Neighbourhoods and local public goods: evidence from a housing lottery

PRE-ANALYSIS PLAN

Simon Franklin*

July 2018

1 Motivation

A common policy response to the problem of a lack of formal housing in African cities is for the state to build large housing estates on the outskirts of cities. While this may alleviate housing shortages, and reduce the proportion of households living in slums, it is not clear that these will new housing projects will be good places to live. Indeed, government built housing doesn't have a good reputation. Public housing projects in Europe and United States have often been described as "failed estates", due to associations with urban decay. In particular, public housing projects are often associated with a lack of maintenance and provision of local public goods. The problem may be even more acute in developing countries settings, where the states' capacity to provide this public goods may be weaker, and the role of maintenance and funding of neighbourhoods may fall to local residents. Yet relatively little is know about the reasons for *why* communities in state-built housing projects have either succeeded or failed in terms of public goods provision.

On the one hand, a literature in sociology places particular emphasis on the design of public housing a major determinant of neighbourhood outcomes. Commentators such as Alice Coleman and Oscar Newman emphasise the role of the design of public housing on neighbourhood outcomes, though from different perspective (see Coleman (1985) and Newman (1972)). Second, others fault the design of tenure arrangements in public housing. A lack of property rights may blunt the incentives of residents to properly invest in the maintenance and management of public spaces, leaving maintenance up to local authorities who may not have the resources or knowledge to keep properties well maintained. In the policy that I study, government built housing is owned by residents (who buy the units from the state). Private

*London School of Economics (email: franklin.simon@gmail.com)

ownership of housing has the potential to improve household investments, but these advantages may be blunted if owners often rent their units out. Third, concentrations of poverty are seen as causes of dysfunctional neighbourhoods. According to this view, housing policy should place more emphasis on creating mixed-income neighbourhoods.[1] Fourth, ethnic segregation and diversity could play a role in the provision of public goods. While segregation may be costly at the level of a city or neighbourhood, the implications of segregation and diversity within smaller geographic areas is not clear, and may differ across groups and contexts (Cutler et al., 2008; Cutler and Glaeser, 1997). Recent work in economics suggests that residents of housing estates with high levels of ethnic diversity can find it hard to coordinate to overcome collective action problems related to maintaining and improving public spaces and resolving disputes and violence (Algan et al., 2016).

Across developing countries, large housing estates are increasingly be built as a response to the challenge of slums and as a response to a lack of supply of formal housing. What kind of neighbourhoods will these estates turn out to be? The discussion above identifies four key factors that could play a role in determining neighbourhood outcomes: 1) building design, 2) poverty rates, 3) ethnic diversity, and 4) owner-occupation rates. To the extend that housing policy can explicitly influence each of these four things, how can housing policies be designed to improve the provision of public goods in these new neighbourhoods?

2 Study design

The Ethiopian governments housing policy presents a unique opportunity to study these questions. Housing units are delivered in five-storey buildings (referred to as "blocks" throughout this document) and awarded to applicant households via lottery. The lottery determines not only which households win apartments, but also the location of apartment that winning households are assigned, down to the door number. This helps to overcome the key identification challenge of endogeneous sorting of households into housing blocks and across neighbourhoods.

The random allocation of households to housing units generates considerable variation in the diversity, average income, and inequality within the housing blocks. I observe, on average, 20 exogenously assigned owner-households per apartment block. This natural variation allows me to study the effects of neighbourhood composition on neighbourhood and public goods outcomes. In addition, different housing block exhibit great variation in the number of owner-occupiers who moved into the housing. This variation is plausibly exogenous to the characteristics of the other owner-occupants of the block, as I find that households very rarely have the opportunity to meet their prospective neighbours before moving in.

In addition, random assignment to housing blocks generates neighbourhood compositions

[1] Recent evidence from the United States suggests that growing up in a high poverty neighbourhoods has a long term negative effect on individual outcomes Chetty et al. (2016).

that are orthogonal to the design and location characteristics of housing blocks (buildings). While housing blocks follow a similar basic design, there is variation in location, building orientation, and the availability of public spaces.

3 Housing Lottery and Data Collection

The Government of Ethiopia assigns new housing units to a pool of applicants via a lottery. Condition on winning the housing lottery, households are assigned to new housing blocks (buildings) at random. The average building is resident to twenty-five households, although this varies across and within different housing projects. The lottery is performed in rounds, annually. I study a lottery that took place in 2015, for which I had access to the list of households registered for the lottery, and the list of over 30,000 households who won the lottery out of about 100,000 households that were on waiting lists, and therefore entered into the lottery. In order to be eligible for the lottery, households need to register. It is thought that as many as 900,000 households have registered for the housing scheme at some point to date. The condition for registration is that the household not already own a property and is a registered resident of the city. Lottery winners are offered the chance to buy a housing mortgage, with a 20% down payment, with the mortgage to be paid off over the next 15-20 years at an interest rate of 9.5%. Almost all households that win the lottery buy the unit. I have verified that the lottery is indeed random, and will provide balance tests to show this.

Household and block listing:
I collected data on all households who own homes in 233 housing blocks (individual buildings), across 6 housing sites, at various times during 2016, in the few months after each housing block was made available to occupy to these owners. First I collected data on all of the owners (regardless of whether these owners lived in the housing blocks or not). Second I collected data on all of the actual occupiers of the housing blocks, for the cases where the unit was being rented out. For all of these households I collect basic household demographic and income data: wealth, employment, migration-status, ethnicity, religion, education, household composition, and some attitudes. In all, I collected data on over 8,000 households, both owners and renters.

Then I collected detailed data on the housing blocks in which households live through direct observation. A team of trained enumerators collected data on the design, layout, and physical condition of each housing block, using a standardized survey instrument. This instrument relies only on the judgment of the enumerators to record key features of each housing block and to score the quality of the housing blocks along key dimensions. I use GPS coordinates of each housing block to measure the distance of each block to roads and other key areas of the city.

Endline household survey:
I then randomly sampled households owning and living in their housing blocks (I sampled

roughly 30% of owner-occupiers living in each housing block), to be interviewed with a detailed survey instrument between January and April 2018. This questionnaire includes detailed modules on residents' perceptions of their neighbourhood, public goods, maintenance, and cooperation, social networks and conflict among neighbours. In all, I sample 900 households across 233 blocks.

Endline building quality survey:
Finally, I collected data on the conditions in and around the housing blocks, using a similar standardized survey instrument to the one used at the time of the block listing. This gives me two measures of housing block conditions, one right after households moved in, and one two years later. For the endline survey I expanded the set of a questions to include the measure of various other public goods investments households are known to make in the new housing blocks. The data from this endline survey will be used to construct my key indicators of neighbourhood quality and public goods investments.

4 Empirical strategy

This project is primarily concerned with processes that happen at the neighbourhood level, in this case taken to be a building block.

4.1 Identification from random assignment of owners

In the paper I will report descriptive statistics on the neighbourhood-level variation in owner and resident composition measures. Because it is the variation in owners' assignment that is random, I will use this to variation to explain block-level outcomes. In other words, I will report ITT estimates of the effect of block owner composition on block outcomes. The composition of block owners could affect block outcomes through distinct channels: 1) through the interactions (conflict, cooperation, management and socializing) of residents living in the housing block. This would be affected by the random assignment of owners if they a) those owners move in to the units they own, or b) they rent out the units to other individuals who have similar characteristics to them. 2) Through the interactions of non-resident owners who may meet and discuss management of the housing block even if they are not residents of the blocks themselves.

Based on interviews with housing winners who report relatively little interaction with other owners who do not live in the blocks, most of the effect is likely to come through the former channel: through the interactions of residents themselves. However, the effect of interactions through ownership could be under-powered if the relationship between the randomly assigned composition of owners is not correlated with the composition of actual residents (if the share of owners who move in is very low and their renters' characteristics are not correlated with theirs) or completely invalid, if the relationship between owner composition and renter composition is non-monotonic. For example, imagine that very poor households are more likely to rent out their units, and when they do they rent out the units to richer households on average than the average wealth of their randomly assigned peers. This could generate a pattern where housing blocks with particularly low income owners actually experience compositions of residents with higher incomes. In order to check this, I will report the correlations of owner and resident neighbourhood composition measures, using data on owners and renters collected at the time that households move in.

I will report tests for random assignment, by regressing the immutable (baseline) characteristics of households on the average characteristics of their peers at baseline. Given that the assignment of owners to units was random, it should be the case that owner characteristics are uncorrelated with the characteristics of their neighbours.

4.2 Block level regressions

I study block-measures of building and neighbourhood quality Y_{bst}, for block b and site s, regressed on a set of block level controls, including the corresponding baseline measure of Y measured at $t-1$ (when households had just moved in), a set of block level fundamentals $\mathtt{L}_b$ (such as block location, orientation and design) and finally block level composition $\widetilde{X}_{bist-1}$, calculated from the characteristics of the owners i assigned to live in that block.

$$Y_{bst} = \alpha_0 + \alpha_1 Y_{bst,t-1} + \beta_1 \cdot \mathtt{L}_{bs} + \beta_2 \cdot \widetilde{X}_{bis,t-1} + \gamma_s + \mu_b. \tag{1}$$

4.3 Household level regressions

I study household level measures of neighbourhood quality and community for household i, in building b, in housing site s, at my endline survey ($t = 1$). I estimate the effects of building fundamentals $\mathtt{L}_{bs}$ and building composition $\widetilde{X}_{bt-1}$, controlling for household level covariates measured at baseline $X_{ib,t-1}$.

$$Y_{bist} = \alpha_0 + \beta_1 \cdot \mathtt{L}_{bs} + \beta_2 \cdot \widetilde{X}_{b,t-1} + \delta \cdot X_{ib,t-1} + \gamma_s + \mu_{bi}. \tag{2}$$

I cluster standard errors at the block level. In the household level regressions I will weight each household observation by the inverse of the number of households surveyed in that block, in order to recover estimates that are representative of the average block effect.[2] I general, I do not include controls for baseline measures of Y_{bist} because the baseline did not measure all of the key components of the outcome of interest at the household level that were measured in the endline survey.

Baseline household covariates, $X_{ib,t-1}$ are used to control for possibility that households

[2]In addition, I will check that my results are robust to weighting for the number of households randomly selected to be surveyed in the block, to account for differences in attrition rates across blocks.

with certain characteristics may respond differently to questions about housing and neighbourhood quality. I include in $X_{ib,t-1}$ individual measures of all variables used to construct block level outcomes $\widetilde{X}_{b,t-1}$: *household income, household poverty, household head ethnicity dummies, household head gender, and a dummy for the household head being under the age of 30*.

Finally, in all regressions I will control for the total percentage of available housing units that are occupied by either the owner or a renter at the time of the baseline data collection (the housing block occupation rate).

4.4 Multiple hypothesis testing

In addition to reporting standard p-values, I will report False Discovery Rate q-values, using the method of Benjamini et al. (2006). I will correct for multiple hypothesis testing across the six primary outcomes of interest, but within each of the three main independent variables of interest. Following Anderson (2008) I will report the minimum q-value at which each hypothesis is rejected.

5 Main Outcomes of interest and independent variables

For the main outcomes of interest, I construct indices for families of outcomes following the procedure proposed by Anderson (2008). I recode variables so that higher values correspond to positive outcomes. I transform each outcome by subtracting its mean and dividing by the standard deviation. I then sum up the variables using weights corresponding to sum of the corresponding row from the inverted covariance matrix of all variables in the family of outcomes.

5.1 Primary outcomes of interest

My main outcome of interest will be a single index of building quality that will comprise the three main indices from the 'building survey' listed below (items 1 to 3). I will then study each of the outcomes below, in turn.

1. Building construction and finishing (building survey).
   - Index of eleven outcomes, including: painting, plastering, clearing of rubble, accessibility, walk-ways, condition of stairs. .
2. Common space and greening (building survey).
   - Index of eleven outcomes, including: planting of trees, community gardens, installation street lights, maintenance and use of communal space.
3. Management of trash, litter, and sewerage (building survey).

- Index of eight outcomes, including: smells of trash and sewerage; visible signs of sewerage, trash, litterl; vandalism and graffiti.

4. Willingness to contribute to public goods (household survey).

    - Index of: Amount of money households would contribute towards project to build a public good, willingness to contribute time to improving neighbourhood or building public goods (index), evaluation of how willing neighbours would be to spend time to improve the neighbourhood or build a public good (index).

5. Engagement in social life and neighbour-goodwill (household survey).

    - Index of 6 outcomes: Number of close friends that live in neighbourhood, attendance of community events, caring/concern among neighbours, advise among neighbours, would attend memorial of neighbours, neighbours would attend memorial for the respondent.

6. Quality of neighbourhood: crime and noise (household survey).

    - Index of: Perceptions of crime (index of 3 crime questions), perceptions of noise (index of 2 questions about noise disturbance).

5.2 Independent variables, household composition (from household listing):

1. **Ethnic fractionalization index**

    - For household level outcomes I will use the individual weighted index defined in Section 5.4 and given by $\widetilde{X}_{ib}^{m}$.
    - For block-level public goods outcomes I will use the overall block-level fractionalization measure.

2. **Block-level poverty rate:** percentage of households with baseline income per capita income below 1000 Birr ($38) per month. By this definition, roughly one third of my sample of owners are poor.

3. **Owner-occupation rate** (as a percentage of total residents).

In addition to these three main household composition variables, I will include controls for additional measures of household composition, to check that my main results are not driven by other correlated household level outcomes. I particular, in any regression where I control for the ethnic fractionalisation of the composition of neighbours, I will check for robust of the results to controlling for the block-level *shares* of ethnic groups included in the measure of ethnic fractionalization. This allows me to control for the possibility that the ethnic fractionalization is correlated with the representation of certain ethnic groups, who may have systemically different outcomes, which would influence block-level outcomes in a way that is not directly the result of fractionalization itself (Kustov and Pardelli, 2018).

I will include tests for the effects of the following block level composition measures, as secondary tests, after investigating the three main independent variables, listed above.

- Percentage of households with a female head.
- Percentage of households with a young head (30 or younger).
- Education: percentage of households with head with tertiary education
- Household income gini coefficient.

5.3 Independent variables, building fundamentals (from baseline building survey):

1. Building orientation
    - Access to a public quad between housing blocks, view of the street of staircase/balconies, access to the street, visibility of entrances ("defensible space").
2. Building remoteness:
    - Index of: distance to shops, distance to transport, distance from the main road, proximity to the periphery of the site.

5.4 Definition of block level dependent variables

If ethnic fractionalization affects block-level outcomes, it may do so through the channel of direct contact between neighbours living in the block. If individuals living in a housing block are more likely to interact with their direct neighbours (that is, neighbour living closer to them geographically *within* the housing block, then the relevant measure of fractionalization would be at a smaller geogrphic level than the level of the entire housing block.

I will construct individual level measures of neighbourhood diversity, using information about geographic proximity. That is, between any two neighbours i and j living in a given housing block b, I will construct a physical distance weight $w_i j$, between the blocks. Using the individual data on households randomly assigned to blocks, I then define the individual diversity measure for household i, along characteristic m, using the following expression:

$$\widetilde{X}_{ib}^{m} = \frac{1}{\sum_{j \neq i} w_{ij}} \sum_{j \neq i} w_{ij} d_{ij}^{m}. \tag{3}$$

where the dyadic distance between pairs differs depending on the household characteristics. For ethnic fractionalisation I use $d_{ij}^{e} = 1(ethnicity_i \neq ethnicty_j)$ (d_{ij}^{m} is equal to one if the households are of a different ethnicity, and zero if they are same ethnicity). I then average across resident households such that $\widetilde{X}_b^e$ is a weighted ethnic fractionalization index:

$$\widetilde{X}_b^m = \frac{1}{N_i} \sum_i \left(\widetilde{X}_{ib}^m \right).$$

To assign the weights w_{ij} between each households within the same block, I will use the physical distance between the households. For each block, I measure the characteristics of each household randomly assigned to the block, as well as the floor housing unit number to which they were assigned. I also collected dyadic data on neighbour-to-neighbour interactions within blocks, between randomly assigned individuals living in the housing blocks. I will use these dyadic data in order to construct weights w_{ij} based on actual physical distance between household dyad-pairs. To do this, I will run a regression of the form

$$c_{ij} = \eta_1 + \eta_2 \cdot \mathbb{1}(f_i = f_j) + \eta_3 \cdot |h_i - h_j| * \mathbb{1}(f_i = f_j) + \eta_4 \cdot |h_i - h_j| * \mathbb{1}(f_i \neq f_j) + \gamma_{f_j} + \epsilon_i \quad (4)$$

Where f_i and h_i are the floor number and house number of house i, respectively. I include floor fixed effects f_j for household j to account for the fact that households living on the lower floors are more likely to be known by everyone in the housing blocks. Here I allow the effect of distance in house numbers to differ according to whether the households are located on the same floor, or on different floors. I then use predicted interactions as weights in equation 3 above. In other words, I set $w_{ij} = \widehat{c_{ij}}$ using estimates from the regression 4.

References

Algan, Y., C. Hémet, and D. D. Laitin (2016). The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation. *Journal of Political Economy 124*(3), 696–733.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484).

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive Linear Step-up Procedures that Control the False Discovery Rate. *Biometrika 93*(3), 491–507.

Chetty, R., N. Hendren, and L. F. Katz (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review 106*(4), 855–902.

Coleman, A. (1985). *Utopia on trial: Vision and reality in planned housing*. Longwood Pr Ltd.

Cutler, D. M. and E. L. Glaeser (1997). Are ghettos good or bad? *The Quarterly Journal of Economics 112*(3), 827–872.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (2008). When are ghettos bad? lessons from immigrant segregation in the united states. *Journal of Urban Economics 63*(3), 759–774.

Kustov, A. and G. Pardelli (2018). Ethnoracial homogeneity and public outcomes: The (non) effects of diversity. *American Political Science Review*, 1–8.

Newman, O. (1972). *Defensible space*. Macmillan New York.