. *Habitat International 28*, 13–40.

Greenwood, J., A. Seshadri, and M. Yorukoglu (2005). Engines of liberation. *The Review of Economic Studies 72*(1), 109–133.

Gronau, R. (1977). Leisure, Home Production, and Work–the Theory of the Allocation of Time Revisited. *Journal of Political Economy 85*(6), 1099.

Harari, M., M. Wong, et al. (2018). Slum upgrading and long-run urban development: Evidence from indonesia. In *2018 Meeting Papers. Society for Economic Dynamics*.

Huchzermeyer, M. (2003). A legacy of control? The capital subsidy for housing, and informal settlement intervention in South Africa. *International Journal of Urban and Regional Research 27*(3), 591–612.

Huchzermeyer, M. (2006). Challenges facing people-driven development in the context of a strong, delivery-oriented state: Joe Slovo Village, Port Elizabeth. In *Urban Forum*, Volume 17, pp. 25–53. Springer.

Jacob, B. A. (2004). Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in chicago. *American Economic Review 94*(1), 233–258.

Jalan, J. and M. Ravallion (2003). Does piped water reduce diarrhea for children in rural India? *Journal of econometrics*.

Keare, D. H. and S. Parris (1982). Evaluation of Shelter Programs for the Urban Poor Evaluation of Shelter Programs for the Urban Poor Principal Findings.

Klonner, S. and P. Nolen (2010). Cell Phones and Rural Labor Markets: Evidence from South Africa. Proceedings of the german development economics conference, hannover 2010.

Kumar, T. (2019). The human capital effects of subsidized homeownership: Evidence from a natural experiment in mumbai.

Lam, D., J. Seekings, and M. Sparks (2006). The Cape Area Panel Study: Overview and Technical Documentation for Waves 1-2-3.

Marx, B., T. Stoker, and T. Suri (2013). The economics of slums in the developing world. *Journal of Economic Perspectives 27*(4), 187–210.

McKenzie, D. and Y. Seynabou Sakho (2010). Does it pay firms to register for taxes? The impact of formality on firm profitability. *Journal of Development Economics 91*(1), 15–24.

Michaels, G., D. Nigmatulina, F. Rauch, T. Regan, N. Baruah, and A. Dahlstrand-Rudin (2017). Planning ahead for better neighborhoods: Long run evidence from tanzania.

Muyeba, S. (2014). Effects of government housing subsidies in khayelitsha, cape town. *CSSR Working Paper No. 346*.

Olsen, E. O. and J. E. Zabel (2014). United States housing policy. In *HandBook of Regional & Urban Economics*.

Pharoah, R. (2012). Fire Risk in Informal Settlements in Cape Town, South Africa. *Disaster risk reduction: cases from urban Africa*, 105.

Picarelli, N. (2019). There is no free house. *Journal of Urban Economics 111*, 35–52.

Pitt, M. M., M. R. Rosenzweig, and N. Hassan (2006). Sharing the Burden of Disease: Gender, the Household Division of Labor and the Health Effects of Indoor Air Pollution in Bangladesh and India. In *Health San Francisco*, Volume 202.

Ravallion, M. and Q. Wodon (2000). Does child labour displace schooling? Evidence on behavioural responses to an enrollment subsidy. *The Economic Journal 110*(462), 158–175.

Rodriques, E., J. Gie, and C. Haskins (2006). Informal Dwelling Count for Cape Town (1993-2005).

Royston, L. (2002). Security of urban tenure in South Africa: Overview of policy and practice. *Holding their Ground. Secure Land Tenure for the Urban Poor in Developing Countries*, 165–181.

Schwebel, D. C., D. Swart, S. K. A. Hui, J. Simpson, and P. Hobeb (2009). Paraffin-related injury in low-income South African communities: Knowledge, practice and perceived risk. *Bulletin of the World Health Organization 87*(9), 700–706.

Seekings, J., T. Jooste, S. Muyeba, M. Coqui, and M. Russell (2010, May). The Social Consequences of 'Mixed' Neighbourhoods. *CSIR Working Paper*.

Tissington, K. (2011). A Resource Guide to Housing in South Africa 1994-2010: Legislation, Policy, Programs and Practice.

Tissington, K., N. Munshi, G. Mirugi-Mukundi, and E. Durojaye (2013). Jumping the Queue: Waiting lists and other myths: Perceptions and practice around housing demand and allocation in South Africa. Technical report, Socio-economic Rights Institute of South Africa.

Tomlinson, M. R. (1998). South Africa's New Housing Policy: An Assessment of the First Two Years, 1994-96. *International Journal of Urban and Regional Research 22*(1), 137–146.

Tshangana, M. (2009, April). Allocation of subsidized housing opportunities to households in the Western Cape by municipalities. International Conference, Exhibition and Housing Awards, Cape Town 2009.

UN Habitat (2003). The Challenge of Slums: Global report on human settlements. *Nairobi, Kenya, UN: Habitat*.

UN Habitat (2010). *State of the World's Cities 2010/11: Bridging the Urban Divide*. Earthscan.

Vorster, J. H. and J. E. Tolken (2008). Western Cape Occupancy Study. Technical report, Department of Sociology and Social Anthropology, Stellenbosch University.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Zwane, A. P. and M. Kremer (2007). What works in fighting diarrheal diseases in developing countries? A critical review. *World Bank Research Observer 22*(1), 1–24.

Online Appendix: Additional Material, Figures and Tables

A.1 A Model of Home Production, Work and Leisure

All told, the time burdens of living in informal dwellings are considerable. In the empirical analysis I seek to evaluate the total effect of receiving housing, which affects the way in which home production happens through many channels. In what follows, I present a simple model of how changes in housing quality could influence home production, which in turn predicts increases in labour hours due to the effect of formalized housing. I do not distinguish between the different channels through which housing could improve home production, which were outlined in the previous section.

The model I use is of the lineage of Becker (1965), since it specifies utility as a function of an unobserved home production input $H(T_h, b)$, which is produced through time at home T_h in combination with the physical housing infrastructure b. Importantly, I assume that home produced goods and services are *not* perfect substitutes with other forms of consumption, as opposed to many of the other models in this literature (Gronau, 1977, for example). This fits with the way I have conceived of home production in informal settings, where the basic needs provided for by housing cannot be taken for granted.

In this model, production at home cannot be traded on the market, it is used within the household. Household utility is a function of home production, consumption and leisure $U(H, C, L)$. Consumption is given by time spent working for wage labour T_w times the prevailing wage w. Leisure, time on home production, and time at work sum to one. With prices normalized to one, household utility is given by:

$$U(H(T_h, b), wT_w, 1 - T_h - T_w,)$$

If the household maximizes utility with respect to its allocation of time between labour, leisure and work at home, the first order conditions are simple:

$$\begin{aligned} U_H \cdot H_{T_h} &= U_L \\ U_C \cdot w &= U_L \end{aligned}$$

The optimizing household would thus choose its optimal time on work at home and labour, given by $T_h^\star = T_h^\star(b, w)$ and $T_w^\star = T_w^\star(b, w)$, respectively. I want to find $\frac{dT_w^\star}{db}$: the impact of upgrading the physical housing infrastructure on wage labour supplied. While one could speculate intuitively about the direction of impact from the FOC's, total differentiation with respect to b, gives a more complete picture, in the general case. With some manipulation this eventually yields:

$$\frac{dT_w^\star}{db}\left[U_{LL} - \frac{wU_{CC} + U_{LL}}{U_{LL}}\left(U_{LL} + U_{HH}H_T + H_{TT}U_H\right)\right] = -[U_{HH}H_bH_T + H_{Tb}U_H]$$

$$\frac{dT_w^\star}{db} = \frac{[U_{HH}H_bH_t + H_{Tb}U_H]}{wU_{CC} + \left(\frac{U_{CC}w + U_{LL}}{U_{LL}}\right)(U_{HH}H_T + H_{TT}U_H)}$$

Assuming diminishing marginal utility for all inputs into the utility function, and a diminishing marginal product of time at home, renders the denominator unambiguously negative.

Turning to the numerator, the first term is clearly negative due to the diminishing marginal returns on home production and the positive returns to housing quality from time spent at home. The sign of the second term hinges on whether or not the marginal utility of time in the home increases or decreases with an improvement in housing quality. If $H_{Tb} = \frac{\partial^2 H}{\partial T \partial b} \leq 0$ the numerator would be negative, and the response of hours in the labour market would be unambiguously positive.

The sign of the H_{Tb} reflects the extent to which the returns to time spent on activities in the home increase or decrease as the home technology improves. In an setting where home production leads to income through the production of goods sold on the market, one might expect a positive sign for H_{Tb}: b acts as production technology that allows households to increase output by working at home more.

I would argue that under my definition of home production, H_{Tb} is likely to be negative in this context. Improved housing is thought to be a labour saving technology, allowing households to reach a desired level of home quality. For instance, providing a better roof and walls would reduce the value of work done on maintaining the home, because nothing really needs to be done to make the structure more secure anymore.

In a sense, the empirical results of the paper provide a test of the sign of H_{Tb}, showing that poor housing quality necessitates increased time spent at home, which could be spent more productivity somewhere else.

A.2 Additional information on first stage estimates

This section covers some additional information related to the estimation of the predicted values for the probability of receiving housing.

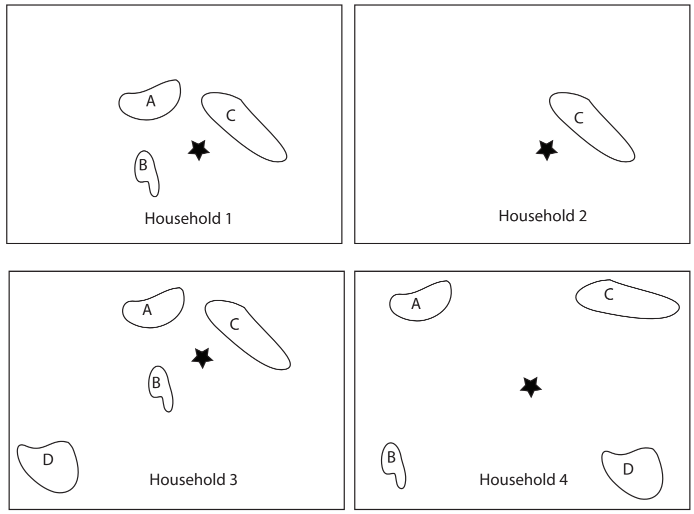
I do three things: I explain the rationale for the use of a non-linear predictor of getting housing in more detail than the main section, as opposed to other linear estimators, or estimators using a single index of housing proximity. Secondly, I explain the time-varying multi-nominal version of the estimator outlined, which explains how I use this framework to predict time varying treatment effects. Thirdly, I present Monte Carlo estimates that show the reliability of the method to recover the true parameters of the data generating process it describes.

A.2.1 Why a non-linear estimator

I estimate the probability of selection into treatment as the joint probability of being selected by a set of neighboring housing projects. Using only the closest housing project as an instrument would completely miss the effect of living in an area with a great number of projects, relative to a house who just has one, very small project nearby. Additionally, the marginal impact of additional housing projects should diminish as more are built: Consider the possible geographical scenarios depicted in Figure A.1 below. We want household 1 to be more likely to be treated than 2 of course, but not three times as likely. After all, we might believe the causal effect of a project as close as C to be a 50% chance of treatment for household 2 which would lead to considerable estimation issues for Househould 1 in a linear model.

Alternatively, summing distances to all projects and estimating a single coefficient would severely penalize household 3 to the benefit of household 1 in the diagram, as 3's extra project D would increase the sum of distances. However, placing a different coefficient on the distance from each project, would also be misleading- a new project's influence should be diminishing in the probability that a household as already been selected. For households like 3, project D

Figure A.1: Hypothetical housing proximity scenarios



is unlikely to influence is probability of treatment at the margin but for households like 4- D could make all the difference.

Trying to incorporate all of these sorts of concerns would involve a great deal of linear restrictions. Instead I estimate a non-linear model that more closely resembles the real world allocation process, and avoids the pitfalls of linear models.

I make two main assumptions in the specification of our function form: firstly, that a household's probability of selection into a individual housing project is a function of a linear combination of a number of proximity instruments, as well as household covariates. I also assume that each project allocates houses independently from the other projects, but that they all do so with the same catchment area; so that parameter on "distance from a project" is constant across households, projects and time.

A.2.2 Time specific predictions

The problem is complicated further when we consider that we have panel data, and need to estimate a probability of being treated for each time period $P(T_{it} = 1)$ in order to generate a time varying probability of treatment. This problem is analogous to finding the joint probability across programs, except that in this case we find the probability across time. Just like one cannot be treated by two projects at one time, one cannot be treated in more than one time period, and once you are treated you stay treated for all ensuing periods. Thus the probability of not being treated at all at time t is simply the product of the probability of not being treated by any projects built in each time period $A_t \in A$, up to and including t. This can be simply rewritten as the sum of of probabilities across all projects built in the past. In this way, estimation exploits my data on the timing of housing construction to more accurately predict when households were treated.

$$
\begin{aligned}
P(T_{it}=1) &= 1-\prod_{j=1}^{t}\prod_{a}^{A_j} P(T_{ia}=0) \\
&= 1-\prod_{a}^{A_t,A_{t-1},\ldots} P(T_{ia}=0) \\
&= 1-\prod_{a}^{A_t,A_{t-1},\ldots} \Lambda(-x_i\beta - dis_{ia}\rho)
\end{aligned}
\tag{10}
$$

In this functional form (10), the probability for a household at each time is an independent event. The probability of having the treatment must be monotonically increasing with time (for each household) as more projects are built. This is the first of two specifications I use to predict treatment.

An alternative method for estimating the probability of treatment, which takes into account the panel nature of the data, is to construct a single likelihood function for each household, which predicts *when* the household was treated. Each household is assigned a value for TD, which takes on 0 if never treated or $t = 1, 2, \ldots, Y$ if it received the treatment in period t. These are mutually exclusive outcomes. We can then estimate a model of **multinomial form**, in this case for a Y period model. At each time period we continue to use the form (5) in the calculations.

$$
\begin{aligned}
P(TD_i=0) &= \prod_{t=1}^{4}\prod_{a}^{A_t} P(T_{ia}=0)) \\
P(TD_i=1) &= 1-\prod_{a}^{A_1} P(T_{ia}=0) \\
P(TD_i=2) &= (1-\prod_{a}^{A_2} P(T_{ia}=0))\prod_{a}^{A_1} P(T_{ia}=0) \\
&\cdots \\
P(TD_i=Y) &= (1-\prod_{a}^{A_Y} P(T_{ia}=0))\prod_{t=1}^{Y-1}\prod_{a}^{A_t} P(T_{ia}=0)
\end{aligned}
\tag{11}
$$

Here the probability of treatment is now conditioned on not having been treated earlier. For instance, dummy variable $TD_i = 2$ (got treated in period 2) is the probability of not being selected in period 1 times the probability of being selected in period 2. These dummy variables and their predicted probabilities must, by definition, must sum to 1.

This dummy indicates the the household actually got the treatment in that period. This is different to the outcome of interest, which is the probability of *having the treatment* at a given time period. This can be backed out by simply summing the predicted dummy variables for all time periods up to the present. The probability of treatment at a given time period then simplifies to the same expression given by (10), although the estimation procedure differs. For instance for could calculate the probability of being treated at time 2 using this framework and get the expression give by (10) for $t = 2$:

$$
\begin{aligned}
P(T_{i2}=1) &= P(TD_i=1)+P(TD_i=2) \\
&= (1-\prod_a^{A_1} P(T_{ia}=0)) + (1-\prod_a^{A_2} P(T_{ia}=0)) \prod_a^{A_1} P(T_{ia}=0) \\
&= 1-\prod_a^{A_1} P(T_{ia}=0) \prod_a^{A_2} P(T_{ia}=0) \\
&= 1-\prod_t^{2}\prod_a^{A_t} P(T_{ia}=0)
\end{aligned}
$$

To summarize, I have two non-linear specifications for the probability of being treated at a given time, equation (10) and equation (11). In both models, probability of being treated in a certain time period depends on the projects built up until that point. While we would expect (11) to be the better estimator in the presence of unobserved individual heterogeneity, in the presence of measurement error, we may get less efficient results because it requires the precise time period in which the household was treated. The difference between these two types of estimators and their bias in the presence of unobservables, is explored in the Monte Carlo section below.

A.2.2.1 Marginal Effects

But first it is useful to have some marginal effects interpretation. This is slightly more complicated than a standard logit framework, but has an intuitive interpretation. We write down the probability of being treated at a particular point using the expression (5) and use the properties of the logistic function, to take the derivative with respect to a particular project b.[51]

$$
P(T_i=1) = 1-\prod_a^{A} \Lambda(-x_i\beta - dis_{ia}\rho) \tag{12}
$$

$$
= 1-\prod_a^{A} \frac{1}{1+exp(x_i\beta + dis_{ia}\rho)} \tag{13}
$$

$$
\begin{aligned}
\frac{\partial P(T_i=1)}{\partial dis_{ib}} &= -\prod_{a\neq b}^{A} \frac{1}{1+exp(x_i\beta+dis_{ia}\rho)} \cdot \frac{-exp(x_i\beta+dis_{ib}\rho)}{(1+exp(x_i\beta+dis_{ib}\rho))^2}\cdot\rho \\
&= \prod_a^{A} \frac{1}{1+exp(x_i\beta+dis_{ia}\rho)} \cdot \frac{exp(x_i\beta+dis_{ib}\rho)}{1+exp(x_i\beta+dis_{ib}\rho)}\cdot\rho \\
&= P(T_i=0)\cdot P(T_{ib}=1)\cdot\rho
\end{aligned}
$$

In this framework the marginal effects of distance to particular project depend on a household's current probability of being treated (negatively) and on the probability of being by the treated by the project in question (positively). This is consistent with the idea that a new

[51] Of course, taking a partial derivative invokes the ceteris paribus assumption. Strictly speaking, this is not plausible in my case. Up until now, I have been discussing a set of distances to projects for each household, these projects will be common to a number of households. So it is hard to imagine the distance from a household to a project changing without it effecting the distance for other households, which in turn would influence the probability of a household being treated

construction has a relatively bigger effect for a household with few existing projects nearby, and that the probability of being treated drops off faster the further away a particular project gets.

Using the results from the estimation of the first stage in Section 4.2.2 and the coefficients in Column (4): the coefficient on distance is 0.672, on distance squared it is 0.0079. Imagine a household close to two projects, with characteristics such that the household has a predicted probability of being selected of 10% for both of the projects. Then imagine that one project (b) was originally located 1km away but is relocated slightly further away the household. Then the probability of that household being treated would fall by over 4% for each kilometer that it moved:

$$\begin{aligned}\frac{\partial P(T_i = 1)}{\partial dis_{ib}} &= P(T_i = 0) \cdot P(T_{ib} = 1) \cdot (-0.672 + 0.0158 \times dis_{ib}) \\ &= (0.9 \times 0.9) \times 0.1 \times (-0.672 + 0.0158 \cdot 1) = \qquad -4.16\%\end{aligned}$$

A.2.2.2 Maximum Likelihood Estimation and Monte Carlo Simulations

In this section I estimate the parameters of the models specified in (10) and (11) using simulated data. Estimation of these models has to be performed using maximum likelihood estimation. I use the Stata **ml** code in order to maximize the log likelihood functions derived from the predicted probability of treatment given by each model, using the Newton-Raphson method.

To perform Monte Carlo tests, I simulate a dataset of $N = 1000$ observations with 3 time periods each. Then for each time household-time observation I generate 5 random project distances (to simulate the construction of houses nearby that household). Each household has a randomly generated household effect x_i that is constant across time and assumed to be unobserved. In addition, each time period has a random effect on the probability of treatment, common to everyone. Then, for each project at each time a latent variable is generated as a function of time effects, the household fixed effects and, of course, the distance from the household to the project. I use a linearly added logistic error term. If this latent variable is greater than zero we consider a household to be "treated" by that project. A household is treated at that time if it is treated by any *one* of the projects, and it remains treated for the ensuing periods.

$$\begin{aligned}y^{\star}_{iat} &= \alpha + \sigma_x x_i + \sigma_\lambda \lambda_t + \rho dis_{iat} + \epsilon_{it} \\ y_{iat} &= \mathbf{1}[y^{\star}_{iat} > 0] \\ y_{it} &= \mathbf{1}[\sum_{a}^{A_t, A_{t-1}\ldots} y_{iat} > 0]\end{aligned}$$

Where the household characteristics and distance variables are generated in the following way:

$$\begin{aligned}&x_i \sim N[0,1], \lambda_t \sim N[0,1] \\ &dis_{iat} \sim U[1,10] \\ &\epsilon_{it} = \frac{exp(\eta_{it})}{1 + exp(\eta_{it})}, \eta_{it} \sim U[0,1]\end{aligned}$$

Having generated simulated values, I recover the parameter of interest, which is ρ, using the models specified. I estimate three different specifications. The first (L_{nt}) uses the functional form (10) but without any attempts to control for time trends. The second (L_t) also uses (10), but controls for time by specifying time dummies λ_t in the latent $y^\star$ form. The third (MNL) is the estimation of (11). I then perform Monte Carlo simulations with 1000 repetitions, for each model, while varying the magnitude of variance of the unobserved effects. The results of these simulations, with different "true" values of ρ, are given in table A.1. The model

Table A.1: Results of Monte Carlo Simulations: Estimated value of ρ with different unobserved fixed and time effects

	$\rho = -1$			$\rho = -0.5$		
	L_{nt}	L_t	MNL	L_{nt}	L_t	MNL
$\sigma_x^2 = 0, \sigma_\lambda^2 = 0$	-1.013	-1.013	-1.019	-0.507	-0.507	-0.513
	(0.148)	(0.148)	(0.137)	(0.078)	(0.078)	(0.067)
$\sigma_x^2 = 1, \sigma_\lambda^2 = 0$	-0.914	-0.915	-0.950	-0.433	-0.433	-0.464
	(0.125)	(0.125)	(0.109)	(0.067)	(0.067)	(0.059)
$\sigma_x^2 = 0, \sigma_\lambda^2 = 1$	-0.997	-1.003	-1.007	-0.494	-0.500	-0.503
	(0.144)	(0.143)	(0.143)	(0.076)	(0.075)	(0.076)
$\sigma_x^2 = 3, \sigma_\lambda^2 = 0$	-0.742	-0.744	-0.821	-0.330	-0.332	-0.383
	(0.099)	(0.099)	(0.089)	(0.061)	(0.061)	(0.055)
$\sigma_x^2 = 0, \sigma_\lambda^2 = 3$	-0.836	-0.904	-0.928	-0.416	-0.481	-0.453
	(0.244)	(0.175)	(0.240)	(0.138)	(0.082)	(0.138)

(N=1000, t=3, R=1000)

performs very well without any fixed effects or time trends, as expected. The introduction of fixed effect biases the estimates towards zero. The bias can be quite considerable when these fixed effects are relatively large, as the example with $\sigma_x^2 = 3$ indicates. The effects are less severe with the introduction of unobserved time effects, but still biased towards zero. The MNL estimator performs better when there fixed effects. Importantly, the inclusion of time controls in the L model does a very good job of recovering the parameters correctly.

A.3 Additional Figures and Tables

Figure A.2: Map of Cape Town and its housing projects in 2009

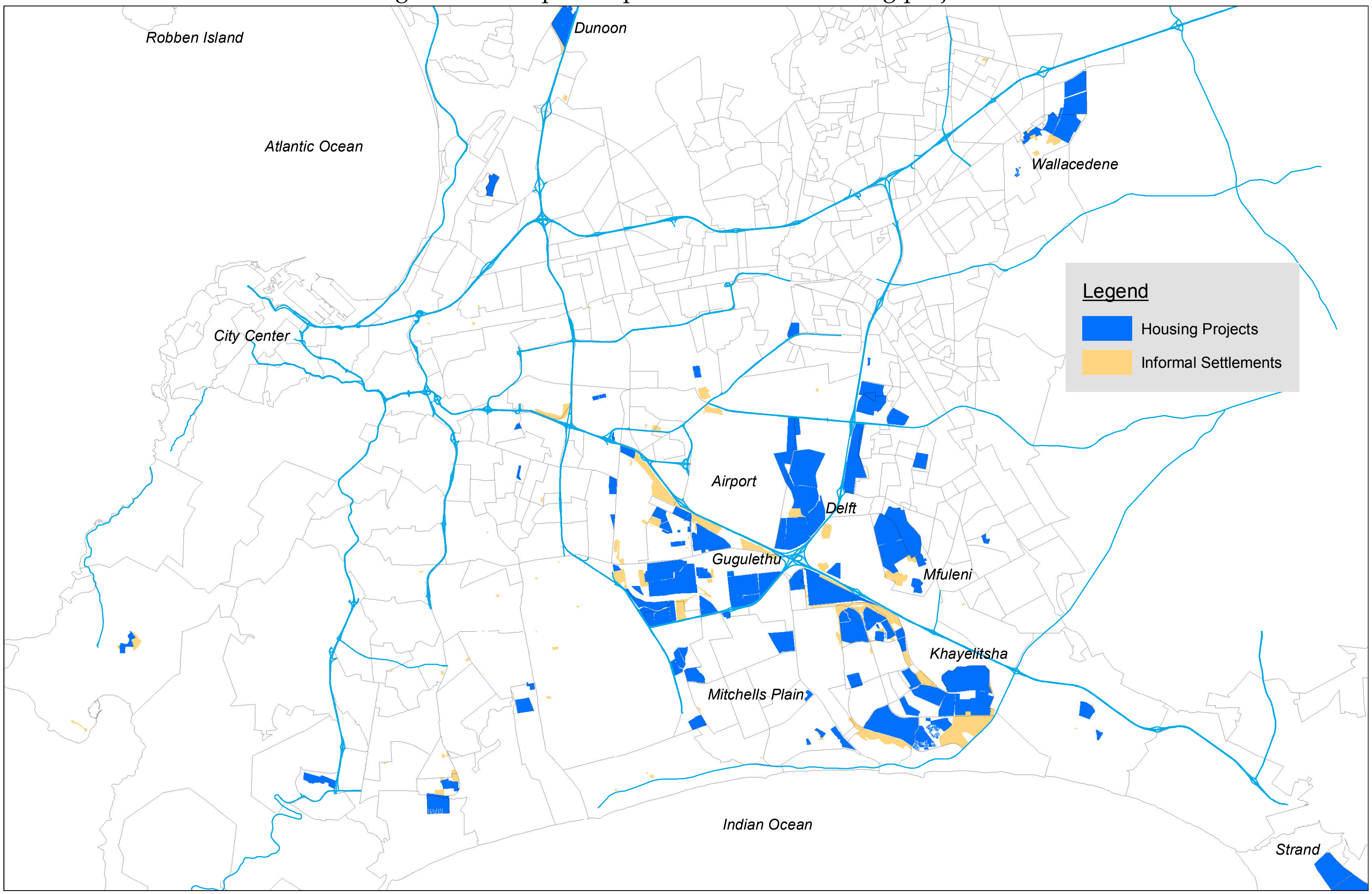


Figure A.3: Kernel density of predicted treatment for treated and untreated groups

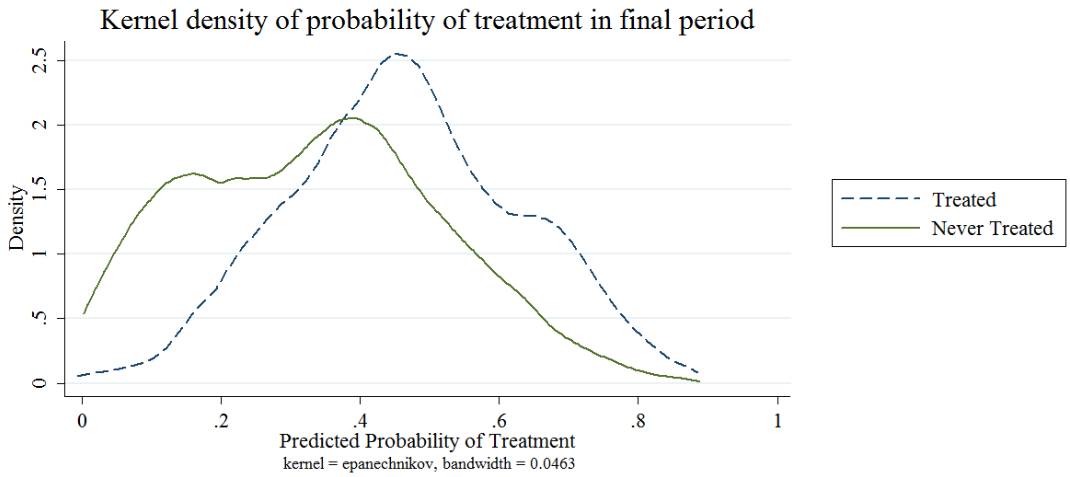


Figure A.4: Predicted probability of treatment in the trimmed sample

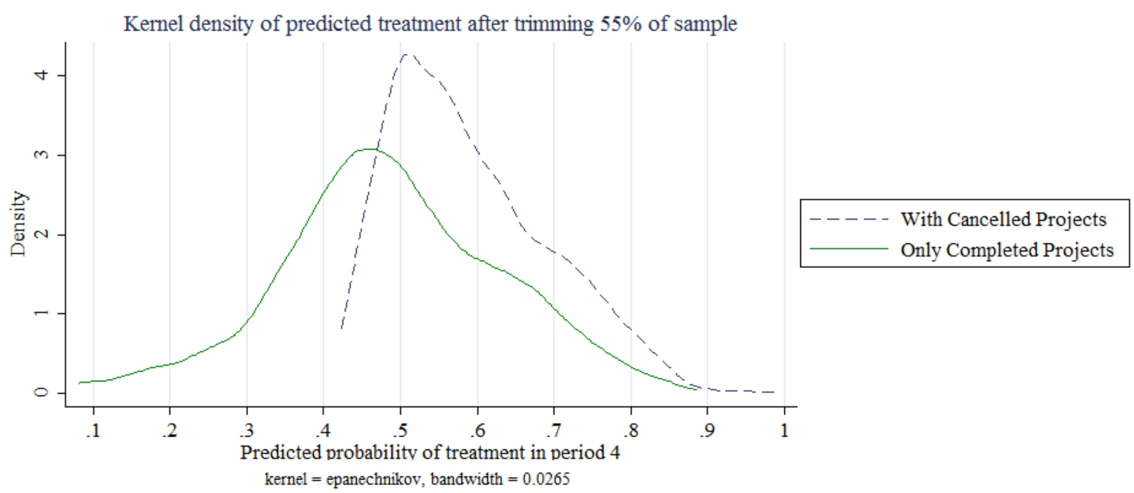


Table A.2: Most common occupations by men in women in study sample

	Percent of jobs			Average Earnings		
Occupation	Men	Women	All	Men	Women	All
Domestic and Related Helpers, Cleaners	7.83	48.42	30.94	1,434	1,238	1,259
Housekeeping and Restaurant Services Wo	6.21	9.93	8.33	1,437	1,299	1,343
Shop Salespersons and Demonstrators	7.02	7.85	7.49	1,563	1,414	1,469
Building Frame and Related Trades Worke	13.43	0.15	5.87	1,681	1,200	1,673
Protective Services Workers	9.97	2.39	5.65	2,017	2,051	2,025
Motor Vehicle Drivers and Related Worke	10.99	0.08	4.78	2,081	600	2,063
Cashiers, Tellers and Related Clerks	1.53	7.08	4.69	3,292	1,910	2,109
Building Finishers and Related Trades W	10.27	0.15	4.51	2,119	1,400	2,103
Market Gardeners and Crop Growers	4.98	0.85	2.63	1,359	962	1,278
Painters, Building Structure Cleaners a	5.49	0.31	2.54	1,701	1,000	1,670
Food Processing and Related Trades Work	2.24	2.62	2.45	1,308	1,015	1,135
Textile, Fur and Leather Products Machi	0.41	3.46	2.15	1,500	1,725	1,707
General Managers	1.12	2.31	1.8	1,930	1,023	1,290
Material-Recording and Transport Clerks	2.54	1	1.67	2,944	1,930	2,582
Metal Moulders, Welders, Sheet-Metal Wo	3.87	0	1.67	2,379	.	2,379
Personal Care and related workers	0.31	2.54	1.58	1,800	1,326	1,370
Messengers, Porters, Doorkeepers and Re	3.26	0.31	1.58	1,391	1,188	1,362
Numerical Clerks	0.81	1.77	1.36	2,683	2,985	2,907
Textile, Garment and Related Trades Wor	1.22	1.39	1.31	2,248	1,338	1,688
Machinery Mechanics and Fitters	2.44	0.23	1.18	2,385	1,867	2,314
Client Information Clerks	0.71	1.39	1.1	2,688	2,580	2,605
Business Professionals	1.22	0.85	1.01	2,400	2,350	2,377
Nursing and Midwifery Professionals	0.2	1.54	0.96	9,000	3,964	4,300
Other Personal Services Workers	0.71	1.15	0.96	1,192	1,268	1,244
Electrical and Electronic Equipment Mec	0.92	0.92	0.92	1,043	600	763
Street Vendors and Related Workers	0.31	1.31	0.88	390	2,223	1,917
Average				1,847	1,428	1,604

Table A.3: Household characteristics in first & last waves, by treatment

	Wave 1 (2002)			**Wave 4 (2009)**		
	Control	Treatment	Diff	Control	Treatment	Diff
From EC	0.712	0.827	0.114***			
			(0.0277)			
Backyard	0.122	0.0742	-0.0478*			
			(0.0198)			
Coloured	0.187	0.0797	-0.107***			
			(0.0228)			
Black	0.790	0.915	0.125***			
			(0.0238)			
Migrant	0.658	0.657	-0.00119			
			(0.0306)			
Shack	1	1	0	0.624	0.188	-0.436***
			(0)			(0.0328)
City distance	22.68	25.12	2.434***	22.51	25.19	2.681***
			(0.394)			(0.436)
Years Ed. head	11.04	11.38	0.341	12.22	12.46	0.238
			(0.214)			(0.239)
Num. Rooms	3.123	3.259	0.136	3.427	3.828	0.401**
			(0.0975)			(0.122)
HH Size	5.189	5.478	0.289	5.549	6.065	0.516*
			(0.149)			(0.221)
Female Head	0.488	0.563	0.0751*	0.533	0.597	0.0649
			(0.0321)			(0.0361)
Age Head	41.66	42.51	0.854	43.95	46.08	2.129*
			(0.758)			(0.947)
Young adult employed	0.112	0.0769	-0.0353	0.465	0.460	-0.00472
			(0.0193)			(0.0360)
% Females Employed	0.442	0.447	0.00439	0.367	0.403	0.0353
			(0.0289)			(0.0277)
Head Employed	0.691	0.632	-0.0596*	0.618	0.562	-0.0554
			(0.0302)			(0.0354)
Health Score	3.858	3.879	0.0214	3.790	4.022	0.233*
			(0.0865)			(0.0903)
Piped Water	0.130	0.107	-0.0233	0.348	0.534	0.185***
			(0.0211)			(0.0351)
Earnings pc	874.9	620.21	-254.75	330.95	400.01	69.06
			(254.74)			(46.177)
Log Income	7.436	7.218	-0.217***	8.194	8.272	0.0784
			(0.0606)			(0.0635)
Obs	713	364		626	344	

Table A.4: Example of first stage from 2SLS with single fitted instrument

Variable	Coefficient		(Std. Err.)
g-hat	0.885	**	(0.088)
femalehead	-0.070	**	(0.021)
hhsize	-0.009		(0.006)
sexratio	0.097	***	(0.051)
youngratio	0.008		(0.046)
femadultcount	0.002		(0.015)
time2	0.024		(0.020)
time3	0.024		(0.025)
time4	0.037		(0.034)
citydis	0.008	*	(0.004)
maxhhed	0.001		(0.003)
hhmaxage	0.002	**	(0.001)
hhgrants	-0.027	***	(0.016)
Intercept	-0.267	*	(0.122)
N	3711		
R^2	0.318		
$F_{(12,158)}$	28.104		

Notes: These are results from the first stage of the household fixed effects regressions used throughout this paper. These results are basic OLS regression of the dummy variable for having government housing on time varying household characteristics, as well as g-hat, the predicted probability of receiving housing from the maximum likelihood estimator of the probability of getting housing. Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.5: Effects of government housing on log per capita total household income using a balanced panel (households observed in all four waves of the data)

	(1) FE	(2) FE	(3) FE	(4) IV	(5) IV	(6) IV
house	0.230***	0.250***	0.248***	0.679**	0.701**	0.644**
	(0.0770)	(0.0751)	(0.0725)	(0.296)	(0.312)	(0.295)
Observations	2,252	2,252	2,248	2,252	2,252	2,248
R-squared	0.196	0.215	0.241	0.186	0.190	0.222
Households	563	563	563	563	563	563
Weak IV F				91.17	82.12	82.05

Notes: Standard errors clustered at the enumeration area level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. house=1 if household reported getting a subsidized house at any point in the past. Dependent variable is the log of total household income.

Table A.6: Heterogeneity in effects on individual female employment

	(1)	(2)	(3)	(4)	(5)
Baseline covariates	1) work 2) no work	1) part time 2) full time	1) over 30 2) under 30	1) Not single 2) Single	1) HH Head 2) Not head
house*cov=1	0.026	0.012	0.010	0.070	-0.002
	(0.043)	(0.070)	(0.039)	(0.048)	(0.051)
house*cov=2	0.007	0.032	0.018	-0.031	0.038
	(0.037)	(0.054)	(0.046)	(0.041)	(0.039)
Observations	3,744	1,181	3,744	3,744	3,744
R-squared	0.165	0.229	0.060	0.029	0.033
Households	1,328	363	1,328	1,328	1,328
Cov1=Cov2 (p)	0.737	0.827	0.898	0.112	0.538
Av Group	2.819	3.253	2.819	2.819	2.819
Mean (cov=1)	0.649	0.613	0.501	0.471	0.552
Mean (cov=2)	0.325	0.671	0.350	0.390	0.375

Notes: Standard errors clustered at the enumeration area level in parentheses. Sample is restricted to women of working age: ages 18 to 60. All regressions include individual fixed effects. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Each Column shows heterogeneous treatment effects by a particular coviarate. Column subtitles specify first a covariate category 1 and then category 2 below that. Results below are then divided by that category. For example, in Column (1) I show heterogeneous treatment effects by whether women *i* was working at baseline. Category 1 looks at women who worked. Category 2 looks at women who did not work at baseline. Therefore results row 1 shows the treatment effects for women who worked at baseline, row 2 shows results for women who did not work at baseline. Similarly, mean outcomes are shown for women who worked "Mean (cov=1) " and who did not work "Mean (cov=2)", respectively.
Column 2 restricts the sample to only women who were working at baseline, and the looks at differential treatment effects according to whether they had part time or full time work.
The row labelled "Cov1=Cov2 (p) " shows the p-values from t-tests of equivalence of the category 1 and category 2 coefficients.

Table A.7: IV estimations of impacts on household income with trimming using cancelled projects

	(1)	(2)	(3)	(4)
Sample Trimmed %	**20**	**40**	**60**	**80**
Instrumental Variables Impacts on Log Total Income				
house	0.886**	1.313**	1.409**	0.639
	(0.422)	(0.571)	(0.659)	(0.544)
Observations	2,880	2,224	1,510	745
Households	829	623	416	206
R-squared	0.209	0.123	0.108	0.420
Average Observations per household	3.474	3.570	3.630	3.617
Weak IV F-statistic	53.35	37.73	14.92	15.83

Notes: Standard errors clustered at the enumeration area level in parentheses. All IV regressions use a non-linear predicted probability of treatment. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Columns (1) - (4) replicate the main estimates in the paper but with the sample restricted to households with increasingly higher *ex ante* probability of treatment according to proximity to *planned* housing projects. Specifically, I generate a predicted probability of treatment based on proximity to completed and cancelled housing projects using my model of housing access based on completed projects only. This predicted probability is a kind of placebo treatment, had the cancelled projects actually gone ahead. Then, in Column 1 I drop the 20% of the sample with the lowest placebo probability of treatment. This, in effect, drops households in areas where no housing projects were planned. In Column 2 I drop the 40% lowest placebo probability of treatment, and so on.

Table A.8: Average household characteristics by major townships

	Gugulethu N=222	Khayelitsha N=608	Other N=247
Closest Project Distance	0.413	0.144	2.560
Distance to the City	18.57	29.75	21.32
Treated	30%	41%	19.4%
Coloured	7.7%	3.6%	48.8%
Log Income	7.245	7.303	7.598

Table A.9: Robustness to including cluster-specific (city-block) time trends

	(1)	(2)	(3)	(4)
	FE	FE	FE	FE
	Log hh income	HH earnings	Female earnings	Male Earnings
house	0.127*	506.0***	225.7*	228.6
	(0.0648)	(149.8)	(127.8)	(161.2)
Observations	3,577	3,699	3,438	2,902
R-squared	0.322	0.165	0.244	0.186
Households	1,064	1,064	1,045	985

Notes: Standard errors clustered at the enumeration area level in parentheses. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. *** p<0.01, ** p<0.05, * p<0.1. All regressions include interactions between time and cluster-fixed effects. There are 159 clusters in the final sample, in total.

Table A.10: Robustness to including area-specific (12 townships) time trends

	(1)	(2)	(3)	(4)
	FE	FE	FE	FE
	Log hh income	HH earnings	Female earnings	Male Earnings
house	0.160***	482.8***	276.1**	209.9*
	(0.0592)	(168.3)	(127.2)	(125.7)
Observations	3,577	3,699	3,438	2,902
R-squared	0.226	0.071	0.030	0.047
Households	1,064	1,064	1,045	985

Notes: Standard errors clustered at the enumeration area level in parentheses. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. *** p<0.01, ** p<0.05, * p<0.1. All regressions include interactions between time and cluster-fixed effects. There are 159 clusters in the final sample, in total.

Table A.11: Robustness to clustering standard errors at the electoral ward level

	(1)	(2)	(3)	(4)
	FE	FE	FE	FE
	Log hh income	HH earnings	Female earnings	Male Earnings
house	0.169***	465.8***	264.6*	191.7
	(0.0575)	(174.3)	(141.5)	(115.6)
Observations	3,577	3,699	3,438	2,902
R-squared	0.222	0.063	0.023	0.042
Households	1,064	1,064	1,045	985

Notes: Standard errors clustered at the enumeration area level in parentheses. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. *** p<0.01, ** p<0.05, * p<0.1. All regressions include interactions between time and cluster-fixed effects. There are 159 clusters in the final sample, in total.

Table A.12: Main FE results only among individuals in clusters where many households were treated

	(1)	(2)	(3)	(4)	(5)	(6)
	FE	FE	FE	FE	FE	FE
	lginc	lginc	lginc	ln_hhtotsal	ln_hhtotsal	ln_hhtotsal
house	0.226***	0.150**	0.144**	0.193***	0.181**	0.176**
	(0.0609)	(0.0588)	(0.0582)	(0.0656)	(0.0722)	(0.0719)
femalehead		-0.297***	-0.239***		-0.221***	-0.203***
		(0.0566)	(0.0596)		(0.0730)	(0.0765)
hhsize		0.133***	0.143***		0.0441***	0.0418***
		(0.0102)	(0.0110)		(0.0124)	(0.0136)
citydis		-0.0118	-0.00929		-0.0216*	-0.0197
		(0.0110)	(0.0108)		(0.0130)	(0.0130)
yamom			-0.0145			0.0753
			(0.0682)			(0.0900)
maxhhed			0.0315***			0.0396***
			(0.00718)			(0.00911)
sexratio			-0.125			-0.184
			(0.142)			(0.184)
youngratio			-0.623***			-0.273*
			(0.112)			(0.143)
Observations	2,388	2,397	2,393	1,895	1,895	1,892
R-squared	0.216	0.301	0.325	0.129	0.147	0.165
Number of hhs	747	758	758	718	718	718

Notes: Clustered standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1. house=1 if household reported getting a subsidized house at any point in the past. Sample restricted to households in EAs in which more than 20% of hhs received housing by 2009.

Table A.13: Replication of the key results without households treated before 2005

	(1) total income	(2) total salary	(3) female salaries	(4) male salaries	(5) female % employed	(6) male % employed
Panel A: Fixed Effects Regressions (Multiple Outcomes)						
house	0.268***	0.321***	0.341***	0.202*	0.0718*	0.0335
	(0.101)	(0.103)	(0.125)	(0.120)	(0.0379)	(0.0627)
Observations	1,900	1,467	966	918	1,844	1,407
R-squared	0.256	0.127	0.172	0.114	0.036	0.045
Number of hhs	498	489	415	385	494	454
Av Group Size	3.815	3	2.328	2.384	3.733	3.099
Panel A: Instrumental Variables Regressions (Multiple Outcomes)						
house	1.401**	1.982**	1.699	-0.154	0.472*	0.298
	(0.698)	(1.004)	(1.127)	(0.978)	(0.260)	(0.310)
Observations	1,900	1,429	858	828	1,835	1,349
R-squared	0.158	-0.120	0.016	0.097	-0.046	0.014
Number of hhs	498	451	307	295	484	396
Av Group Size	3.815	3.169	2.795	2.807	3.791	3.407
Weak IV F	27.57	19.95	13.75	6.726	31.46	23.61

Notes: Clustered standard errors in parentheses.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. house=1 if household reported getting a subsidized house at any point in the past. All regressions include controls for time-varying household characteristics. Dependent variable in Columns (1)-(4) are the log of total income, salaries of household members, salaries of female household members and salaries of male household members, respectively. Dependent variable "% employed" gives the proportion of female and male household members currently employed, in columns (5) and (6) respectively.

Table A.14: Replication of impacts on Log Income within communities and sub-samples

	(1) FE	(2) IV	(3) FE	(4) IV	(5) FE	(6) IV
	Khayelitsha		Gugulethu		Other	
house	0.262***	1.243*	-0.0150	0.247	0.191***	0.564*
	(0.0844)	(0.733)	(0.0920)	(0.595)	(0.0604)	(0.314)
Obs	1,471	1,470	730	726	3,582	3,561
R^2	0.276	0.168	0.364	0.357	0.290	0.275
	Out of Project		Poor		Black	
house	0.268**	0.988*	0.116*	0.839	0.178***	0.914**
	(0.102)	(0.574)	(0.0696)	(0.562)	(0.0641)	(0.432)
Obs	1,389	1,377	1,733	1,731	3,015	2,998
R^2	0.328	0.294	0.504	0.451	0.264	0.203
HH Chars	Yes	Yes	Yes	Yes	Yes	Yes
Weak IV F		19.61		13.07		95.24

Notes: Clustered standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1. house=1 if household reported getting a subsidized house at any point in the past. All regressions include controls for time-varying household characteristics. Dependent variable is log of total household income.

Table A.15: Impacts of government housing on household membership and composition.

	(1) FE Male adults	(2) FE Female Adults	(3) FE All children	(4) FE Under 3s	(5) FE Gender ratio	(6) FE Adult-child ratio	(7) FE Max yrs ed
house	0.116	0.086	0.085	0.114**	0.003	0.004	0.138
	(0.074)	(0.076)	(0.097)	(0.058)	(0.011)	(0.013)	(0.216)
Observations	3,492	3,492	3,492	3,526	3,492	3,492	3,490
R-squared	0.066	0.091	0.072	0.041	0.003	0.142	0.032
Households	955	955	955	955	955	955	955
Mean dep var	1.342	1.723	2.276	0.429	0.541	0.351	11.60

Notes: Clustered standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1. house=1 if household reported getting a subsidized house at any point in the past. Dependent var is the log of the sum of all monthly salaries earned by members of the household. All regressions include controls for time varying household characteristics and household fixed effects.

Table A.16: Impact of government housing on living conditions and asset ownership in the home (fixed-effects estimates)

	(1)	(2)	(3)	(4)	(5)
	FE	FE	FE	FE	FE
Panel A: Impact on Household Access to Services					
	Shack	Electricity	Toilet	Piped	Piped In
house	-0.579***	0.000640	0.00995	0.102***	0.242***
	(0.0206)	(0.0208)	(0.0210)	(0.0206)	(0.0219)
Observations	3,750	3,789	3,789	3,789	3,789
R-squared	0.480	0.041	0.077	0.028	0.157
Households	1,095	1,095	1,095	1,095	1,095
Panel B: Impact on Ownershipo of Household Appliances and Fuel Use					
	Stove	Fridge	Microwave	Paraffin	Fire
house	0.0445*	0.0908***	0.0575**	-0.0439*	-0.0171
	(0.0267)	(0.0259)	(0.0250)	(0.0237)	(0.0193)
Observations	3,749	3,749	3,748	2,942	2,949
R-squared	0.149	0.064	0.158	0.002	0.036
Number of hhs	1,095	1,095	1,095	1,095	1,095

Notes: Clustered standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1. house=1 if household reported getting a subsidized house at any point in the past. All regressions include controls for time-varying household characteristics. Dependent variables are dummy variables if the household has: Panel A, Col (1) an informal dwelling, Col (2) access to Electricity, Col (3) access to a flushing toiler, Col (4) access to piped water inside the house or nearby, Col (5) access to piped water in the home. Panel B, Col (1) a stove, col (2) a refrigerator, Col (3) a microwave , Col (4) used a paraffin stove for cooking or heating, Col (5) experienced a fire in the home.

Table A.17: Effects of government housing on feelings of being unsafe at home at night

	(1)	(2)	(3)	(4)
	FE	FE	IV	IV
	unsafe	unsafe	unsafe	unsafe
house	-0.325***	-0.134**	-1.711**	-1.169*
	(0.0548)	(0.0643)	(0.818)	(0.655)
HH controls	No	Yes	No	Yes
Observations	1,408	1,376	1,352	1,344
R^2	0.039	0.103	-0.631	-0.210
Weak IV F			8.723	11.47

Notes: Clustered standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1. house=1 if household reported getting a subsidized house at any point in the past. All regressions include controls for time-varying household characteristics. Dependent variable is a dummy variable = 1 if the household reports feeling unsafe at night in the home.

Table A.18: Heterogeneity in impacts on log household income by baseline characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Covariate	Children	School child	Child grant	HH size	Fem head	Fem employ	Distance	House	Assets
Cov =1	3+	1+	Yes	5+	Yes	Yes	>25km	High	High
Cov =2	<3	none	No	<5	No	No	<25km	Low	Low
house*cov=1	0.228***	0.216***	0.196**	0.150**	0.109	0.212***	0.134*	0.143	0.163**
	(0.075)	(0.072)	(0.097)	(0.074)	(0.080)	(0.067)	(0.075)	(0.087)	(0.080)
house*cov=2	0.028	0.038	0.140*	0.144	0.181**	0.040	0.185*	0.171**	0.140
	(0.091)	(0.093)	(0.073)	(0.103)	(0.086)	(0.101)	(0.103)	(0.080)	(0.085)
Observations	3,604	3,604	3,604	3,604	3,604	3,604	3,604	3,604	3,604
R-squared	0.207	0.208	0.203	0.206	0.208	0.207	0.201	0.201	0.202
Households	1,075	1,075	1,075	1,075	1,075	1,075	1,075	1,075	1,075
Cov1=Cov2 (p)	0.0815	0.116	0.640	0.962	0.518	0.131	0.685	0.810	0.838
Mean (cov=1)	6.224	6.263	6.220	6.272	6.315	6.462	6.341	6.379	6.511
Mean (cov=2)	6.522	6.492	6.417	6.511	6.417	6.198	6.403	6.356	6.247

Notes: Standard errors clustered at the enumeration area level in parentheses. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. *** p<0.01, ** p<0.05, * p<0.1 Each Column shows heterogeneous treatment effects by a particular coviarate. Column subtitles specify first a covariate category 1 and then category 2 below that. Results below are then divided by that category. For example, in Column (1) I show heterogeneous treatment effects by whether the household had three or more childern in the household at baseline. Category 1 looks at households with three or more children. Category 2 looks at households with two or fewer households. Therefore results row 1 ("house* cov=1") shows the treatment effects for households with three or more children. Similarly, mean outcomes are shown (in Column 1) for households with many children "Mean (cov=1) " and with fewer children"Mean (cov=2)", respectively.

Table A.19: Heterogeneity in impacts on log household income by Baseline Assets

	(1)	(2)	(3)	(4)	(5)
Covariate	Index	Washing Machine	Stove	Fridge	Microwave
Cov =1	High	Yes	Yes	Yes	Yes
Cov =2	Low	No	No	No	No
house*cov=1	0.163**	-0.041	0.182**	0.163**	0.013
	(0.080)	(0.148)	(0.074)	(0.075)	(0.155)
house*cov=2	0.140	0.168**	0.129	0.148	0.162**
	(0.085)	(0.067)	(0.085)	(0.095)	(0.066)
Observations	3,604	3,604	3,604	3,604	3,604
R-squared	0.202	0.202	0.201	0.202	0.204
Households	1,075	1,075	1,075	1,075	1,075
Cov1=Cov2 (p)	0.838	0.215	0.601	0.904	0.384
Mean (cov=1)	6.511	6.792	6.466	6.466	6.731
Mean (cov=2)	6.247	6.308	6.248	6.241	6.316

Notes: Standard errors clustered at the enumeration area level in parentheses. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. *** p<0.01, ** p<0.05, * p<0.1 Each Column shows heterogeneous treatment effects by a particular coviarate. Column subtitles specify first a covariate category 1 and then category 2 below that. Results below are then divided by that category.

Table A.20: Heterogeneity in impacts on log household income by baseline housing conditions

	(1)	(2)	(3)	(4)	(5)
Covariate	Index	Tapped water	Electricity	Big house	Improved walls
Cov =1	High	Yes	Yes	Yes	Yes
Cov =2	Low	No	No	No	No
subhet1	0.143	0.198***	0.180***	0.109	0.143*
	(0.087)	(0.070)	(0.065)	(0.081)	(0.074)
subhet2	0.171**	0.016	0.013	0.193**	0.175*
	(0.080)	(0.133)	(0.192)	(0.076)	(0.096)
Observations	3,604	3,604	3,604	3,604	3,604
R-squared	0.201	0.203	0.202	0.201	0.202
Households	1,075	1,075	1,075	1,075	1,075
Cov1=Cov2 (p)	0.810	0.226	0.418	0.401	0.797
Mean (cov=1)	6.379	6.388	6.384	6.360	6.397
Mean (cov=2)	6.356	6.308	6.291	6.367	6.335

Notes: Standard errors clustered at the enumeration area level in parentheses. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. *** p<0.01, ** p<0.05, * p<0.1 Each Column shows heterogeneous treatment effects by a particular coviarate. Column subtitles specify first a covariate category 1 and then category 2 below that. Results below are then divided by that category.

Table A.21: Impact of treatment on household grants and remittances

	(1)	(2)	(3)	(4)
	FE	FE	FE	FE
	hhgrants	hhgrants	totalsend	totalrec
house	-0.00495	-0.0245	-484.3	-80.95
	(0.0323)	(0.0303)	(629.7)	(197.3)
Observations	3,731	3,723	1,854	1,854
R-squared	0.009	0.140	0.006	0.029
Number of hhs	1,077	1,077	1,062	1,062
HH Chars	No	Yes	Yes	Yes
Av Group	3.464	3.457	1.746	1.746

Notes: Clustered standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1. house=1 if household reported getting a subsidized house at any point in the past. All regressions include controls for time-varying household characteristics. Totalsend and totalrec refer to the total amount of remittances sent and received. hhgrants is dummy for whether households received any one of the household grants such as disability benefits, or the childcare grant.

Table A.22: Estimates of longer-run treatment effects using only houseohlds that received housing before wave 2

	(1)	(2)	(3)	(4)
	FE	FE	FE	FE
	Log hh income	HH earnings	Female earnings	Male Earnings
house	0.229**	512.8	316.5	180.3
	(0.111)	(454.2)	(329.2)	(415.1)
Observations	1,229	1,294	1,203	1,040
R-squared	0.335	0.060	0.016	0.050
Households	650	652	642	601

Notes: Standard errors clustered at the enumeration area level in parentheses. house=1 if household reported getting a subsidized house at any point in the past. Dependent Variable is the log of total household income. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. All regressions include interactions between time and cluster-fixed effects. There are 159 clusters in the final sample, in total.