Unpacking Neighborhood Effects: Experimental evidence from a large scale housing program in Brazil

Carlos Alberto Belchior Gustavo Gonzaga Gabriel Ulyssea*

UZH PUC-Rio UCL

June 30, 2022

Abstract

Abstract: A large literature argues that the neighborhood of residence plays a key role in determining individuals' labor market outcomes. This paper explores a large-scale, randomized housing program in Brazil to estimate the magnitude of neighborhood effects on individual labor market outcomes. We combine the double-randomization into the housing program and across different neighborhoods, with a partial identification approach, to unpack the overall effect into neighborhoods' labor market access, quality of peers, and amenities. We find that neighborhood effects are quite heterogeneous, ranging from no impact at all to a 7.5 percentage decrease on formal employment. We show that the neighborhood's distance to job opportunities is responsible for 82% to 93% of the negative impact on formal employment, while network quality plays only a very limited role in explaining the labor market effects.

Keywords: *Neighborhood effects, Employment, Mechanisms.*

*Belchior: Department of Economics, UZH (email: carlosalbertobdc@gmail.com). Gonzaga: Department of Economics, PUC-Rio (email:gonzaga@econ.puc-rio.br). Ulyssea: Department of Economics, University College London (email: g.ulyssea@ucl.ac.uk).

> One of the greatest mistakes of my administration was to build [public housing] units in the city periphery, which tremendously increased mobility costs.
>
> *Eduardo Paes, Rio de Janeiro's mayor*

1 Introduction

Public housing is a widely used welfare program in both developed (del Pero et al., 2016) and developing countries (Barnhardt et al., 2017a). These programs can have far reaching impacts on beneficiaries, as vary beneficiaries' peers and social environment. Indeed, a large literature has emphasized the importance of neighborhood quality in determining several important outcomes of individuals' welfare, such as health, education, and labor market prospects (Chyn and Katz, 2021). However, still little is known about the mechanisms through which neighborhood quality may affect individuals, in particular their labor market outcomes.

This paper tackles this question by investigating the effects of a large housing program in Rio de Janeiro, *Minha Casa, Minha Vida* (MCMV), which provided subsidized houses for individuals in peripheral regions of this Brazilian city. We leverage the fact that the program introduced a double-randomization in the allocation of houses in of the lotteries in early 2015: not only beneficiaries were randomly allocated to the group that received the houses, but they were also randomly assigned to six different housing projects in three different neighborhoods. Crucial to our research design, these neighborhoods had quite different characteristics and offered substantially different bundles of traits.

We link information for lottery participants with rich administrative data - for the universe of formal labor market workers (*Relação Anual de Informações Sociais*) - and with data for the Brazilian Single Registry of social programs (*Cadastro Único*) for both pre and post-treatment periods. This data allow us to track the labor market outcomes, place of residence, housing characteristics, and monthly expenses for lottery participants and their family members.

We show that being drafted to the program, in fact, significantly changes individuals' lives. About half of the drafted individuals actually takes-up treatment. These individuals moved to the housing project, and the vast majority was still living there at the end of 2020, six years later. The housing provided by the program has *higher* quality than beneficiaries' previous residences but are located in neighborhoods with *lower* income, employment, and education than the ones they lived before. We analyze the program pooled effects on employment probability and show that receiving the housing subsidy decreased formal employment probability by approximately two percentage points, which is close to previous results in the literature (Jacob and Ludwig, 2012; Van Djik, 2019). We find no effect on wages, conditional on labor market participation.

Next, we show that the employment response of receiving a house is radically different depending on the neighborhood that individuals were drafted to. Estimates range from no effect at all to more than a seven percentage point drop in formal employment probability. Since the assignment to neighborhoods was also random, these differences in treatment effects cannot be explained by self-selection. We also rule out that this disparity is due to chance, differences in composition of the compliers that took up treatment in different neighborhoods, or differences in the houses' market values. Therefore, we conclude that the housing project's neighborhood characteristics must have significant consequences for employment outcomes.

Finally, we unpack *why* neighborhoods are important for labor market outcomes. There are two broad sets of models that rationalize the importance of neighborhoods (Topa and Zenou, 2015). First, it is possible that social interactions are an important channel to find jobs (Picard and Zenou, 2018). According to this hypothesis, if more educated and employed individuals interact within one's neighborhood, the higher the chance of finding a job - we call this the "network quality" theory. Second, the neighborhood proximity to jobs may be the most important determinant. If a neighborhood is far away from job centers, job search costs are higher, and this may decrease employment rates (Kain, 1968; Zenou, 2013) - this is usually referred to as the "spatial mismatch" theory. Also present (although less often emphasized) in the theoretical literature on neighborhood effects, are changes in other neighborhood characteristics, such as amenities and crime rates that may also influence residents' labor market outcomes.

The neighborhood effects literature mainly used income or poverty rates as sufficient statistics for neighborhood quality (Chyn and Katz, 2021). However, this type of measure is only informative about the "network quality" hypothesis. In order to test which of these potential mechanisms is the most important, we combine the double randomization feature of the program with partial identification tools.

We show that individuals randomly assigned to neighborhoods with higher income had actually worst labor market outcomes than those assigned to neighborhoods with lower income. Thus, using the most common approach in the literature and focusing only on the income dimension, we would conclude that better neighborhoods actually hurt beneficiaries.

Instead, we implement a different approach. We assume that the program directly affects employment, which might encompass the house wealth shock, for example, and an indirect effect, mediated through neighborhood characteristics individuals are drafted to. Then, differences in employment outcomes for individuals drafted to different neighborhoods must be explained by different neighborhood effects.

We argue that the neighborhood effects on residents are mediated through a wide vector of neighborhood characteristics: network quality, labor market access, crime rates, and amenities. This vector incorporates the main explanations present in the theoretical literature but also other additional mechanisms.

We show that by imposing theoretical sign restrictions on the effects of these mechanisms on labor market outcomes, we can provide meaningful bounds on the importance of these mechanisms. We also show that these bounds are algebraically equivalent to a Two-Sample Two-Stage Least Squares (TS2SLS) regression of formal labor market outcomes on a subset of the mechanisms instrumented by which neighborhood individuals were drafted to.

We reach two important conclusions. First, we show that labor market access is a necessary mechanism. If the labor market access was not relevant, no other combination of mechanisms could generate the observed patterns of employment outcomes. Second, we calculate that labor market access accounts for most of the variation in employment outcomes. Our bounds suggest that this mechanism explains between 82% and 93% of total neighborhood effects on formal employment. Thus, our analysis provides strong evidence in favor of the spatial mismatch hypothesis.

We contribute to two strands of literature. First, we contribute to the literature that evaluates the importance of individuals' residence neighborhoods on welfare. This literature showed consistent benefits for children growing up in better neighborhoods (Kling et al., 2007; Chetty et al., 2016; Chetty and Hendren, 2018; Nakamura et al., 2021; Hwang, 2022). Results for contemporaneous effects of place of residence on adults have been much more mixed. Kling et al. (2007), Chetty et al. (2016), and Chyn (2018) find that exogenous relocation of adults to better neighborhoods does not affect their labor market outcomes. On the other hand, Van Djik (2019) estimates that receiving a house in a high-income neighborhood enhances employment and earnings.

However, to the best of our knowledge we are the first to disentangle neighborhood effects on employment outcomes into these different mechanisms emphasized in the theoretical literature (Topa and Zenou, 2015). Indeed, the relative importance of each potential mechanism remains an open question (Chyn and Katz, 2021). Our results point to the importance of labor market access to explain neighborhood effects, in particular in the context of developing countries. The insights on the relative importance of underlying mechanisms may contribute to reconciling previous mixed effects of the neighborhood on labor market outcomes.

Second, we contribute to the much scarcer literature that studies the effects of housing programs on beneficiaries in developing countries (Barnhardt et al., 2017b). We also contribute specifically to the literature that evaluates the effects of the *MCMV* program on a range of outcomes (Rocha, 2018; Pacheco, 2019). Generally, our pooled estimates for the effects of the housing programs on labor market outcomes are in line with this literature. We find small negative effects of the program on employment. We contribute to this literature by examining the mechanisms behind these negative treatment effects.

This paper is organized in six sections besides this introduction. Section 2 provides a description of the MCMV program and the respective lotteries. Subsequently, section 3 presents our database and some descriptive statistics. Section 4 presents our empirical strategy, pooled, and neighborhood-specific results. Section 5 discusses the importance of neighborhoods. Finally, section 6 concludes the paper.

2 Background: A large scale housing program

The *Minha Casa, Minha Vida* program (MCMV) was launched in 2009, with the goal of reducing the housing deficit in Brazil. Initially, it aimed at building and financing one million houses across the country. Later, in 2011, the program's second phase was launched - aiming to provide two million more houses. By 2020, around 5 million houses had been delivered, of which around 2 percent were in the city of Rio de Janeiro, where we draw our data from.

Most of the individuals benefited by the program received houses or apartments at near-zero cost. The government subsidized between 90% and 95% of the house value. The benefited family paid the remaining value of the house during the following ten years through monthly payments, which could not exceed 10% of their total household income. Generally, these individuals provided monthly payments of approximately R$ 50 (approximately US$ 15 in 2015). In order to be eligible for most of the benefits of the program[1], individuals should have up to R$ 1.600,00 of monthly income (484 US$ in 2015), be Brazilian, older than 18, and should not own a home or have had access to home financing.

The federal government supplied funds for building houses in specific locations

[1]For richer individuals, the *MCMV* also provides financing for buying homes.

according to the measured housing deficit in each city, while the local governments provided necessary public infrastructure. Then, private construction firms presented projects to financial intermediaries (public banks) – subject to minimum criteria established at the national level. In large cities such as Rio de Janeiro, the housing projects were built in peripheral regions of the city to reduce land costs. The average cost per house was approximately R$ 63.000 in Rio de Janeiro (19.000 US$ in 2015). Since the house was almost fully subsidized by the government, the program generated a significant wealth shock to beneficiaries.

Local governments had some discretionary power to allocate the houses within the program. However, the federal government required that at least 44 percent of the available houses to be allocated to the pool of eligible beneficiaries through lotteries. Other 6% of the available houses were reserved for elderly or disabled individuals and local governments should allocate the remaining 50% to individuals in vulnerable conditions, but without necessarily relying on lotteries.

2.1 The MCMV program in Rio de Janeiro

In Rio de Janeiro, the local housing secretary created a program-specific registry of potential beneficiaries. It included individuals from the Brazilian Single Registry of social programs who lived in the municipality and individuals who the secretary actively prospected[2]. In particular, there was an emphasis on individuals who lived in slums with high environmental risk. The program-specific registry reached more than 600,000 individuals in its peak in 2015, and it was used to define the pool of potential beneficiaries to be drafted in the housing lottery.

As an attempt to avoid frauds, the *MCMV* lottery was linked to a well-established federal lottery run by a federal public bank, *Caixa Econômica Federal*. Between 2011 and 2014, the city government ran six lotteries associated to six different housing projects and distributed 8,507 houses. For these lotteries, individuals who were drafted could choose which housing project to go (among those being offered in that batch) by order of arrival. This allocation method led to a considerable burden on the municipal bureaucracy.

In the first lottery of 2015, in an attempt to simplify and make the allocation process more transparent, the city of Rio changed the allocation method. It moved to a double-randomization design, in which individuals were randomly allocated into the program and and then across different housing projects. We focus on this single lottery conducted in January of 2015 to leverage this double-randomization design.[3]

[2]The municipal Housing Secretary actively looked for individuals living in dangerous conditions and offered subscription in the registry.

[3]Later in 2015, the housing secretary opted to start drafting only individuals that lived close to

The lottery proceeded as follows: all potential beneficiaries in the program registry were ordered alphabetically and received a lottery number equivalent to their position in the list. Then, six numbers ranging from 1 to 999 were drafted from the federal lottery. Individuals were considered to be drafted if the last three digits of their lottery number corresponded to the drafted number. By design, the number of drafted individuals was equal across housing projects. However, the number of available houses was not. For the cases that the number of drafted individuals exceeded the number of available houses, houses were offered in alphabetical order and the remaining drafted individuals were included in a wait list. The design of the lottery generated wait lists of different sizes across housing projects.

The housing projects

The lottery we focus on drafted houses in six different housing projects over three different neighborhoods, *Santa Cruz*, *Campo Grande*, and *Cosmos*. A total of 2,580 houses were drafted in this lottery, being 1,500 in *Santa Cruz*, 860 in *Campo Grande*, and 220 in *Cosmos*.

As can be seen in Figure 2, all three neighborhoods are located in the west region of the city, which is very far from the city center (about 50 km). Also, housing projects in different neighborhoods are far from each other (about 13 km). If an individual wished to move from one housing project to another using public transportation, it would take her one hour and a half on average and, if she wished to go from one housing project to the center of the city, this would take on average two and a half hours.

In Table 1, we show some selected characteristics of the neighborhoods individuals were drafted to. The Table shows several interesting facts. All neighborhoods with MCMV housing projects are in the low end of the Rio de Janeiro neighborhood distribution in all displayed dimensions. Second, these neighborhoods have heterogeneous characteristics. For instance, there is a gap of 26 percentiles in the income between the neighborhood with lowest. Third, there is a substantial mismatch between the comparative qualities of each neighborhood. In section 7, we will leverage this heterogeneity to explore the relative importance of different empirical mechanisms. In Appendix A, we show additional comparisons between *MCMV* and other Rio de Janeiro's neighborhoods.

the housing project.

3 Data

We use four main data sources in this paper. First, the administrative data from *Minha Casa, Minha Vida* program lotteries, which provide information about the pool of potential beneficiaries in the city of Rio de Janeiro. These data were obtained from the municipal housing secretary and contain the names, a time-invariant identifier – the *Cadastro de Pessoa Física* (CPF) – for all individuals who participated, the housing project that each individual was randomized to, and where it was built. Program take up was not 100 percent for drafted individuals. This could happen because they missed important deadlines or incorrectly reported information (in which case they could not receive a house), or because they were no longer interested in receiving a house. Even though we do not have direct information on take-up, the rules of the program stipulated that individuals who won the lottery but did not receive a house should be automatically added to the set of potential beneficiaries of future lotteries. We can thus infer which individuals actually received a house as those drafted in a given lottery who do not appear in subsequent ones.

Our second source of data is the RAIS data, a matched employer-employee, administrative data set from the Ministry of Labor. All formal firms in Brazil are required to fill in the information about their workers annually, and the Brazilian ministries use this information for unemployment insurance and other social programs, including those conceded by the *MCMV* program. The data thus contain the universe of formal firms and workers in Brazil. We use a restricted access version of the database, in which the CPF identifier is also available, and therefore we can merge this information with the data from the housing lotteries.[4]. The RAIS data allow us to draw precise monthly information on which individuals participated in the formal labor market as well as contract characteristics, such as hours and wages. We focus on the period of 2003-2017. Because the RAIS is a matched employer-employee data set, we have information on workers' firms and, particularly important to our analysis, we know their firms' exact address. Using this information, we have geo-referenced all formal jobs in the city of Rio de Janeiro (and neighbor municipalities) for our period of analysis.

Third, we use confidential information from the Single Registry of Social Programs, the *CadUnico*.[5] We obtained information for the years 2012 to 2020. This database contains individual-level information about family composition, house-

[4]Since RAIS gathers data from all formal contracts in one year, some individuals appear multiple times in the same year. When this is the case, we keep only the contract with the higher total annual wages

[5]We thank the Ministry for Social Development for the concession of this database, according to the process 71000.000372/2018-11.

hold characteristics, income, some labor market characteristics, and household expenses. Each year, the Single Registry data contains the most recent updates for all individuals registered since 2003. The *CadUnico* also contains the CPF identifier, which allows us to merge its information with the two previous data sets.

Fourth, we use non-identified data from the Deccenial Demographic Census to compute a wide range of neighborhood characteristics. We use the 2010 Census, which is the latest available information.[6]. We aggregate the individual-level information at the census tract level, which is the smallest area for which there is representative information in the sample questionnaire.[7] From the Census data, we compute average income, education, labor market outcomes, the share of single-parent households, household quality, and labor market access for all neighborhoods in Rio de Janeiro. We complement the Census data with information on the number of schools, public daycare facilities, public hospitals, public parks, and crime rates, which are available from different sources[8]. We also scrape data for market prices of houses being sold in the neighborhoods of the *MCMV* housing projects. We use these different variables to provide a richer characterization of the amenities and public good provision in Rio de Janeiro's neighborhoods. In Appendix A, we provide more details about these data sources and how we define the different variables.

Samples

We work with two complementary samples obtained by merging the data sources discussed in the previous section. The first is our *labor market sample (lotteries+RAIS)*, which is obtained by merging the lotteries administrative data with RAIS. We keep in the sample only individuals who participated in the formal labor market in any year before the lotteries (2003-2014), which encompasses more than 80% of the universe of individuals in the MCMV registry. We keep information on their most recent participation in the baseline period. We then use the three years of endline data (2015 to 2017) to construct post-lottery labor market outcomes.[9] In order to maximize statistical power, our main specification stacks all three endline years.

The second sample used is the one of *families (lotteries + CadUnico)*. In

[6]Due to budget problems and the Covid-19 pandemic, the Census was not conducted in 2020.

[7]Only a very limited set of information is available in the full census, we therefore rely on the sample that contains a broad set of variables.

[8]Data for public facilities comes from the City Hall. Data for crimes comes from the Rio de Janeiro State Public Security Secretary (ISP).

[9]We consider 2015 to be an endline year since the lottery was implemented in the very first days of the year.

this alternative sample, we merge lottery participants with the Single Registry of Social Programs (*CadUnico*) and again only use individuals' most recent information for both the baseline (2002-2014) and endline (2015-2020) periods.[10] We are able to find a little more than 40 percent of MCMV participants either before or after the lotteries.

These two samples offer different costs and benefits to our analysis, and we see them as complementary. In the labor market sample, we have a much lower rate of attrition and monthly information on individuals. However, we only have information on formal labor market outcomes and socioeconomic characteristics of individuals. The sample of families has a substantially higher attrition, and constitutes a selected sample of lottery participants. By design, the CadUnico captures poorer individuals, but also has a larger fraction of non-white and females.[11] Despite these limitations, this alternative sample provides valuable additional information on beneficiaries, such as the place of residence, informal labor market outcomes, and information on family members.

Given that we focus on labor market outcomes as the main outcomes of interest, we use the labor market sample for most of our empirical analysis. We use the sample of families in complementary exercises. Table 2 shows the main descriptive statistics for control and drafted individuals in both samples. Panel A shows the socioeconomic variables, while Panel B summarizes the labor market characteristics.

Table 2 shows three important pieces of evidence. First, the randomization was successful, as there are no statistically significant differences between the control and drafted groups, either in the labor market nor the families samples. Second, take up was around 50 percent in the labor market sample and 63 percent in the families sample (first row). Third, the families sample that is matched to the Single Registry is highly selected. In this sample, not only the treatment take-up is higher, but the share of females is higher, individuals are less educated and poorer. In Appendix A, we show descriptive statistics for additional variables and in Appendix B, we show that the lottery also balanced the characteristics of individuals drafted to each neighborhood and their families' characteristics.

[10] We exclude information from the three months immediately following the lottery in order to avoid updates directly induced by the program.

[11] Non-white individuals are over represented among the poor, and many social programs – such as the *Bolsa Família* (the flagship conditional cash transfer program in Brazil) – have women as their main benefit recipient.

4 Program and Neighborhood Effects

In this section we estimate the overall housing program effects and separately by the three different neighborhoods. For that, we leverage the double-randomization designed described in Section 2. Next, in Section 5 we discuss a partial identification approach to unpack the determinants of the neighborhood effects we discuss in this section.

4.1 Empirical Strategy

Throughout this section, we estimate both the intent-to-treat (ITT) and treatment-on-the-treated (TOT) parameters. To estimate the ITT, we use the following reduced-form specification:

$$y_i = \beta_0 + \beta_1 * D_i + \boldsymbol{X}_i \boldsymbol{\beta_2} + \epsilon_i \tag{1}$$

where y_i is the outcome of interest, D_i is a dummy variable for individuals who win the lottery to receive a house within the *MCMV* program, $\boldsymbol{X}_i$ is a vector of covariates that includes gender, race, and schooling, and ϵ_i is an error-term.

To estimate the TOT, we replace the dummy for winning the lottery, D_i, with a dummy H_i that equals one if individual i actually receives a house. The H_i is of course an endogenous variable, so we instrument it with the dummy for winning the lottery and estimate this regression using a two-stage least square (2SLS) estimator. In our main specification, we stack all three endline years (2015 to 2017), while year-by-year effects are reported in the Appendix C. We do so to increase the sample size and power. All standard errors are clustered at the individual level.

We also take advantage of the double-randomization and estimate ITT and TOT for each of the three neighborhoods in which the 6 housing projects were constructed. For the TOT, we estimate the following:

$$y_i = \gamma_0 + \gamma_1 * H_{i,k} + \boldsymbol{X}_i \boldsymbol{\gamma_2} + \epsilon_i, \;\; k = 1, ..., 3 \tag{2}$$

where $H_{i,k}$ equals one if individual i actually receives a house in neighborhood k, which we instrument with a dummy variable for individuals who win the lottery to receive a house and are randomized into neighborhood k, denoted by $D_{i,k}$.

4.2 Results

We start by examining whether drafted individuals actually received a house and moved. To do so, we calculate the fraction of drafted individuals living in each neighborhood in the baseline and in the endline using the families' sample. As Figure 3 shows, at baseline individuals' houses were located all over the city and were not particularly concentrated in the neighborhoods where the *MCMV* housing projects are located. In contrast, we observe a massive concentration of drafted individuals living in these neighborhoods after the lottery. In Appendix B, we show that individuals started moving to the housing project six months after the lottery and the fraction of beneficiaries' that reported living in the *MCMV* housing project remained stable until the end of 2020.

Table 3 provides the first piece of evidence on how the housing lottery affected individuals' outcomes. The table shows that winning the housing lottery increases the housing quality (as measured by the number of rooms and bedrooms), significantly decreases the rent paid by beneficiaries an increases transportation costs. However, the neighborhood quality individuals live in decreases substantially on dimensions such as average income, share of individuals employed or formally employed and formal jobs availability.

Table 4 turns to the overall program effects on formal labor market outcomes. It shows both the ITT and TOT estimates for the probability of being formally employed, monthly hours worked, number of months formally employed in the year, the probability of keeping the same job one year after the lottery, and the log of individuals' formal wages.[12] We find overall negative treatment effects on formal employment. Focusing on the TOT, treated individuals have a 1.9 percentage points lower probability of being formally employed, work less hours and spend less time formally employed in a given year. Given the baseline employment distribution, these results suggest an decrease in aggregate employment probability of approximately 1.0 percentage point. Beyond the difference in employment probability, we also estimate that drafted individuals that continue to be employed are 4.1 percentage points less likely to keep working in the same firm. We find no significant differences in wages conditional on employment.

In Appendix C, we show that the program did not significantly affect informal labor market outcomes of beneficiaries and provide more detailed dynamic effects of the program. We also show that the negative pooled effects of the program are concentrated on white and more educated individuals. Finally, we show that the pooled estimates are slightly smaller if we use an ANCOVA or a difference-in-difference specification.

[12]We include the zeros in the hours and months formally employed, while wages are measured conditional on employment.

Neighborhood effects

Table 5 shows the neighborhood-specific treatment effects, which are obtained estimating regression 2. These results show that the pooled results in Table 4 mask substantial heterogeneity. Treatment effects on formal employment range from no effect to very large decreases in employment probability that reach more than 7 percentage points. We formally test whether neighborhood-specific coefficients are equal to each other and reject this null hypothesis for all dependent variables ($p < 0.1$).

The beneficiaries were randomly allocated to these different neighborhoods and received identical houses. Even though one can be tempted to immediately attribute these differences in treatment effects to differences in neighborhood characteristics, there is a host of other potential explanations that could be driving the results. The first could be sampling variation. We thus implement a "pooling factor metric" (Gelman and Hill, 2007) to decompose observed heterogeneity into sampling and true heterogeneity.[13] We can see that the sampling variance heterogeneity can explain only between 15-18% of the total observed heterogeneity. These results suggest that it is extremely unlikely that the observed heterogeneity can be explained by chance.

Second, even though the assignment to different neighborhoods was random, the instrumental variable estimator identifies the causal effect of receiving a house on compliers (Angrist et al., 1996). Thus, it is possible that differences in compliers across neighborhoods could be explaining the differences in treatment effects. In Table 6, we examine this hypothesis by comparing the characteristics of compliers in each neighborhood. At first glance, take-up rates appear to be very heterogeneous across neighborhoods, which is due to the number of houses being offered in each neighborhood. Once we exclude the wait-list, take-up rates are very close across groups. Also, once we condition on beneficiaries who accepted the house, characteristics are very similar and statistically equal across neighborhoods. In Appendix D, we formally show that differences in compliers' characteristics cannot explain the heterogeneity in neighborhood-specific treatment effects. We implement Angrist and FernÃ¡ndez-Val (2010) methodology and show that, if compliers in neighborhoods 1 and 2 had the same characteristics as compliers in neighborhood 3, the difference in neighborhood-specific treatment effects would

[13]Let $\hat{\beta}_n$ be the estimated treatment effect of receiving a house at neighborhood n. Let $\hat{\sigma}^2_\beta$ be the variance of estimated treatment effects. Also, let $\hat{se}_n$ be the standard errors associated with each treatment effect. Then, we calculate the fraction of observed heterogeneity explained by sampling errors as:

$$\omega = \frac{1}{3}\Sigma_{n=1}^{3}\frac{\hat{se}_n^2}{\hat{se}_n^2 + \hat{\sigma}^2_\beta}$$

be even slightly higher than shown in Table 5.

A third potential explanation is that, even though houses are identical, their market value varies substantially across neighborhoods. If this was the case, the heterogeneity in estimated treatment effects could be reflecting the magnitude of the wealth shock received. To examine this hypothesis, we collect two different measures of houses market values across neighborhoods. The first comes from scraped data on house characteristics and respective prices across all three *MCMV* neighborhoods (we provide details on the scraping procedure in Appendix A). Despite being prohibited, some individuals sell their *MCMV* homes, which allows us to directly compare the market value of the *MCMV* houses.[14] The second measure comes from Census data, which includes individual-level information on rents paid. The advantage of using Census information is that it provides a representative sample of individuals paying rent in those neighborhoods, as well as a wide array of house characteristics that is not available in our scraped data.[15]

We use these data to estimate the following simple regression:

$$log(y_h) = \alpha_0 + \alpha_1 * D_{1,n} + \alpha_2 * D2, n + \boldsymbol{\alpha} * \boldsymbol{X}_h + \epsilon_h \quad (3)$$

where y_h is the outcome of interest for house h (either the house selling price or rent being paid), $D_{k,n}$ denotes a dummy for a house located in neighborhood $k = 1, 2$, and $\boldsymbol{X}_h$ is a vector for available house characteristics.

We run this regression for three different samples: (i) all houses; (ii) only houses with similar characteristics to the *MCMV* ones; and (iii) *MCMV* houses being sold online. Results are shown in Table 7. Houses in neighborhoods 1 and 2 have higher prices than houses in neighborhood 3, but differences are quite small: for the whole sample, prices in neighborhoods 1 and 2 are approximately one percentage point higher (Column 1).

If we focus on houses with similar characteristics to the *MCMV* houses, the price difference becomes much smaller and it completely vanishes when we restrict to the sample of *MCMV* houses. The results for rents paid indicate that there are no statistically significant differences between neighborhoods. In sum, the difference in wealth shocks received by beneficiaries appears to be quite small across neighborhoods. Therefore, it is very unlikely that this difference is able to explain the large difference in neighborhood-specific treatment effects we found

[14] We can identify those houses either by the combination of addresses and housing characteristics or by the description of the houses. The number of *MCMV* houses being sold is small, approximately 80 houses relative to a total of approximately 15,000 houses provided by the program in those neighborhoods.

[15] The Demographic Census includes information on the number of rooms and bathrooms, dummies for the supply of potable water, and energy, household adequacy, the type of the house (whether a house, apartment, house in a condo, etc.) and the external material of those houses.

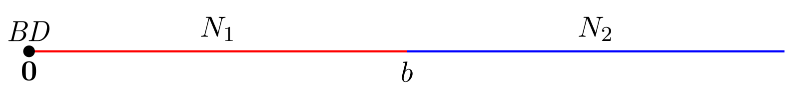
in Table 5.

5 Unpacking neighborhood effects

5.1 Theoretical framework

We start this section by illustrating the potential mechanisms that might mediate the relation between place of residence and employment using a toy model similar to Picard and Zenou (2018). This discussion will also rationalize our empirical approach.

Consider a unit width linear city with workers residing along it. The city locations are indexed by x. Workers might live in two different neighborhoods, $N_i \in [0,1]$, $i = 1,2$, defined as the support set along the city line. Workers are *a priori* uniformly allocated along the city and neighborhoods' population sizes are given by P_i. The city has a central business district (BD) located at $x = 0$ and all job opportunities are concentrated there. The border between the the two neighborhoods is given by b. We also assume that each neighborhood has an exogenous characteristic α_i that can directly affect individual outcomes (more details below). We can represent the city as follows:



All workers receive the same wage - $w \in (0,1)$[16], conditional on being employed, and have the same preferences:

$$U_i(x) = e_i(x)(w - t * x) - C_i(x) \tag{4}$$

where $e_i(x)$ is the employment probability, t is a commuting cost to the business center, and $C_i(x)$ is the cost of social interactions, discussed below. Employed individuals might lose their job with probability γ. When individuals are unemployed, they search for a job with success probability $\pi(x)$.For each neighborhood, the total employment is defined by $E_i = \int_{N_i} e_i(x)dx$.

In this model, the main channel for finding jobs are information flows from social interactions. Individuals interact only with other individuals that live in their neighborhood. Workers choose how many individuals to socially interact with, $n_i(x)$, and randomly meet individuals living in the same neighborhood as them.

[16]We bound wages to one, so that other quantities of the model below adjust accordingly.

Let neighborhood average income be denoted by $m_i = w * \frac{E_i}{P_i}$. We define job search success probability as:

$$\pi_i(x) = \alpha_i * n_i(x) * m_i \tag{5}$$

that is, the probability of finding a job increases with the number of individuals they decide to meet, the average income on their neighborhood and exogenous neighborhood characteristics, represented by α_i.

The role of neighborhood average income captures the idea that higher quality peers in one's neighborhood – i.e. with higher employment rate and, consequently, higher income – are likely to provide better information about available jobs, therefore increasing the likelihood of success in finding a job. In Picard and Zenou (2018), α_i is modeled as an increasing function of the neighborhood population size, representing the network size provided by the neighborhood. In this model, we allow the probability of job search success to be affected by a broader range of neighborhood characteristics, which may include: crime rates, amenities, public good provision, etc. These are treated as exogenous in the model. Individuals incur in a constant cost c to interact with others, so that the total cost of social interaction is simply $C_i(x) = c * n_i(x)$. Workers choose the number of social interactions that maximizes their total utility:

$$\max_{n_i(x)} \quad e_i(x)(w - t * x) - n_i(x) * c$$

Equilibrium and characterization of neighborhood effects

Note that in equilibrium, we must have that:

$$e_i(x) = \frac{\pi(x)}{\gamma + \pi(x)} \tag{6}$$

We can combine equations (5) and (6), so that we can write employment probability as:

$$e_i(x) = \frac{\alpha_i * n_i(x) * m_i}{\gamma + \alpha_i * n_i(x) * m_i} \tag{7}$$

Thus, the first-order condition for the individual problem should be:

$$e_i^*(x) = \left(1 - \sqrt{\frac{\gamma * c_i(x)}{\alpha_i(w - t * x) * m_i}}\right) \tag{8}$$

which implies the following expected unemployment probability:

$$u_i^*(x) = \sqrt{\frac{\gamma * c_i(x)}{\alpha_i(w - t * x) * m_i}} \tag{9}$$

An equilibrium in this model is a spatial distribution of employment, $e^*(x)$; social interactions, $n^*(x)$; and aggregate employment rate for each neighborhood, $E_i = \int_{N_i} e_i^*(x)dx$. This model is flexible enough to accommodate a range of interesting different equilibria. We will focus on the equilibrium where $\frac{E_2}{P_2} > \frac{E_1}{P_1}$.[17] That is, we focus on an equilibrium where individuals in the distant neighborhood are more likely to be employed[18].

Now, consider a policy experiment in this model. Suppose that we move one individual from x_1 to x_2, where $x_1 < b < x_2$, that is, moving the individual from neighborhood 1 to neighborhood 2 and far away from the central business district. We can evaluate the impact of this experiment by comparing the unemployment ratio before and after the residential change:

$$\frac{u_2^*(x)}{u_1^*(x)} = \underbrace{\sqrt{\frac{(w - t * x_1)}{(w - t * x_2)}}}_{(a):>1} * \underbrace{\sqrt{\frac{m_1}{m_2}}}_{(b):<1} * \underbrace{\sqrt{\frac{\alpha_1}{\alpha_2}}}_{(c):<1} \tag{10}$$

where the three different mechanisms above can be described as follows::

(a): Spatial mismatch mechanism: This term is greater than one and reflects the fact that individuals who move far away from the business center pay higher mobility and interaction costs, which increases the likelihood of unemployment.

(b): Network quality mechanism: it is smaller than one, as individuals move to a neighborhood with a higher share of employed peers and, therefore, higher income. This is the main focus of the empirical literature on neighborhood effects.

(c): The direct effect of neighborhood exogenous characteristics on employment, possibly encompassing its infrastructure, crime rates, among other neighborhood characteristics, which are fixed in the model.

The net effect of these forces is *a priori* ambiguous and depends on the relative strength of these different forces, therefore being ultimately an empirical question. In the next section we propose a partial identification approach to try to disentangle the relative importance of these different mechanisms.

[17] The condition for this equilibrium to prevail is that the exogenous characteristics of the neighborhood 2 to have a large enough effect on employment probability. Formally, this equilibrium will prevail if:

$$\sqrt{\frac{\gamma}{\alpha_1}} * \int_0^b \sqrt{\frac{1}{w - t * x}} dx > \sqrt{\frac{\gamma}{\alpha_2}} * \int_b^1 \sqrt{\frac{1}{w - t * x}} dx$$

[18] As pointed out by Picard and Zenou (2018), this might correspond to a city such as New York or Chicago where disadvantaged workers reside close to the job center while privileged workers live at the outskirts of the city

5.2 Separating Mechanisms

To quantify the relative importance of the potential mechanisms highlighted in the previous section, we use a similar approach to the one proposed in Dix-Carneiro et al. (2018). As in their setting, we need to impose some structure to the problem in order to make progress. We assume that the effect of receiving a MCMV house on the employment outcome of interest can be described by the following equation:

$$y_{in} = \alpha * D_i + \mathcal{M}_n \beta + \varepsilon_{in} \tag{11}$$

where y_{in} is the outcome of interest for individual i living in neighborhood n, D_i is a dummy variable for winning the lottery and $\mathcal{M}_n$ is a vector of mechanisms.

This formulation implies that we decompose the effects of the treatment into a direct effect of receiving a house – D_{in}, which incorporates the wealth shock – and an indirect effect that is mediated through neighborhood characteristics, $\mathcal{M}_n$. One way to parse-out the direct and indirect effects is to explore variation of treatment effects across neighborhoods:

$$\Delta E(y_n) = \Delta \mathcal{M}_n \beta + E(\Delta \varepsilon_n) \tag{12}$$

The mechanism analysis usually requires that $E(\Delta\varepsilon_n|\Delta\mathcal{M}_n) = 0$. This is a demanding condition, which requires that two strong assumptions hold. First, the individual neighborhood choice must be exogenous, which is the fundamental empirical challenge in neighborhood effects literature (e.g. see Van Djik (2019)). As we argue below, the double randomization of houses and neighborhoods in our application introduces additional exogenous variation to aid identification. Second, the total mediation assumption must hold (this is a common hypothesis in mechanism analysis, see Acharya et al. (2016)). That is, there is no omitted mechanism that might simultaneously explain neighborhood effects and that correlates with the included mechanisms.

Even though we cannot exhaust all potential mechanisms, we are able to include the main theoretical mechanisms emphasized by the literature. We consider labor market access, which captures the proximity to jobs and corresponds to the spatial mismatch assumption (Kain, 1968; Picard and Zenou, 2018); the network quality in the neighborhood, as emphasized in (Chetty et al., 2016; Chetty and Hendren, 2018; Van Djik, 2019). We also consider other neighborhood characteristics that are not so frequently emphasized by the theoretical literature, but may also influence labor market outcomes, such as the availability of amenities (Roback, 1988; Krupka and Donaldson, 2007; Moretti, 2010); crime rates (Grogger, 1997, 1998; Huang et al., 2004; Freedman et al., 2018); and the neighborhood population size, which proxies individual's network diversity (Picard and Zenou, 2018). We thus assume that the vector of potential mechanisms that might mediate neighborhood effects is given by:

$$\mathcal{M}_n = \{MA_n, NQ_n, Am_n, Cr_n, PP_n\}$$

where MA_n is labor market access, NQ_n is the network quality, Am_n is the number of amenities, Cr_n is the crime rate, and PP_n is the population size.

We measure each of these variables as follows. Market access is measured as the average distance between residences in a neighborhood and low-skill formal job availability, weighted by the number of residents and jobs. Network quality is measured as average household income, which tends to be very correlated with other measures, such as education, household quality, etc (Chyn and Katz, 2021). Amenities are measured as the principal component of the presence of public parks, schools, and daycare centers in each neighborhood. Crime rates are measured as the principal component of the incidence of several crimes (robbery, burglary, and homicides) in each neighborhood. Finally, network diversity is measured as the census tract population size. We provide more details on data sources and definitions on Appendix A.

To fix ideas, we use neighborhood 3 as the reference group. Using expression 11, the difference in average outcomes between neighborhoods 1 and 2 relative to neighborhood 3 can be expressed as follows:

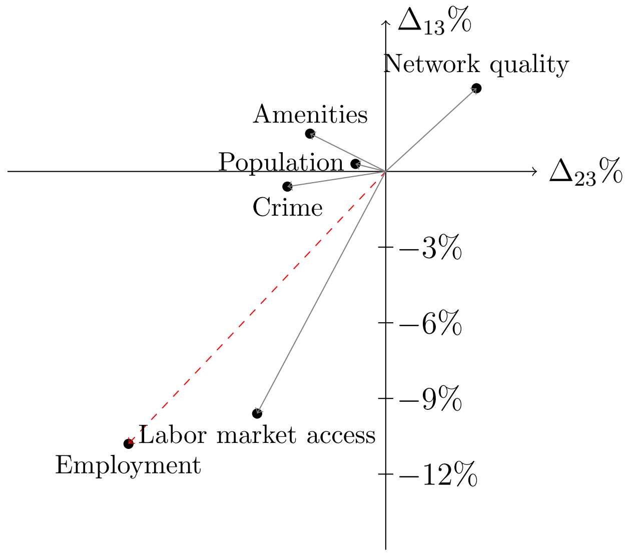
$$\begin{aligned}\Delta y_n = \beta^m \begin{bmatrix}\Delta MA_1 \\ \Delta MA_2\end{bmatrix} + \beta^n \begin{bmatrix}\Delta NQ_1 \\ \Delta NQ_2\end{bmatrix} + \beta^a \begin{bmatrix}\Delta Am_1 \\ \Delta Am_2\end{bmatrix} + \\ + \beta^c * \begin{bmatrix}\Delta Cr_1 \\ \Delta Cr_2\end{bmatrix} + \beta^p * \begin{bmatrix}\Delta PP_1 \\ \Delta PP_2\end{bmatrix} + \Delta\varepsilon_n\end{aligned} \tag{13}$$

To point identify the effects of all different mechanisms, we would need to have randomization into different neighborhoods that differ along these different dimensions. However, it is possible to make progress by imposing theoretical restrictions on the signs of the β coefficients in equation 13 to shed light on the relative importance of these mechanisms. We restrict the coefficients of labor market access and network quality to not having an opposite sign to the theoretical prediction, $\beta^m, \beta^n \geq 0$. As for amenities and population, we also assume that they cannot harm employment outcomes $\beta^a \geq 0$, $\beta^p \geq 0$ while crime rates cannot improve labor market outcomes, $\beta^c \leq 0$.

We can illustrate the argument by plotting the differences in estimated treatment effects and changes in mechanisms across neighborhoods. We do so in Figure

1, with the differences between neighborhoods 1 and 3 in the horizontal axis and neighborhoods 2 and 3 in the vertical axis:

Figure 1: Relative treatment effects x differences on mechanisms



Note: This Figure percent changes in employment outcomes and potential mechanisms for individuals drafted to neighborhoods one and two, relative to individuals drafted to neighborhood three. The former is highlighted as the dashed red arrow, while the black arrows represent the latter.

We start by noting that the employment treatment effect should be generated by a positive combination of the vectors displayed in the figure above. Thus, the first conclusion from Figure 1 is that it is not possible to generate the observed employment effects without relying on the vector for labor market access. That is, labor market access must play a significant role in explaining the employment effects of the program. Second, it is possible to impose upper and lower bounds on β^m (the parameter of labor market access). We can do so by expressing the neighborhood-specific effect as positive linear combination of β^m and β^c (the crime coefficient) or β^a (amenities). The former provides the lower bound, while the latter constitutes the upper bound.

We show in Appendix E that it is possible to derive bounds for the importance of market access. Let θ be the vector of treatment effects. Then, we can show that it is possible to write the lower and upper bounds as:

$$\frac{-\theta_1 \Delta Cr_2 + \theta_2 * \Delta Cr_1}{\Delta MA_1 \Delta Cr_2 - \Delta MA_2 \Delta Cr_1} \leq \beta_m \leq \frac{-\theta_1 \Delta Am_2 + \theta_2 * \Delta Am_1}{\Delta MA_1 \Delta Am_2 - \Delta MA_2 \Delta Am_1} \tag{14}$$

We can estimate these bounds by replacing the population parameters by the estimates in the equation above. Alternatively, we show in Appendix E that it is possible to recover these bounds using a Two-Sample Two-Stage Least Squares

(TS2SLS) of employment outcomes on labor market access and the other relevant mechanism (crime for the upper bound and amenities for the lower bound) instrumenting the mechanisms by dummies of being drafted to neighborhoods one and two[19].

It is also possible to use this equivalence to a particular TS2SLS estimator to conduct inference. In Table 8, we show estimates of the lower and upper bounds for β^m and the robust standard-errors for the TS2SLS estimator derived by Pacini and Windmeijer (2016). Again, details on estimates of the variance are provided in Appendix E.

We can see that labor market access explains most of the neighborhood effects on labor market outcomes - between 82% and 93% of total effects. The estimated sample uncertainty around the bounds is relatively small.

6 Conclusions and implications

In this paper, we examined the effects of a large housing program in Brazil on employment probability and other labor market outcomes. We used the lotteries from the *Minha Casa Minha Vida* program that took place within the Rio de Janeiro municipality to identify the causal impacts of the housing program on the recipients. The program not only randomly allocated houses to beneficiaries but, in a specific lottery, randomly drafted beneficiaries to six housing projects in three different neighborhoods.

We found that the program decreased formal employment by almost two percentage points. We also found effects on the intensive margin of employment, number of employed months, and monthly hours but found no effect on wages, conditional on employment. Next, we show that these pooled estimates uncover substantial heterogeneity. Estimates for employment probabilities range from zero effects to a seven percentage point drop in employment probability. We show that the observed heterogeneity in neighborhood-specific treatment effects cannot be explained by sampling variation or different compliers taking-up treatment in each neighborhood.

Then, we proceed to explore why we observe such heterogeneity in neighborhood-specific treatment effects. We combine the random assignments to neighborhoods with partial identification tools to infer more important mechanisms to determine employment outcomes. We find that relocating beneficiaries to neighborhoods with low income or poverty rates does not affect their employment rates while relocating them to neighborhoods with low labor market access dramatically de-

[19] See Angrist and Krueger (1992) and Inoue and Solon (2010) for a formal derivation of the properties of the TS2SLS estimator.

creases their employment. We show that most of the variation (82% to 93%) in treatment effect can be only be attributed to the proximity to formal jobs.

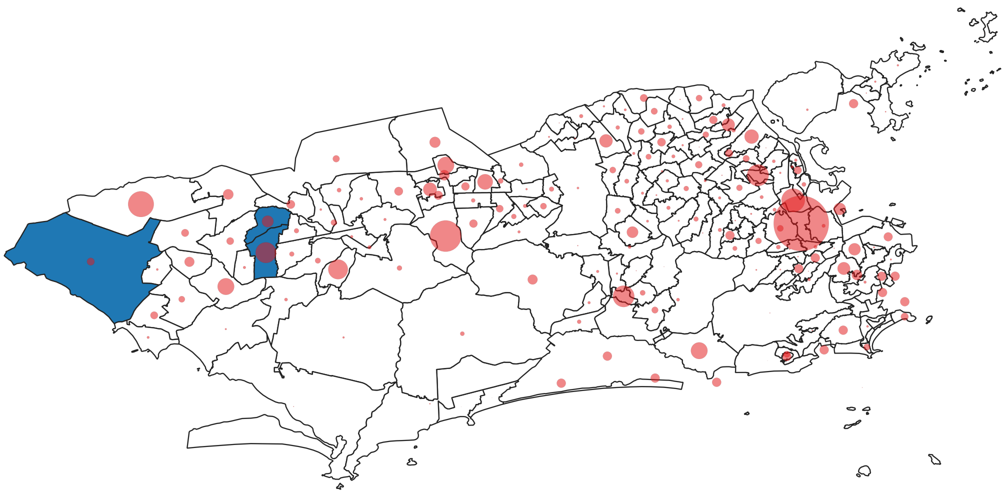
Our results have three important implications. First, similarly to the previous literature, we showed that the housing assistance program adversely affects labor market outcomes. Second, we showed that the neighborhoods where houses are provided are an important determinant of short and medium-run labor market outcomes. The final implication of this paper is that the increased mobility cost generated by the distance where the houses are offered relative to job opportunities in the underlying neighborhood characteristic influences labor market outcomes. These results have direct implications for housing programs' design and other public policies that affect individuals' spatial location.

Tables and Figures

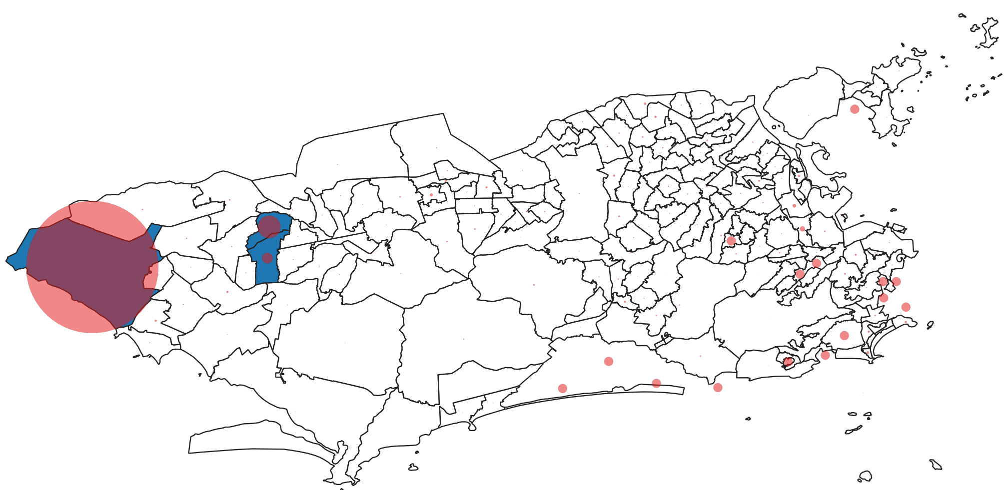
Figure 2: MCMV neighborhoods (hashed areas) and the center of Rio de Janeiro city (red area)

Note: This Figure shows the Rio de Janeiro map and its census tracts. The census tracts where the *MCMV* housing projects are located are crosshatched and the city center, which is equivalent to a business district, is marked in red.

Figure 3: Relative number of drafted individuals by neighborhood



(a) Baseline



(b) Endline

Note: This Figure shows the Rio de Janeiro map and its census tracts. The census tracts where the *MCMV* housing projects are located are highlighted in blue. In both panels, we represent the fraction of individuals that were drafted living in each census tract as the size of the red bubbles. In Panel A, we restrict the sample to updates of drafted individuals before the lottery and, in Panel B, to updates after the lottery.

Table 1: Comparisons of neighborhood characteristics (measured in percentiles)

	Neighborhood 1	Neighborhood 2	Neighborhood 3
Income	32	18	6
Wages	30	31	25
Education	30	15	6
Single parent households	29	20	10
Market access	15	17	33
Formal jobs (RAIS)	29	27	26
Amenities	20	32	25
Crime rates	10	13	14

Note: This table presents the percentiles of several neighborhood characteristics relative to Rio de Janeiro's neighborhood distribution. Income is measured as average household income from all sources, wages is the average labor compensation for working adults, education is the number schooling years, and single parent households is a dummy for households with children and only one parent living in the household. These three variables are measured using Census data. Market access is the average distance between residences and low-skill formal job posts, weighting both the number of residences and job posts in each postal code. This is calculated using both RAIS and Single Registry data. Formal jobs is the total number of formal jobs located in each neighborhood. The level of amenities is calculated as the principal component of the total number of public parks, schools, and day care centers located in each neighborhood. Crime rates are measured as the principal component of the incidence of several crimes (robbery, burglary, and homicides).

Table 2: Descriptive statistics and balancing tests

	Labor market sample		Families sample	
	Control	Drafted Diff.	Control	Drafted Diff.
Treated	0	0.507***	0	0.631***
		(0.007)		(0.007)
Panel A: socioeconomic characteristics				
White	0.38	-0.001	0.269	0.003
		(0.007)		(0.010)
Male	0.445	-0.001	0.200	0.017
		(0.007)		(0.009)
Disability	0.126	0.006	0.169	0.011
		(0.005)		(0.011)
Schooling	3.180	0.001	2.887	0.058
		(0.025)		(0.049)
Panel B: labor market characteristics				
Ever employed	1	0	0.661	-0.001
				(0.011)
Hours	41.904	-0.059	42.372	0.094
		(0.071)		(0.111)
Wages	1267.74	-23.737	666.91	22.617
		(23.839)		(22.668)
Tenure	45.55	1.131	24.395	2.244
		(1.029)		(1.198)
Distance to previous job (km)	57.63	1.288	68.66	0.656
		(1.121)		(1.356)
Last employment year before 2015	2012.96	-0.033	2012.54	-0.062
		(0.032)		(0.077)
Joint test (p-value)		0.59 (0.81)		1.15 (0.33)
Observations	503,880		264,537	

Note: This table presents variable means for the control group (columns 1 and 3) differences for the drafted group (columns 2 and 4). Columns 1 and 2 show data for the baseline labor market sample and columns 3 and 4 show information for the updates of the families sample in the post-lottery period. Treated is a dummy for individuals that were drafted and did not return to the lottery registry. White, Males, and Disability are all dummies for individual individuals characteristics. Schooling is a five-level index indicating the highest level of instruction that ranges from no schooling to post-graduation. Ever employed is a dummy for individuals that held a formal job in any year between 2002 and 2014. Hours shows the number of monthly hours registered in contract in the last year individuals' held a formal job.

Table 3: Effects of winning the lottery on housing and neighborhood characteristics

	ITT	Control mean
Panel A: Housing characteristics		
Lived in a MCMV neighborhood	0.345***	0.030
	(0.013)	
Number of rooms	0.412***	4.006
	(0.035)	
Number of bedrooms	0.193***	1.364
	(0.018)	
Rent (R$)	-36.70***	118.09
	(6.428)	
Transportation expenses	6.545***	9.524
	(0.118)	
Panel B: Neighborhood characteristics		
Percentile of income (census)	-9.537***	37.396
	(0.826)	
Percentile of individuals employed (census)	-3.876***	38.789
	(0.876)	
Percentile of individuals formally employed (census)	-3.487***	47.658
	(0.895)	
Percentile of formal jobs (RAIS)	-5.328***	36.765
	(0.777)	

Note: This table presents treatment effects of winning the housing lottery in the *MCMV* program (column 1) and the mean for the control group (column 2). Results are shown for the families' sample. In Panel A, variables come from the Single Registry and are defined at the individual-level. Lived in a *MCMV* neighborhood is a dummy for individuals that reported living in the same postal code as one of the *MCMV* housing projects. Number of rooms and bedrooms are reported in individuals' updates to the Single Registry. Rent and transportation expenses are reported as nominal values in R$ at the moment of the Single Registry update. We deflate nominal prices using the Price Index for the Wide Consumer (IPCA) for January of 2021. In Panel B, variables are defined at the neighborhood level. Income is measured as average household income from all sources, fraction of individuals employed and formally employed is defined as the fraction of working-age adults living in each neighborhood that report being employed or formally employed in census data. Formal jobs is the total number of formal jobs located in each neighborhood. All variables in Panel B are normalized to percentiles relative to the Rio de Janeiro's neighborhood distribution. Clustered standard-errors at the individual-level are shown in parenthesis. * $p<0.1$, ** $p<0.05$, and *** $p<0.01$.

Table 4: Effects on formal employment outcomes

	Control mean	ITT	TOT	Number of observations
Formal employment	0.63	-0.010*	-0.019*	1,511,631
		(0.05)	(0.010)	
Monthly hours	26.35	-0.467**	-0.987**	1,511,631
		(0.244)	(0.423)	
Number of months employed	5.91	-0.113*	-0.206*	1,511,631
		(0.572)	(0.114)	
Kept the same job one year after the lottery	0.73	-0.024**	-0.041**	344,307
		(0.011)	(0.020)	
Log(wages)	7.59	-0.009	-0.017	749,633
		(0.010)	(0.018)	

Note: This table presents the mean for the control group (column 1), ITT (column 2), and TOT (column 3) estimates. Results are shown for the labor market sample. Observations are stacked for the years of 2015, 2016, and 2017. Formal employment is a dummy for being formally employed at any point in each year. The variable monthly hours measures the number of contractual hours of individuals formally employed (0 if the individual was not employed). Number of months employed is the total number of months in each year individuals were employed (0 if not employed). Log(wages) is the natural logarithm of the average wages individuals received during the year while employed (individuals that were not formally employed in a certain year were dropped from the analysis). Wages are reported as nominal R$. We transform this variable in real values using a broad Consumer Price Index (IPCA) for January of 2021. We include controls for age, race and schooling. Clustered standard-errors at the individual-level are shown in parenthesis: * $p<0.1$, ** $p<0.05$, and *** $p<0.01$.

Table 5: Heterogeneous treatment effects by neighborhood

	Employment	Hours	# of months
Neighborhood 1	-0.066***	-2.878***	-0.774**
	(0.024)	(1.036)	(0.281)
Neighborhood 2	-0.074***	-3.466***	-0.889**
	(0.024)	(1.029)	(0.241)
Neighborhood 3	0.012	0.333	0.100
	(0.014)	(0.603)	(0.164)
Equality of coefficients (p-value)	0.06	0.08	0.05
Pooling factor metric	0.16	0.15	0.18
Observations		1,511,631	

Note: This table presents neighborhood-specific treatment effects of receiving a house of the *MCMV* program. Results are shown for the labor market sample. Observations are stacked for the years of 2015, 2016, and 2017. Formal employment is a dummy for being formally employed at any point in each year. The variable monthly hours measures the number of contractual hours of individuals formally employed (0 if the individual was not employed). Number of months employed is the total number of months in each year individuals were employed (0 if not employed). We include controls for age, race, and schooling. We also show an F test for the joint equality of treatment effects across neighborhoods and a pooling factor metric that decomposes the observed heterogeneity in coefficients that is due to sampling uncertainty. Clustered standard errors at the individual level are shown in parenthesis: * p<0.1, ** p<0.05, and *** p<0.01.

Table 6: Balancing tests for compliers

	Neighborhood 1	Neighborhood 2	Neighborhood 3	Joint p-value for differences
Panel A: take-up				
Received house	0.41	0.65	0.56	0.00
Received house (excluding wait list)	0.57	0.65	0.64	0.06
Panel B: socioeconomic characteristics				
White	0.36	0.40	0.35	0.06
Male	0.45	0.44	0.45	0.94
Schooling	3.76	3.75	3.78	0.76
Panel C: labor market characteristics				
Ever employed	1	1	1	
Hours	41.85	41.84	41.91	0.96
Wages	1232.21	1161.68	1223.15	0.65
Tenure	51.01	48.89	44.26	0.15
Last employment year before 2015	2012.74	2012.72	2012.89	0.27
Joint Hypothesis (p-value)		0.31		

Note: This table presents variable means for individuals who received houses in each neighborhood for the baseline labor market sample. White and Males are dummies for individual characteristics reported by the firm in RAIS. Schooling is a five-level index indicating the highest level of instruction that ranges from no schooling to post-graduation. Ever employed is a dummy for individuals that held a formal job in any year between 2002 and 2014. Variable hours indicates the number of monthly hours registered in contract in the last year individuals' held a formal job in 2014. Wages show the average monthly compensation registered in the contract for individuals who appeared in RAIS last year. Tenure is the number of months since the individual was hired. The last employment year before 2015 is the last baseline year the individuals held a formal job. Robust standard errors for the difference in means are shown in parenthesis. The joint test line shows the F statistic and respective p-value for the joint test that all differences in means (except for the treated variable) are equal to zero. * p<0.1, ** p<0.05, and *** p<0.01.

Table 7: Differences in house prices and rents across neighborhoods

	All	MCMV Similar	MCMV houses
	(1)	(2)	(3)
Panel A: Houses prices			
Neighborhood 1	0.012**	0.008*	0.006
	(0.006)	(0.005)	(0.021)
Neighborhood 2	0.008*	0.006	0.008
	(0.005)	(0.005)	(0.023)
Observations	1,254	842	83
Panel B: Rents			
Neighborhood 1	-0.048	0.068	
	(0.031)	(0.051)	
Neighborhood 2	0.022	-0.033	
	(0.022)	(0.047)	
Observations	566	154	

Note: This table presents the correlation between the log of house prices and dummies for residences located in the *MCMV* neighborhoods. We estimate results for the full sample (Column 1), houses of similar characteristics to *MCMV* ones (column 2), and to the sample of *MCMV* houses (Column 3). We control for the area of the houses, the number of rooms, the number of bedrooms, and dummies for the presence of garage and and lobby in columns (1) and (2). Robust standard-errors for the difference in means are shown in parenthesis: * p<0.1, ** p<0.05, and *** p<0.01.

Table 8: Bounds on the effects of labor market access on employment outcomes

	Lower bound LB^{β_m}	Upper bound UB^{β_m}
Fraction explained by labor market access	0.821	0.934
	(0.031)	(0.026)

Note: This Table shows the lower and upper bounds for the fraction of total employment effects explained by labor market access. The fraction is calculated as $f^m = \frac{\beta^m * (\Delta MA_1 + \Delta MA_2)}{\theta_1 + \theta_2}$. Both bounds and robust standard-errors are estimated using the TS2SLS.

References

Acharya, A., M. Blackwell, and M. Sen (2016). Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review 110*(3), 512â529.

Angrist, J. and I. FernÃ¡ndez-Val (2010). Extrapolate-ing: External validity and overidentification in the late framework. In *Advances in Economics and Econometrics*, Chapter 11, pp. 401–434.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*(434), 444–455.

Angrist, J. D. and A. B. Krueger (1992). The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples. *Journal of the American Statistical Association 87*(418), 328–336.

Barnhardt, S., E. Field, and R. Pande (2017a). Moving to opportunity or isolation? network effects of a randomized housing lottery in urban india. *American Economic Journal: Applied Economics 9*(1), 1–32.

Barnhardt, S., E. Field, and R. Pande (2017b, January). Moving to opportunity or isolation? network effects of a randomized housing lottery in urban india. *American Economic Journal: Applied Economics 9*(1), 1–32.

Bisbee, J., R. Dehejia, C. Pop-Eleches, and C. Samii (2017). Local instruments, global extrapolation: External validity of the labor supplyâfertility local average treatment effect. *Journal of Labor Economics 35*(S1), S99–S147.

Chetty, R. and N. Hendren (2018, 02). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *The Quarterly Journal of Economics 133*(3), 1107–1162.

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

Chyn, E. (2018, October). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review 108*(10), 3028–56.

Chyn, E. and L. F. Katz (2021). Neighborhoods matter: Assessing the evidence for place effects. *Journal of Economic Perspectives 35*(4), 197–222.

del Pero, A., W. Adema, V. Ferraro, and V. Frey (2016). Policies to promote access to good-quality affordable housing in oecd countries.

Dix-Carneiro, R., R. R. Soares, and G. Ulyssea (2018, October). Economic shocks and crime: Evidence from the brazilian trade liberalization. *American Economic Journal: Applied Economics 10*(4), 158–95.

Freedman, M., E. Owens, and S. Bohn (2018, May). Immigration, employment opportunities, and criminal behavior. *American Economic Journal: Economic Policy 10*(2), 117–51.

Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Grogger, J. (1997). Local violence and educational attainment. *The Journal of Human Resources 32*(4), 659–682.

Grogger, J. (1998). Market wages and youth crime. *Journal of Labor Economics 16*(4), 756–791.

Huang, C.-C., D. Laing, and P. Wang (2004). Crime and poverty: A search-theoretic approach*. *International Economic Review 45*(3), 909–938.

Hwang, Y. (2022). Bounding omitted variable bias using auxiliary data with an application to estimate neighborhood effects. *working paper*.

Inoue, A. and G. Solon (2010, 08). Two-Sample Instrumental Variables Estimators. *The Review of Economics and Statistics 92*(3), 557–561.

Jacob, B. A. and J. Ludwig (2012, February). The effects of housing assistance on labor supply: Evidence from a voucher lottery. *American Economic Review 102*(1), 272–304.

Kain, J. F. (1968, 05). Housing Segregation, Negro Employment, and Metropolitan Decentralization*. *The Quarterly Journal of Economics 82*(2), 175–197.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Krupka, D. J. and K. Donaldson (2007). Wages, rents and heterogeneous moving costs.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more t in experiments. *Journal of Development Economics 99*(2), 210–221.

Mogstad, M. and A. Torgovitsky (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics 10*(1), 577–613.

Moretti, E. (2010). Local labor markets. *NBER working papers* (15947).

Nakamura, E., J. Sigurdsson, and J. Steinsson (2021). The gift of moving: Intergenerational consequences of a mobility shock. *Review of Economics Studies* (forthcoming).

Pacheco, T. (2019). PolÃtica habitacional e a oferta de trabalho : EvidÃªncias de sorteios do minha casa, minha vida. *working paper*.

Pacini, D. and F. Windmeijer (2016). Robust inference for the two-sample 2sls estimator. *Economics Letters 146*, 50–54.

Picard, P. M. and Y. Zenou (2018). Urban spatial structure, employment and social ties. *Journal of Urban Economics 104*, 77–93.

Roback, J. (1988). Wages, rents, and amenities: Differences among workers and regions. *Economic Inquiry 26*(1), 23–41.

Rocha, G. M. (2018). PolÃtica habitacional e a oferta de trabalho : EvidÃªncias de sorteios do minha casa, minha vida. *working paper*.

Rubin, D. B. (1981). The Bayesian Bootstrap. *The Annals of Statistics 9*(1), 130 – 134.

Topa, G. and Y. Zenou (2015). Neighborhood and network effects. In *Handbook of Regional and Urban Economics*, Volume 5, Chapter 9, pp. 561–624.

Van Djik, W. (2019). The socio-economic consequences of housing assistance. *working paper*.

Zenou, Y. (2013). Spatial versus social mismatch. *Journal of Urban Economics 74*, 113–132.

APPENDIX

A Additional Data Description

In this Appendix, we provide additional details on the data used in the paper. Appendix A1 describes auxiliary data sources. Appendix A2 provides a detailed description of the variable definitions. Appendix A3 shows additional descriptive statistics for neighborhoods.

A.1 Additional data description

In the main text, we provide a detailed description of four main data sources: the *MCMV* lotteries, the matched employer-employee data (RAIS), the Single Registry, and the Census Data. This section of the Appendix provides additional details on other data sources used in the paper. We describe three additional data sources: city hall records, the Rio de Janeiro Public Security Institute (ISP-RJ) statistics, and sites that specialize in selling houses.

First, the Rio de Janeiro City Hall maintains registries of public facilities operating in the city and their respective addresses. We collected data for daycare facilities, parks, schools, and health facilities. Data for health facilities include hospitals and emergency care units. Not all of these facilities are run by the municipal government, as some of these facilities fall under the state government's responsibility. We georeference all of these facilities and associate them to the census tract they belong to.

The ISP-RJ is an autarchy linked to the Rio de Janeiro State Public Security Secretary. It is responsible for collecting data and performing statistical analysis to inform public policies to prevent crime. The ISP-RJ collects monthly data on crimes provided by police stations and the geographical location of these crimes and publicizes these statistics. We aggregated crimes for auto thefts, robberies, rapes, kidnappings, and murders for all months from February of 2015 to December of 2017, which is the same endline period we use for the labor market results, and associate each crime with its respective neighborhood.

Finally, we also collect data for houses and apartments being sold in Rio de Janeiro. We scrape data for the most common websites used to sell houses in Brazil: https://www.zapimoveis.com.br/, https://www.loft.com.br, and https://www.vivareal.com.br/. We collected the data for house prices in February of 2022.

We keep in the sample only houses in the same neighborhoods as *MCMV* housing projects and drop duplicate housed being announced in more than one website. We also collect information for characteristics that are typically advertised alongside the house, such as total area, number of rooms, number of bathrooms, etc. Also, despite being prohibited by program rules, a small number of individuals still sold houses in *MCMV* housing projects. We can identify those houses either because, in some cases, the homeowner explicitly advertises the house as being originated in the *MCMV* program or by using a combination of address and house characteristics.

A.2 Additional variable description

In Table A.1, we summarize all variables used in the main results of the paper. Some of these variables are too complex to be described in the Table, so we discuss them below.

Table A.1: Additional variables' description

Variable	Source	Reference	Description
Neighborhood income	Census	Table 1, 3	Percentiles of household per capita income adjusted by sample weights
Neighborhood wages	Census	Table 1	Percentiles for wages in all job types for adult-age population adjusted by sample weights
Neighborhood education	Census	Table 1	Percentiles for education of neighborhood residents. Education is a five-level index for the highest level of education achieved: 1) no education, 2) up to middle school; 3) up to highschool, 4) up to college, 5) post-graduate.
Single-parent households in Neighborhoods	Census	Table 1	Percentiles of single-parent households
Formal Jobs	RAIS	Table 1, 3	Percentiles for the total number of formal jobs in firms located in each neighborhood
Market access	RAIS and SR	Table 1 and Figure 3, 4	see separate description
Amenities	city hall records	Table 1 and Figure 3, 4	see separate description
Crime rates	ISP-RJ	Table 1 and Figure 3, 4	see separate description
Treated	Lottery	Table 2	dummy for drafted individuals that do not participate in other *MCMV* lotteries in the future
White	RAIS and SR	Table 2, 6	dummy for individuals that report being white
Male	RAIS and SR	Table 2, 6	dummy for individuals that report being male
Disability	RAIS and SR	Table 2	dummy for individuals that report having some disability
Education	RAIS	Table 2, 6	Five-level index for the highest level of education achieved: 1) no education, 2) up to middle school; 3) up to highschool, 4) up to college, 5) post-graduate.
Ever employed	RAIS	Table 2, 4, 6	dummy formally employed between 2002 and 2014
Hours	RAIS	Table 2, 4, 5, 6	Weekly number of hours registered in formal contract
Tenure	RAIS	Table 2, 6	Number of months since the worker was hired in the current contract
Distance to previous job (km)	RAIS	Table 2	Number of kilometers between housing project the individual was drafted to and the last job site before the lottery. For the control group, we choose the median distance across all different housing projects
Last employment year before 2015	RAIS	Table 2, 6	Last year individual held a formal job in the baseline period
Number of neighborhood residents	SR	Figure 1	Total number of individuals that were drafted and report living one neighborhood postal-code (either before of after the lottery)
Lived in a MCMV neighborhood	SR	Table 3	dummy for individuals that reported living in a postal-code inside a *MCMV* neighborhood
Number of rooms	SR	Table 3	Total number of rooms individuals report having in their house
Number of bedrooms	SR	Table 3	Total number of bedrooms individuals report having in their house
Rent (R$)	SR	Table 3	Monthly rent expenses measured in brazilian reais. Values are deflated by the Price Index for the Wide Consumer (IPCA) for January of 2021
Transportation expenses (R$)	SR	Table 3	Monthly expenses on all types of transportation measured in brazilian reais. Values are deflated by the Price Index for the Wide Consumer (IPCA) for January of 2021. Variable is only available for 2020 updates
Formal employment	RAIS	Table 4, 5, 6	dummy for working at any point in a year
Month employment hours	RAIS	Table 4, 5	number of months the worker was employed during each year
log(wages)	RAIS	Table 4	natural logarithm of average wages during the year only for employed workers
House prices	selling sites	Table 7	prices of houses being offered during February 2022 in any of the *MCMV* neighborhoods
Number of rooms	selling sites	Table 7	number of rooms being advertised alongside the house being sold
Number of bathrooms	selling sites	Table 7	number of bathrooms being advertised alongside the house being sold
Area	selling sites	Table 7	total area being advertised alongside the house being sold
Garage	selling sites	Table 7	dummy for a garage spot being advertised alongside the house being sold
Lobby	selling sites	Table 7	dummy for the presence of a lobby being advertised alongside the house being sold

Now, we discuss three other variables that could not be easily described in the Table above: measures of labor market access, amenities, and crime rates.

For the former, we want to measure how close or distant average individuals in a certain neighborhood are from job opportunities. In order to do that, we combine information from the Single Registry with data for RAIS.

First, we georeference all firms that formally employ individuals in Rio de Janeiro city, count the number of low-skill employed individuals, and associate the number of workers at each postal code. Next, we georeference all low-skill individuals living in each neighborhood and, similarly, associate them to the respective postal code. In both cases, we define low-skill workers as individuals with high school or less.

Let h be the location of a housing site, j_c be the location of a establishment that employs low-skill workers, $d(.,.)$ be a distance function, N_h be total number of housing sites, r_h be the number of residents in h, and w_c be the total number of workers in c. Then, we calculate the market access for neighborhood n as:

$$M_n = \sum_c \sum_{h \in n} \frac{d(h, j_c) * w_{hc}}{N_h}$$

where:

$$w_{hc} = \frac{r_h + w_c}{\sum_c \sum_{h \in n} (r_h + w_c)}$$

Our measure of market access is the weighted average distance between residences and jobs. We decide to concentrate on workers with high school or less because this restriction encompasses almost all individuals in the lotteries sample, and the geographical distribution of jobs and residences can be very different from the population as a whole. That is why we opted to use Single Registry data instead of the Census data since in the latter; we do not have available socioeconomic characteristics for the whole sample. Instead of considering a less coarse geographical measure, we also decided to associate jobs and residents to the respective postal code to reduce the dimensionality of the distance function calculation.

For the other two measures of amenities and crimes, we rely on a principal component analysis. Consider that we want to observe a latent neighborhood characteristic and that we have m measurements for this latent variable stacked in vectors: $X = (v_1, ...v_m)$. We standardize these measurements so that they have zero mean. Then, we compute the covariance matrix of these correlations and their respective eigenvectors and eigenvalues. Finally, we keep the first principal component for each variable of interest.

For the crime variable, we choose five different measurements: auto thefts, thefts, rapes, kidnappings, and murders. For amenities, we have four different measurements: public parks, schools, daycare facilities, and health facilities. The first principal component explains about 60% of the total variance in crime measurements and about 50% of the total variance in the measurements of amenities. Then, we standardize these principal components as percentiles relative to the Rio de Janeiro neighborhood distribution.

A.3 Additional descriptive statistics

In Table A.2, we show additional descriptive statistics for other variables at the neighborhood level. In Table A.3, we compare the characteristics of the *MCMV* neighborhoods relative to other Rio de Janeiro neighborhoods.

Table A.2: Descriptive statistics for additional variables and all Rio de Janeiro neighborhoods

Variable	Average
Average distance (km)	32.43
Number of parks	0.37
Number of daycare facilities	1.05
Number of hospitals	0.28
Number of schools	0.98
Average auto thefts/per year	200.22
Average thefts/per year	2216.12
Average rapes/per year	15.12
Average murders/per year	12.14
Average kidnappings/per year	0.05

Note: This table shows averages for several different variables. The average distance is based on RAIS and Single Registry data, the number of public facilities comes from city hall records, data for crimes is based on ISP records. For all variables, we show data for all Rio de Janeiro neighborhoods.

Table A.3: Comparison of *MCMV* neighborhoods with other Rio de Janeiro neighborhoods

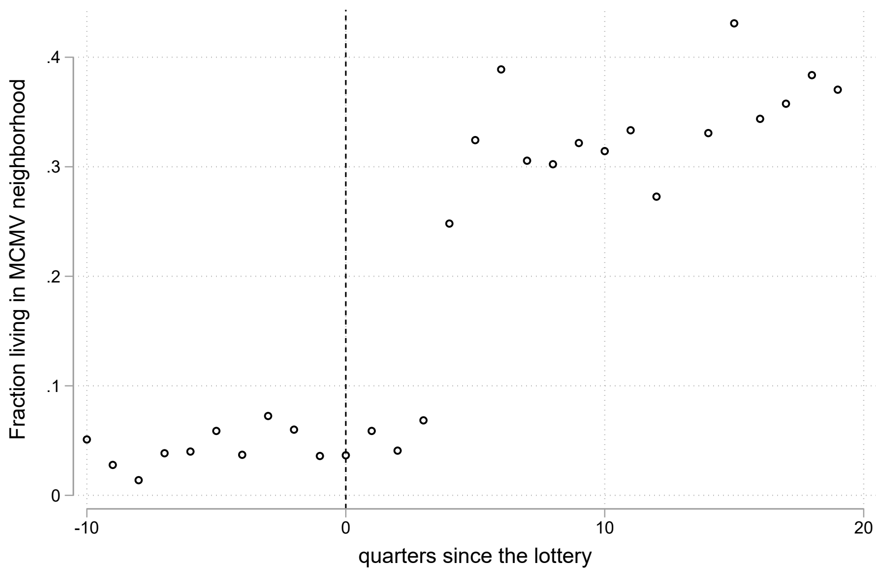
	MCMV neighborhoods	All other neighborhoods
Age	31.92	33.67
White	0.34	0.47
Education	1.83	1.89
Literacy	0.89	0.88
Employed	0.48	0.48
Formally employed	0.24	0.23
Wages (R$)	935.15	1392.48
Single parent household	0.79	0.81
Number of individuals in the household	3.63	3.65
Rent	288.42	445.60
Water supply	0.95	0.99
Owns a car	0.31	0.38

Note: This Table shows averages for several variables in individuals living in a *MCMV* neighborhood (column 1) or any other Rio de Janeiro neighborhood (column 2). All variables come from the Census data. Age contemplates all individuals, including children, white is a dummy for individuals that report being white, employed is a dummy for individuals working in the week of the Census survey, formally employed is a dummy for individuals working in the week of the Census survey, wages are nominal wages measured in R$ of 2010. Single parent households is a dummy for families with children but only one parent, number of individuals in the household measures the number of components in the family, and rent is the monthly amount paid by households that rent a house. Water supply is a dummy for households that have access to potable water, and owns a car is a dummy for households owning at least one car.

B Additional Balancing Tests and First-Stage Results

In this Appendix, we show additional balancing and first-stage results. Figure B.1 shows that drafted individuals started moving to the housing projects almost a year after the lottery, and the fraction that lived there remained relatively unchanged until the end of 2020. Table B.1 shows that the allocation process was strict, and being drafted to a specific neighborhood increases only the probability that the drafted individuals were living in that neighborhood. Table B.2 shows that not only the drafted group is balanced relative to the control group, but also that individuals drafted to each neighborhood have similar characteristics to each other. Finally, in Table B.3 we show that the lottery was not only successfully balanced participant characteristics but also for their family members. For this last sample, we only have Single Registry characteristics available.

Figure B.1: Timing of changes to housing project for drafted individuals



Note: This Figure shows average fraction of drafted individuals in the sample of families living in a *MCMV* neighborhood by quarter.

Table B.1: Neighborhood-specific moving patterns

	Lived in neighborhood 1	Lived in neighborhood 2	Lived in neighborhood 3
Drafted to neighborhood 1	0.104***	-0.003	-0.004
	(0.015)	(0.004)	(0.003)
Drafted to neighborhood 2	0.002	0.220***	0.004
	(0.004)	(0.015)	(0.004)
Drafted to neighborhood 3	0.001	0.002	0.292***
	(0.002)	(0.002)	(0.010)
Observations		373,903	

Note: This table presents regressions of dummies for living in one of each of the *MCMV* neighborhoods on dummies that indicate being drafted to specific neighborhoods. All results are shown for the families' sample. We show robust standard-errors in parenthesis. * p<0.1, ** p<0.05, and *** p<0.01.

Table B.2: Lottery-specific balancing tests

	Neighborhood 1	Neighborhood 2	Neighborhood 3	Joint p-value for differences
Panel A: socioeconomic characteristics				
White	0.38	0.42	0.37	0.23
Male	0.45	0.42	0.45	0.54
Schooling	3.79	3.87	3.81	0.18
Panel B: labor market characteristics				
Ever employed	1	1	1	
Hours	42.00	41.59	41.90	0.35
Wages	1236.94	1238.29	1322.34	0.54
Tenure	50.99	48.84	48.59	0.81
Last employment year before 2015	2012.65	2012.92	2012.85	0.16
Joint Hypothesis (p-value)		0.97		

Note: This table presents variable means for individuals drafted to each neighborhood (Columns 1 to 3). In column 4, we present the p-value for the joint test for the null hypothesis that means are equal across groups. All columns show data for the labor market sample. White and Males are dummies for individuals' characteristics. Schooling is a five-level index indicating the highest level of instruction that ranges from no schooling to post-graduation. Ever employed is a dummy for individuals that held a formal job in any year between 2002 and 2014. Hours shows the number of monthly hours registered in contract in the last year individuals' held a formal job. Wages show average monthly compensation registered in contract for the last year individuals appeared in RAIS. Tenure is the number of months since the individual was hired. The last employment year before 2015 is the last baseline year the individuals held a formal job. In all columns, characteristics for the labor market sample come from RAIS. The joint test line shows the F statistic and respective p-value for the joint test that all means are equal across groups and variables. * p<0.1, ** p<0.05, and *** p<0.01.

Table B.3: Descriptive statistics and balancing tests for families

	Sample of families	
	Control mean	Drafted group difference
Treated	0	0.646***
		(0.007)
White	0.23	-0.002
		(0.007)
Male	0.25	-0.002
		(0.007)
Schooling	2.167	0.032
		(0.025)
Disability	0.156	0.006
		(0.005)
Joint F test (p-value)	0.32	
Observations	1,922,721	

Note: This table presents variable means for individuals that were not drafted in the lottery and their family members (column 1) as well as the differences for the drafted group (column 2). All columns show data for the families' sample. Treated is a dummy for households where the drafted beneficiary took-up treatment. White and Males, and Disability are dummies for individuals' characteristics. Schooling is a five-level index indicating the highest level of instruction that ranges from no schooling to post-graduation. In all columns, characteristics come from the Single Registry. The joint test line shows the F statistic and respective p-value for the joint test that all difference in means are equal to zero (except for the treated variable). * p<0.1, ** p<0.05, and *** p<0.01.

C Additional Results

In this section of the Appendix, we present additional results. Table C.1 shows that the program did not significantly alter the beneficiaries' informal employment but reduced the probability that individuals received a transfer from the largest Brazilian cash transfer by two percentage points. Table C.2 show that informal employment outcomes are not significantly affected by the neighborhood individual were drafted to. Table C.3 shows the dynamic formal employment response to the *MCMV* house. We find negative effects for all years, but larger decreases in employment probability are concentrated on year one and year three after the lottery. In this specification, we have less power than in the pooled sample, so we cannot reject the null hypothesis that most of these differences are equal to zero. Table C.4 shows the effects of the *MCMV* program on formal employment probability separately for 1) man and woman; 2) white and non-white individuals, and; 3) High and Low education individuals. We find that the program's effects are concentrated on high education and white individuals and that there are no significant differential impacts by gender.

In Table C.5, we show alternative estimates for the pooled sample. In column (1), we control for formal employment in the baseline period similar to an ANCOVA specification. We find that estimated treatment effects are slightly smaller, and the respective standard error is slightly higher than in our main estimates, which is expected since the outcome variable is strongly serially correlated (McKenzie, 2012). In column (2), we show standard difference-in-difference estimates much closer to the pooled estimates shown in the paper. Finally, in Table C.6, we show that the lottery did not significantly affect labor market outcomes of family members. Table C shows that the *MCMV* program does not affect employment outcomes for individuals that are not formally employed at any point before the lottery (and are not included in the labor market sample).

Table C.1: Effects of receiving a MCMV house on other outcomes

	Control mean	ITT	TOT	Number of observations
Informal employment	0.305	-0.002	-0.003	373,903
		(0.006)	(0.008)	
Number of months employed informally	2.692	0.054	0.082	373,903
		(0.055)	(0.082)	
Bolsa FamÃlia benficiary	0.24	-0.014***	-0.021***	373,903
		(0.006)	(0.009)	

Note: This table presents the mean for the control group (column 1), ITT treatment effects of receiving a house of the *MCMV* program on several variables (column 2), TOT treatment effects (column 3), and the number of observations used in each regression (column 4). Results include all updates in the families sample. Informal employment is a dummy for individuals that reported working without *carteira asssinada*. Number of months employed informally is the total number of months in each year individuals were informally employed (0 if not employed or formally employed). Bolsa FamÃlia beneficiary is a dummy indicating that the household received any kind of Bolsa FamÃlia transfer. Clustered standard-errors at the individual-level are shown in parenthesis. * p<0.1, ** p<0.05, and *** p<0.01.

Table C.2: Heterogeneous treatment effects on informal labor market outcomes by neighborhood

	Employment	# of months
Neighborhood 1	-0.007	-0.125
	(0.031)	(0.312)
Neighborhood 2	0.009	-0.015
	(0.026)	(0.260)
Neighborhood 3	-0.015	0.042
	(0.015)	(0.162)
Equality of coefficients (p-value)	0.76	0.97
Control mean	0.31	2.77
Observations	640,571	

Note: This table presents neighborhood-specific treatment effects of receiving a house of the *MCMV* program on informal employment. Results are shown for the families' sample. Observations are stacked for the years of 2015, to 2020. Informal employment is a dummy for individuals reporting to be informally employed in the Single Registry. Number of months employed is the total number of months in each year individuals were employed (0 if not employed on the informal market). We include controls for age, race, and schooling. We also show an F test for the joint equality of treatment effects across neighborhoods. Clustered standard errors at the individual level are shown in parenthesis: * p<0.1, ** p<0.05, and *** p<0.01.

Table C.3: Dynamic treatment effects of receiving a *MCMV* house

	ITT	TOT	Control mean
One year after lottery (2015)	-0.013	-0.025	0.68
	(0.011)	(0.020)	
Two years after lottery (2016)	-0.007	-0.014	0.63
	(0.008)	(0.015)	
Three years after lottery (2017)	-0.012*	-0.022*	0.57
	(0.007)	(0.013)	
Observations	503,880		

Note: This table presents ITT treatment effects of receiving a house of the *MCMV* program on several variables (column 1), TOT treatment effects (column 2), and the control group mean (column 3). Results are shown for the labor market sample. Formal employment is a dummy for participating in the formal labor market at any point in each year. Each line represents the treatment effects for a certain year relative to the lottery. We include controls for age, race and schooling. Clustered standard-errors at the individual-level are shown in parenthesis. * p<0.1, ** p<0.05, and *** p<0.01.

Table C.4: Heterogeneous treatment effects of the MCMV on formal employment

	(1) Control mean	(2) ITT	(3) TOT	(4) Number of observations
Woman	0.60	-0.009	-0.017	828,489
		(0.007)	(0.014)	
Man	0.68	-0.009	-0.018	683,142
		(0.008)	(0.014)	
Non-white	0.60	-0.008	-0.017	568,488
		(0.007)	(0.014)	
White	0.66	-0.019***	-0.037***	943,143
		(0.007)	(0.012)	
Low education	0.54	0.001	0.002	407,190
		(0.004)	(0.008)	
High education	0.67	-0.019**	-0.037**	1,104,450
		(0.006)	(0.011)	

Note: This table presents the mean formal employment for the control group (column 1), ITT treatment effects of receiving a house of the *MCMV* program on several variables (column 2), TOT treatment effects (column 3), and the number of observations in each regression (column 4). Results are shown for the labor market sample. Formal employment is a dummy for participating in the formal labor market at any point in each year. Each line shows results for a different group: males, females, non-white individuals, white individuals, low education, and high education. Low education is defined as not have attending high school and high education is defined as at least attending highschool. We include controls for age, and schooling. Clustered standard-errors at the individual-level are shown in parenthesis. * p<0.1, ** p<0.05, and *** p<0.01.

Table C.5: Effects of MCMV house on formal employment with alternative specification

	ITT	
	(1) Baseline control	(2) Difference-in-difference
Receiving a house	0.008	-0.009*
	(0.006)	(0.004)
Formal employment	0.951***	
	(0.002)	
Control mean	0.63	
Number of observations	1,511,631	

This table presents two additional ITT estimates for the effects of being drafted to the *MCMV* program on employment outcomes. The first column controls for age, race, schooling, and a dummy for being formally employed at any point in 2014. The second column does not control for previous formal employment, but we control for dummies for drafted individuals and post-treatment period and report results for the interaction between drafted and post-treatment dummies. Results are shown for the labor market sample. Observations are stacked for the years of 2015, 2016, and 2017. Clustered standard-errors at the individual-level are shown in parenthesis.

Table C.6: Effects of receiving a MCMV house on employment outcomes of beneficiaries' family members

	Control mean	ITT	TOT	Number of observations
Formal employment	0.45	-0.007	-0.010	1,922,721
		(0.007)	(0.010)	
Number of months employed	4.32	-0.010	-0.156	1,922,721
		(0.098)	(0.144)	

Note: This table presents the mean for the control group (column 1), ITT treatment effects of receiving a house of the *MCMV* program on several variables (column 2), TOT treatment effects (column 3), and the number of observations used in each regression (column 4). We include in the sample all family members of lottery participants that had any update in the post-lottery period. We also restrict the sample to those individuals with available individual identifying information (CPF) on the Single Registry. Formal employment is a dummy for participating in the formal labor market at any point in each year. Number of months employed is the total number of months in each year individuals were employed (0 if not employed). We use RAIS administrative variables as outcomes of interest. Clustered standard-errors at the individual-level are shown in parenthesis. * $p<0.1$, ** $p<0.05$, and *** $p<0.01$.

Table C.7: Effects on formal employment outcomes for individuals that were never employed

	Control mean	ITT	TOT	Number of observations
Formal employment	0.03	0.005	0.010	648,684
		(0.005)	(0.009)	
Monthly hours	1.47	-0.053	-0.102	648,684
		(0.262)	(0.504)	
Number of months employed	0.19	0.038	0.074	648,684
		(0.037)	(0.070)	
Log(wages)	7.17	-0.037	-0.058	13,040
		(0.052)	(0.081)	

Note: This table presents the mean for the control group (column 1), ITT (column 2), and TOT (column 3) estimates. Results are shown for individuals that are not formally employed at any point between 2003 and 2014. Observations are stacked for the years of 2015, 2016, and 2017. Formal employment is a dummy for being formally employed at any point in each year. The variable monthly hours measures the number of contractual hours of individuals formally employed (0 if the individual was not employed). Number of months employed is the total number of months in each year individuals were employed (0 if not employed). Log(wages) is the natural logarithm of the average wages individuals received during the year while employed (individuals that were not formally employed in a certain year were dropped from the analysis). Wages are reported as nominal R$. We transform this variable in real values using a broad Consumer Price Index (IPCA) for January of 2021. We include controls for age, race and schooling. Clustered standard-errors at the individual-level are shown in parenthesis: * p<0.1, ** p<0.05, and *** p<0.01.

D ExtrapoLATE

In section 6 of the main text, we showed that individuals drafted to different neighborhoods experienced very different impacts of the program. We argued this was due to different neighborhood characteristics. However, one alternative explanation is that these results are driven by a different set of individuals accepting to go to each neighborhood.

We already argued in Table 6 that the characteristics of different groups of compliers are similar. Recently, several different papers discussed how to assess this question more formally and extrapolate IV estimates to different populations, such as Bisbee et al. (2017) and Mogstad and Torgovitsky (2018). In this section of the Appendix, we use the model suggested by Angrist and FernÃ¡ndez-Val (2010) to formally show that differences in the characteristics of the compliers cannot explain differences in neighborhood-specific treatment effects.

Appendix D1: Extrapolation model

Let y_i^0 and y_i^1 be the formal employment potential outcomes for individual i. Also, let D_i^n be a dummy for individual drafted to neighborhood n, and H_i^n a dummy for individuals that accepted the house in neighborhood n. Besides the usual instrumental variable hypothesis (exclusion restriction and monotonicity), we make two important assumptions:

Assumption 1: $\mathbb{E}[y_i^1 - y_i^0 | D_i^n, x] = \mathbb{E}[y_i^1 - y_i^0 | x], \quad n = 1, 2, 3.$

Assumption 2: For a finite set, $\mathcal{X} = \{x_1, ..., x_K\}, P[x \in \mathcal{X}] = 1$.

The first assumption is the conditional effect ignorability of the instrument (CEI). This assumption states that the treatment effect heterogeneity in the causal effects is entirely due to differences in observable differences in compliers in characteristics (x). This hypothesis is similar to conditional independence assumptions in matching estimators. Since we are interested in extrapolating the neighborhood-specific treatment effects to another complier population with different observable characteristics, this is a natural assumption. The second assumption is that all covariates of interest are discrete. This second hypothesis is not necessary for identification, but it significantly eases estimation.

Now, let:

$$\Delta^n(x) = \mathbb{E}[y_i^1 - y_i^0 | \mathrm{x} = x, D_i^n = 1]$$

be the causal effect of receiving a *MCMV* house on employment for compliers with characteristics x.

Then, note that we can write the local average treatment effect (LATE) for one of the instruments and the whole sample as:

$$\Delta^n = \mathbb{E}[y_i^1 - y_i^0 | D_i^n = 1] = \mathbb{E}\big[\mathbb{E}[y_i^1 - y_i^0 | D_i^n = 1, \mathrm{x} = x] | D_i^n = 1\big]$$

where the equality follows from the law of iterated expectations. Then, by the

CEI assumption, we can write:

$$\mathbb{E}\big[\mathbb{E}[y_i^1 - y_i^0|D_i^n = 1, \mathrm{x} = x]|D_i^n = 1\big] = \mathbb{E}\big[\mathbb{E}[y_i^1 - y_i^0|\mathrm{x} = x]|D_i^n = 1\big]$$

Using the definition of $\Delta^n(x)$, we have that:

$$\mathbb{E}\big[\mathbb{E}[y_i^1 - y_i^0|\mathrm{x} = x]|D_i^n = 1\big] = \mathbb{E}\big[\Delta^n(x)|D_i^n = 1\big]$$

Finally, by using the Bayes rule, we can write:

$$\mathbb{E}\big[\Delta^n(x)|D_i^n = 1\big] = \sum_{x=\mathcal{X}} \Delta^n(x)\omega^n(x)$$

where:

$$\omega^n(x) = \frac{\mathbb{E}[H_i^n|D_i^n = 1, x]}{\mathbb{E}[H_i^n|D_i^n = 1]}$$

That is, under the CEI assumption, we can write the LATE for instrument D_i^n as a weighted average of LATE for each value of x where weights are given by the size of the first-stage for $\mathrm{x} = x$, relative to the size of the first-stage for the whole sample.

Now, lets use the decomposition above to compare the LATE two instruments. To fix ideas, lets compare LATE for neighborhoods 1 and 3. In Table 5, we show that the LATE for neighborhood 1 is negative and very large, while for neighborhood 3 is close to zero. Note that by using the decomposition above:

$$\begin{aligned}\Delta^3 - \Delta^1 = \sum_{x=\mathcal{X}} \Delta^3(x)\omega^3(x) - \sum_{x=\mathcal{X}} \Delta^1(x)\omega^1(x) \implies \\ \implies \Delta^3 - \Delta^1 = \sum_{x=\mathcal{X}} \big[\Delta^3(x) - \Delta^1(x)\big]\omega^3(x) + \\ + \sum_{x=\mathcal{X}} \Delta^1(x) * \big[\omega^3(x) - \omega^1(x)\big]\end{aligned}$$

The difference between the two LATE can be decomposed into two terms. The first one reflects differences in neighborhood-specific treatment effects. The second term reflects differences in compliers characteristics. Therefore, we can derive the following condition:

$$\Delta^3(x) = \Delta^1(x) \iff \sum_{x=\mathcal{X}} \Delta^1(x) * \big[\omega^3(x) - \omega^1(x)\big] = 0$$

It is straightforward to test the validity of this condition. We can compare the LATE for neighborhood 1 (Δ^1) with $\sum_{x=\mathcal{X}} \Delta^1(x)\omega^3(x)$, which is the an average of the estimated treatment effects of being drafted to neighborhood 1, weighted by the complier population of neighborhood 3. If re-weighting Δ^1 by the characteristics of neighborhood 3 complier population is enough to bridge the difference to Δ^3, then all differences should be explained by the different populations. However, if this is not the case, and Δ^1 and Δ^3 remain different after the re-weighting process, then neighborhood-specific treatment effects should be different.

Appendix D2: Estimation and inference

The assumption of discrete covariates eases a lot the estimation process. We divide our sample into cells for each possible value of $x \in \mathcal{X}$. Then, we estimate $\hat{\Delta}^n(x)$ using a Wald estimator for each cell. Similarly, we estimate $\hat{\omega}^n(x)$ as the take-up rates in cell x, divided by the take-up rate in the whole sample.

We include five relevant variables in the estimation process. They are dummies for: males, white individuals, older than the median of the sample individuals, individuals with completed highschool, and workers employed in firms that were larger than the median of the sample. The combination between these five variables gives us thirty-two different cells. We decide to use only five variables and to discretize some of them (education, age, and firm size) because by making cells even less coarse would significantly reduce the variation in instruments and take-up rates for a relevant fraction of those cells, preventing us to estimate LATE.

The inference process is more complicated than estimation. Angrist and FernÃ¡ndez-Val (2010) suggest that we can estimate asymptotic distributions analytically using a GMM-type estimator or approximate them numerically by bootstrap procedures. We opt for the latter.

However, standard bootstrapping procedures are not well suited for this estimation. As mentioned above, we divided our sample in lots of discrete mutually exclusive cells and there is little instrument variation in some of them. This problem is aggravated by the resampling procedure of the traditional bootstrap because for some realizations of resampling, we would not have enough drafted and treated observations to identify the estimates of $\Delta^n(x)$ and $\omega^n(x)$ for some cells. This would imbalance the sample of cells being used across different bootstrapped samples and generate practical problems in the estimation algorithm.

Instead, we implemented a Bayesian bootstrap (Rubin, 1981). We follow these steps:

1. We draw a vector of N random draws for a gamma distribution with parameters $\Gamma(1,1)$. Then, we associate each draw to a different observation in our sample and normalize these draws by their sum: $w_i = \frac{d_i}{\sum_i^N di}$.
2. For each cell, we estimate $\Delta_1^n(x)$ and $\omega_1^n(x)$ using w_i as weights for the observations. Then, we calculate reweighted LATE as: $\Delta_1^1(x) * \omega_1^3(x)$ and $\Delta_1^2(x) * \omega_1^3(x)$.
3. We repeat the procedure 50 times and collect the vector of reweighted LATE for each replication: $[\Delta_1^n(x) * \omega_1^3(x), ..., \Delta_{50}^n(x) * \omega_{50}^3(x)], \quad n = 1, 2$.
4. Finally, we estimate the mean and standard-deviation from the vector of coefficients that we collect in the first-step.

The advantage of the Bayesian bootstrap is that we never draw a a zero weight for any observation in any replication. Therefore, we can guarantee that we are able to maintain our sample fixed and we always have enough treatment variation to allow the estimation of within-cell first-stage and treatment effects.

Appendix D3: Extrapolation results

In Table D, we show in Panel A the LATE of being drafted by to neighborhoods 1 to 3, identically, to Table 5. This is equivalent to estimating the LATE for each cell and re-weighting them by their own complier population ($\omega^n(x)$). In Panel B, we show weight all LATE by the complier population of neighborhood 3. We also test the hypothesis that the re-weighted LATE are equal to the original LATE or if the re-weighted LATE for neighborhoods 1 and 2 are equal to the one in neighborhood 3.

Table D.1: Original and re-weighted neighborhood-specific treatment effects on employment

	Neighborhood 1	Neighborhood 2	Neighborhood 3
Estimate	$\Delta^1(x)$	$\Delta^2(x)$	$\Delta^3(x)$
Weights	$\omega^1(x)$	$\omega^2(x)$	$\omega^3(x)$
Panel A: Original LATE			
	-0.066***	-0.074***	0.012
	(0.024)	(0.024)	(0.014)
Estimate	$\Delta^1(x)$	$\Delta^2(x)$	$\Delta^3(x)$
Weights	$\omega^3(x)$	$\omega^3(x)$	$\omega^3(x)$
Panel B: Re-weighted LATE			
	-0.085***	-0.081***	0.012
	(0.008)	(0.003)	(0.014)
$P\big(\hat{\Delta}^n = \Delta^n(x)\omega^3(x)\big)$	0.44	0.00	
$P\big(\hat{\Delta}^3 = \Delta^n(x)\omega^3(x)\big)$	0.00	0.00	

We can see that once we weight LATE for neighborhoods 1 and 2 by the compliers that took-up treatment in neighborhood 3, the gap in neighborhood-specific treatment effects becomes even larger than in the original estimates. For neighborhood 1, we cannot reject the hypothesis that the re-weighted LATE is equal to the original LATE and we can reject the hypothesis that it is equal to the LATE for neighborhood 3. For neighborhood 2, on the other hand, we can reject both the hypothesis that the re-weighted LATE is equal to the original one and that is equal to the LATE for neighborhood 1.

E Additional Details on the Mechanisms Analysis

In this section of the Appendix, we provide additional details used in the main text for the mechanism analysis.

Appendix E1: Obtaining bounds for β^m

Let $\boldsymbol{\theta}$ be the vector of treatment effects of the program on employment and $Cone(v_1, ..., v_n)$ be the set that contains all positive linear combinations of the vectors $v_1, ..., v_n$. Then, as in the main paper, we can write:

$$\boldsymbol{\theta} = \beta^m \begin{bmatrix} \Delta MA_1 \\ \Delta MA_2 \end{bmatrix} + \beta^n \begin{bmatrix} \Delta NQ_1 \\ \Delta NQ_2 \end{bmatrix} +$$
$$+\beta^a \begin{bmatrix} \Delta Am_1 \\ \Delta Am_2 \end{bmatrix} + \beta^c * \begin{bmatrix} \Delta Cr_1 \\ \Delta Cr_2 \end{bmatrix}$$

We are ultimately interested in the fraction of employment effects explained by labor market access. That is:

$$f^m = \frac{\beta^m * (\Delta MA_1 + \Delta MA_2)}{\theta_1 + \theta_2}$$

Note that f is strictly increasing in β^m. Therefore, we can provide bounds for f by deriving bounds for β^m. To do that, we make some simplifying assumptions that restrict population parameters (mechanisms and effects on employment). All of them translate the assumption that population parameters are not too far away from the estimates in Figure XX of the main text.

Assumption 1: $\boldsymbol{\theta} \in Cone(\Delta MA, \Delta NQ, \Delta Am, \Delta Cr)$

Assumption 2: $\boldsymbol{\theta} \notin Cone(\Delta NQ, \Delta Cr, \Delta Am)$

Assumption 3: $Cone(\Delta MA, \Delta Cr, \Delta Am) = Cone(\Delta MA, \Delta Am)$

Assumption 4: $\frac{\theta_1}{\theta_2} \notin \left[\frac{\Delta NQ_1}{\Delta NQ_2}, \frac{\Delta MA_1}{\Delta MA_2}\right]$

Assumption 1 guarantees that there is a solution for the mechanism equation. Assumptions 2 and 3 imply that there is no solution for the mechanism equation with $\beta^m = 0$. Assumption 4 is not necessary, but it simplifies the analysis by assuring that $\beta^n = 0$.

Next, note that, since $\beta^m > 0$, then

$$\beta^n = \frac{\theta_2 \Delta NQ_1 - \theta_1 \Delta NQ_2}{\Delta MA_2 \Delta NQ_1 - \Delta MA_1 \theta_2}$$

Therefore, assumption 4 rules-out the possibility that $\beta^n > 0$. We must have that the vector of labor market outcomes should be generated by a combination of $\Delta MA, \Delta Cr, \Delta Am$.

Then, we can write:

$$\boldsymbol{\theta} = \beta^m \begin{bmatrix} \Delta MA_1 \\ \Delta MA_2 \end{bmatrix} + \beta^a \begin{bmatrix} \Delta Am_1 \\ \Delta Am_2 \end{bmatrix} + \beta^c * \begin{bmatrix} \Delta Cr_1 \\ \Delta Cr_2 \end{bmatrix}$$

If we solve the linear equation system, we must have that:

$$\beta^m = \frac{\theta_1 * \Delta Am_2 - \theta_2 * Am_1 + \beta^c * (\Delta Cr_2 * \Delta Am_1 - \Delta Cr_1 * \Delta Am_2)}{\Delta MA_2 * \Delta Am_1 - \Delta MA_1 * \Delta Am_2} =$$
$$= \frac{\theta_1 * \Delta Cr_2 - \theta_2 * Cr_1 + \beta^a * (\Delta Cr_1 * \Delta Am_2 - \Delta Cr_2 * \Delta Am_1)}{\Delta MA_2 * \Delta Cr_1 - \Delta MA_1 * \Delta Cr_2}$$

Fixing ideas, consider that:

$$\frac{\Delta MA_1}{\Delta MA_2} > \frac{\Delta Cr_1}{\Delta Cr_2} > \frac{\Delta Am_1}{\Delta Am_2}$$

Then, β^m is strictly increasing in $\beta_c \in \mathbb{R}_{>0}$ and is strictly decreasing in $\beta_m \in \mathbb{R}_{>0}$. It follows that the lower bound for β_m is such that:

$$\beta^c = 0 \implies LB^{\beta_m} = \frac{\theta_1 * \Delta Am_2 - \theta_2 * Am_1}{\Delta MA_2 * \Delta Am_1 - \Delta MA_1 * \Delta Am_2}$$

Similarly, the upper bound fro β_m is such that:

$$\beta^a = 0 \implies UB^{\beta_m} = \frac{\theta_1 * \Delta Cr_2 - \theta_2 * Cr_1}{\Delta MA_2 * \Delta Cr_1 - \Delta MA_1 * \Delta Cr_2}$$

The upper and lower bound occurs when only one other mechanism (either crime rates or amenities) is relevant. If we switch the inequality assumed above, which expression is the lower and upper bounds are switched, but the expressions remain the same.

We can estimate the lower and upper bounds replacing the population parameters in the equations above by their sample analogues. That is:

$$\hat{LB}^{\beta_m} = \frac{\hat{\theta}_1 * \Delta \hat{Am}_2 - \hat{\theta}_2 * \hat{Am}_1}{\Delta \hat{MA}_2 * \Delta \hat{Am}_1 - \Delta \hat{MA}_1 * \Delta \hat{Am}_2}$$

and

$$\hat{UB}^{\beta_m} = \frac{\hat{\theta}_1 * \Delta \hat{Cr}_2 - \hat{\theta}_2 * \hat{Cr}_1}{\Delta \hat{MA}_2 * \Delta \hat{Cr}_1 - \Delta \hat{MA}_1 * \Delta \hat{Cr}_2}$$

Alternatively, we show in the next section that we can estimate these bounds using a Two-Sample Two- Stage Least squares (TS2SLS).

Appendix E2: Equivalence of UB^{β_m} and LB^{β_m} to the TS2SLS estimator

Dix-Carneiro et al. (2018) have shown that bounds on the relative importance of mechanisms are algebraically equivalent to a particular Two-Stage Least Squares (2SLS) estimator. Our application is more complicated because labor market outcomes, instruments and mechanisms are not jointly observed in the same data. Nonetheless, we show below that we can still recover the upper bound UB^{β_m} from a Two-Sample Two-Stage Least squares regression of Δy_{in} on ΔMA_n and ΔCr_n and instruments are given by being drafted to neighborhood 1 (D_{i1}) and neighborhood 2 (D_{i2}).

The structural model of interest in the regression described above is:

$$\Delta y_{in} = \beta \Delta \tilde{\mathcal{M}}_n + \epsilon_{in}$$

where $\tilde{\mathcal{M}}_n = \{MA_n, Cr_n\}$ is a restriction of the complete set of mechanisms. However, we observe two different samples. In sample 1, we observe $\{y^1_{in}, \boldsymbol{D}^1_{in}\}$, for $i = 1, ..., n_1$. In sample 2, we observe $\{\tilde{\mathcal{M}}^2_{jn}, \boldsymbol{D}^2_{jn}\}$, for $j = 1, ..., n_2$. The superscripts indicate the sample that the variable is available. Also, note that the interpretation of $\boldsymbol{D}^1_{in}$ and $\boldsymbol{D}^2_{in}$ are different. In sample 1, the dummies indicate that individual i was drafted to neighborhood one or two. In sample 2, dummies indicate that individual j was living in neighborhood one or two.

The first-stage equation can be written as:

$$\tilde{\mathcal{M}} = \begin{bmatrix} D^2_{i1} & D^2_{i2} \end{bmatrix} * \begin{bmatrix} \Delta MA^2_1 & \Delta Cr^2_1 \\ \Delta MA^2_2 & \Delta Cr^2_2 \end{bmatrix} + u$$

In order to simplify notation, we can write $\boldsymbol{D}^2 = \begin{bmatrix} D^2_{i1} & D^2_{i2} \end{bmatrix}$ and $\boldsymbol{b}^2 = \begin{bmatrix} \Delta MA^2_1 & \Delta MA^2_2 \\ \Delta Cr^2_1 & \Delta Cr^2_2 \end{bmatrix}$. Then, the predicted values of the first-stage can be written as: $\hat{\tilde{\mathcal{M}}} = \boldsymbol{D}^2 * \hat{\boldsymbol{b}}^2$.

The usual 2SLS estimator is obtained by the linear projection of Δy on the predicted values of the first-stage. The estimator can be written as:

$$\boldsymbol{\beta}^{2SLS} = (\hat{\boldsymbol{b}}^{1'} \boldsymbol{D}^{1'} \boldsymbol{D}^1 \hat{\boldsymbol{b}}^1)^{-1} \hat{\boldsymbol{b}}^{1'} \boldsymbol{D}^{1'} \Delta y^1$$

However, this estimator is not feasible in our context because we do not observe $\hat{\boldsymbol{b}}^1$. Instead, we can generate cross-sample fitted values as

$$\hat{\tilde{\mathcal{M}}}_{12} = \boldsymbol{D}^{1'} (\boldsymbol{D}^{2'} \boldsymbol{D}^2) \boldsymbol{D}^{2'} \tilde{\mathcal{M}}$$

Then, we can estimate the TS2SLS by replacing predicted values in the $\boldsymbol{\beta}^{2sls}$ with the cross-sample predicted values:

$$\boldsymbol{\beta}^{TS2SLS} = (\hat{\tilde{\mathcal{M}}}'_{12} \hat{\tilde{\mathcal{M}}}_{12})^{-1} \hat{\tilde{\mathcal{M}}}'_{12} \Delta y^1 = (\hat{\boldsymbol{b}}^{2'} \boldsymbol{D}^{1'} \boldsymbol{D}^1 \hat{\boldsymbol{b}}^2)^{-1} \hat{\boldsymbol{b}}^{2'} \boldsymbol{D}^{1'} \Delta y^1$$

Angrist and Krueger (1992) and Inoue and Solon (2010) have derived the formal properties of this estimator and have shown that this estimator consistently estimate β under common instrumental variable assumptions.

The reduced form estimate the model is:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{D}^{2'} \boldsymbol{D}^2)^{-1} \boldsymbol{D}^{2'} \Delta y \implies (\boldsymbol{D}^{2'} \boldsymbol{D}^2) \hat{\boldsymbol{\theta}} = \boldsymbol{D}^{2'} \Delta y$$

Using the definition of cross-sample fitted values, we can rewrite the TS2SLS estimator as:

$$\boldsymbol{\beta}^{TS2SLS} = (\hat{\boldsymbol{b}}^{2'})^{-1}\hat{\boldsymbol{\theta}} = \begin{bmatrix} \Delta\hat{MA}_1^2 & \Delta\hat{Cr}_1^2 \\ \Delta\hat{MA}_2^2\Delta & \Delta\hat{Cr}_2^2 \end{bmatrix}^{-1} * \hat{\boldsymbol{\theta}}$$

By inverting the matrix $\hat{\boldsymbol{b}}$, we can write:

$$\boldsymbol{\beta}^{TS2SLS} = \frac{1}{\Delta\hat{MA}_1^2 * \Delta\hat{Cr}_2^2 - \Delta\hat{MA}_2^2 * \Delta\hat{Cr}_1^2} * \begin{bmatrix} \Delta\hat{Cr}_2^2 & -\Delta\hat{Cr}_1^2 \\ -\Delta\hat{MA}_2^2\Delta & \Delta\hat{MA}_1^2 \end{bmatrix} * \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}$$

Finally, note that the first element of the $\boldsymbol{\beta}^{TS2SLS}$ matrix is equivalent to the upper bound for β_m:

$$\hat{\boldsymbol{\beta}}^{TS2SLS}[1,1] = \frac{\hat{\theta}_1 * \Delta\hat{Cr}_2^2 - \hat{\theta}_2 * Cr_1}{\Delta\hat{MA}_2 * \Delta\hat{Cr}_1 - \Delta\hat{MA}_1 * \Delta\hat{Cr}_2} = \hat{UB}^{\beta_m}$$

Similarly, we can estimate the lower bound for β^m (LB^{β_m}) in a TS2SLS regression of Δy_{in} on ΔMA_n and ΔAm_n and instruments are given by being drafted to neighborhood 1 (D_{i1}) and neighborhood 2 (D_{i2}).

This equivalence is very useful because it provides a natural way to conduct inference on the bounds of β_m. We can directly apply the robust variance estimator derived by Pacini and Windmeijer (2016). We can estimate:

$$\hat{V}(\hat{\boldsymbol{\beta}}^{TS2SLS}) = \frac{1}{n_1}\left(\hat{\boldsymbol{b}}^2(n_1 * \hat{V}(\hat{\boldsymbol{\theta}})) * \hat{\boldsymbol{b}}^{2'}\right) +$$
$$+\frac{n_1}{n_2}(\hat{\boldsymbol{\beta}}^{'TS2SLS} \otimes \hat{\boldsymbol{b}}^{2'}) * (n_2 * \hat{V}(vec(\hat{\boldsymbol{b}}^2))) * (\hat{\boldsymbol{\beta}}^{TS2SLS} \otimes \hat{\boldsymbol{b}}^2)$$

Finally, once we obtain the variance of β_m, it is straightforward to conduct inference on the upper bound of the fraction of labor market treatment effects explained by market access (f^m). We can estimate:

$$\hat{V}(\hat{f}^m) = \left(\frac{\Delta MA_1 + \Delta MA_2}{\theta_1 + \theta_2}\right)^2 * \hat{V}(\hat{\boldsymbol{\beta}}^{TS2SLS})[1,1]$$