I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16113

Unpacking Neighborhood Effects: Experimental Evidence from a Large-Scale Housing Program in Brazil

Carlos Alberto Belchior
Gustavo Gonzaga
Gabriel Ulyssea

MAY 2023

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16113

Unpacking Neighborhood Effects: Experimental Evidence from a Large-Scale Housing Program in Brazil

Carlos Alberto Belchior
UZH

Gustavo Gonzaga
PUC-Rio

Gabriel Ulyssea
UCL and IZA

MAY 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Unpacking Neighborhood Effects: Experimental Evidence from a Large-Scale Housing Program in Brazil*

This paper investigates the impacts of neighborhoods on the economic outcomes of adults. We exploit one of the world's largest housing lottery programs and administrative data linking lottery registration, formal employment, and access to social programs in Brazil. Receiving a house has positive impacts on housing quality and reduces household expenditures but has negative effects on beneficiaries' neighborhood characteristics. On average, the program has a negative impact on the probability of being formally employed but no effect on the quality of jobs. Poorer individuals, however, experience better formal employment outcomes and lower welfare dependency. We find no differential impacts by distance to beneficiaries' previous homes or jobs. Leveraging a double-randomization design to allocate houses, we show that there are significant differences in effects across neighborhoods and we propose a framework to estimate the relative importance of potential underlying mechanisms. Network quality, amenities and crime play a very limited role, while labor market access explains 82-93% of the observed differences in neighborhood effects.

JEL Classification: H75, I38, O18, R23, R38
Keywords: neighborhood effects, housing programs, labor markets

Corresponding author:
Gabriel Ulyssea
Department of Economics
Oxford University
Manor Road
Oxford OX1 3UQ
United Kingdom
E-mail: gabriel.ulyssea@economics.ox.ac.uk

* We thank Imran Rasul, Kirill Borusyak, Nava Ashraf, Oriana Bandiera, Claudio Ferraz, and participants of numerous seminars, conferences and workshops for helpful comments and discussions. We also thank the Ministry for Social Development for granting access to the identified, individual-level data from the Registry of Social Programs (*Cadastro Único*), according to the process 71000.000372/2018-11, and the Ministry of Labor for granting access to the identified version of the matched employer-employee data (RAIS).

1 Introduction

Housing policies aimed at low-income households are common in both developed (Van Djik, 2019) and developing countries (Barnhardt et al., 2017). An important rationale for these programs is the longstanding hypothesis that relocating disadvantaged households to less deprived areas could enhance their economic prospects and promote upward mobility. Indeed, a substantial body of literature suggests that neighborhoods may affect the contemporaneous outcomes of adults through various channels,[1] such as exposure to crime and violence, peer quality, and access to jobs (Chyn and Katz, 2021). The latter goes back to the "Spatial Mismatch Hypothesis" (Kain, 1968), which emphasizes the proximity to high-quality jobs as a key determinant of neighborhood effects.

However, experimental studies of within-city relocation have produced mixed findings, with most showing no significant effects (and even some negative effects) of relocating households to "better" neighborhoods (e.g. Kling et al., 2007; Ludwig et al., 2013; Barnhardt et al., 2017; Franklin, 2019; Van Djik, 2019; Chyn and Katz, 2021).[2] While these results may suggest limited potential for neighborhoods to shape the economic outcomes of adults, a key empirical challenge lies in effectively accounting for the multidimensional nature of neighborhood quality. Poverty rates or other income measures are often used as proxies, which may not fully capture all relevant dimensions. As a result, empirical evidence regarding the mechanisms through which neighborhoods may impact the economic outcomes of adults remains scarce.

This paper investigates the impacts of the *Minha Casa, Minha Vida* (*MCMV*) program in Brazil, one of the largest housing programs in the world. From 2009 to 2018, it provided approximately 5 million houses and benefited around 14.7 million people, which corresponds to 7 percent of the country's population. We focus on the city of Rio de Janeiro, where a double-randomization design was implemented in 2015 to allocate houses within the program. Specifically, beneficiaries were first drafted to receive a house, and then randomly assigned to one of six housing projects in three different neighborhoods. We leverage this double-randomization feature along with unique data availability to estimate the overall program effects, as well as neighborhood-specific treatment effects. Motivated by the substantial differences in impacts across the three neighborhoods, we propose an empirical framework to evaluate the relative importance of different potential mechanisms in explaining these neighborhood effects.

To do that, we link administrative data on lottery participants with matched employer-employee data on the universe of formal jobs and formal establishments, the RAIS (*Relação*

[1] We distinguish between contemporaneous effects on adults and those that are a consequence of exposure to better neighborhoods during childhood (e.g. Chetty et al., 2016).

[2] Importantly, Van Djik (2019) and Pinto (2021) also document heterogeneous effects across beneficiaries in the Netherlands and the U.S., respectively.

Anual de Informações Sociais). This gives us the complete formal employment history of more than 80 percent of all lottery participants up to ten years before and three years after the lottery.[3] This constitutes our *main sample*. Importantly, the RAIS also contains the exact address of all formal establishments and jobs in Rio de Janeiro and surrounding municipalities. This allows us to compute a labor market access measure at the census tract level, taking into account distances from all residential zip codes to all available formal jobs in the city and surrounding municipalities.

We also link the lottery data to the Single Registry of Social Programs (*Cadastro Único*), which is designed to gather information on all potential beneficiaries of social programs in Brazil, and therefore covers a more disadvantaged population. This dataset provides us with information on place of residence, housing characteristics, informal labor market outcomes, and access to social welfare programs for around 40 percent of potential beneficiaries in the housing lottery data. This constitutes our *disadvantaged sample*. In addition to these administrative data, we use non-identified data from the Demographic Census and other sources to obtain a rich characterization of all neighborhoods in Rio de Janeiro. These data sources provide us with information at the census tract level on the characteristics of its residents, crime, amenities (e.g. parks, nurseries and schools), property prices and rents.

Our results show that nearly 50 percent of those who are drafted take up treatment and relocate, which is comparable to take-up rates in other settings (e.g. Kling et al., 2007; Barnhardt et al., 2017; Franklin, 2019). Furthermore, the vast majority of beneficiaries continue to reside in their assigned housing projects six years after the lottery (end of 2020). Our findings also reveal important trade-offs associated with accepting a *MCMV* house. On the one hand, it provides highly subsidized home ownership,[4] better housing quality (e.g. larger number of rooms and bedrooms), and a substantial reduction in rent payments, even after accounting for modest increases in transportation costs. On the other hand, the housing projects are situated in neighborhoods within the bottom tercile of the city's neighborhood quality distribution in all dimensions considered. Therefore, on average, receiving a *MCMV* house implies moving to neighborhoods with lower average income, higher crime rates, lower labor market access and worse average employment outcomes relative to recipients' previous residences.

In terms of average labor market impacts, taking up a *MCMV* house reduces the probability of being formally employed by 1.7 percentage point, which is consistent with prior research (e.g. Jacob and Ludwig, 2012; Van Djik, 2019). Similarly, we observe a negative effect on the number of months formally employed in a given year. Conditional on being

[3]Even though Brazil is a high informality country, transitions in and out of the formal sector are very common, which explains the high matching rate between these two datasets.

[4]The federal government subsidizes 90-95 percent of the house value, but beneficiaries are not allowed to sell their homes. This could limit some of the positive consequences of the implied wealth shock, such as using the house as collateral for loans and increased credit access.

employed, we do not find any effects on the quality of jobs held. There are no significant effects on wages, occupational rank, or quality of the firm where individuals are employed. However, we do find evidence that individuals who receive a house are more likely to switch jobs, and they tend to move to jobs closer to the housing projects or in a neighboring municipality (the housing projects are close to the western city border). Thus, we find evidence that individuals tend to adjust and find jobs closer to the housing projects, but with no change in the average quality of the jobs held.

Interestingly, the negative average labor market effects turn positive when we focus on the disadvantaged sample. These individuals show a 2.3 percentage point increase in the probability of being formally employed, with corresponding positive effects on the number of months formally employed in a given year. Consistent with these positive labor market effects, the probability of being a beneficiary of *Bolsa Família*, the main conditional cash transfer program in Brazil, decreases by 6 percentage points. We find no statistically significant effects on informal employment, which indicates that the improvements in economic self-sufficiency that lead to a lower participation in the *Bolsa Família* program seem to be driven by the positive effects on formal employment.

Previous studies have argued that disrupted social networks can be important barriers to employment. Indeed, if residential mobility has direct negative effects by removing individuals from their relevant social networks, then it can be a counteracting force against the gains from moving to a more affluent neighborhood (Harding et al., 2021). Even though there is some qualitative evidence in support of this hypothesis (e.g. Turney et al., 2006; Barnhardt et al., 2017), the quantitative evidence is scarcer (see Kling et al., 2007; Harding et al., 2021, for exceptions in the context of the Moving to Opportunity program in the U.S.). We investigate this issue by examining employment effects by distance from participants' previous jobs (main sample) or homes (disadvantaged sample). We find no evidence of heterogeneous effects. To the extent that greater distances imply greater disruption of previously established social networks, these results indicate that this is not a significant factor shaping the program's effects.

The average effects discussed above also hide substantial heterogeneity across the three neighborhoods in which the housing projects are located. In the main sample, the neighborhood-specific treatment effects range from zero to a 6.9 percentage point drop in the probability of being formally employed, with corresponding negative effects on the number of months employed in a formal job. Turning to the disadvantaged households, we find the same gradient of effects across the three neighborhoods but shifted up (consistent with the average results): effects range from negative to an increase of 3.2 percentage points in the probability of formal employment. Given that assignment to neighborhoods was random, these differences cannot be explained by self-selection. We show that they are also not driven by differences in houses' market values across neighborhoods, which would imply heterogeneous

wealth effects. Finally, we implement the model of Angrist and Fernandez-Val (2010), and show that these heterogeneous effects cannot be explained by differences in the observable characteristics of compliers.

In the final part of the paper, we investigate the potential underlying mechanisms that could explain these heterogeneous effects across neighborhoods. We start by discussing a simple model based on Picard and Zenou (2018), which illustrates how different neighborhood characteristics may mediate the relationship between place of residence and labor market outcomes. Guided by the model and existing literature (see Chyn and Katz, 2021, for a recent review), we focus on four main mechanisms: (i) the average quality of peers in the neighborhood, which proxies for network quality (e.g. Picard and Zenou, 2018); (ii) labor market access, which is closely related to the spatial mismatch hypothesis (Kain, 1968; Zenou, 2013); (iii) neighborhoods' amenities and infrastructure; and (iv) crime rates.

We propose an empirical framework that allows us to estimate bounds on the relative importance of these mechanisms to explain the differences in employment effects across neighborhoods. We show that these bounds and their standard errors can be estimated using a Multi-Sample Two-Stage Least Squares (MS2SLS) estimator, which extends the Two-Sample Two-Stage Least Squares (Angrist and Krueger, 1992; Inoue and Solon, 2010) to the case where multiple endogenous variables come from different samples. The results show that the observed patterns of variation in effects across the three neighborhoods cannot be generated by any combination of mechanisms that does not include labor market access. Furthermore, labor market access emerges as the dominant factor, explaining 82–93 percent of the estimated differences in employment effects across neighborhoods.

These bounds are precisely estimated and robust to the use of alternative measures of network quality, as well as to expanding the vector of mechanisms considered. In particular, the estimated bounds remain largely unchanged when we include labor market variables that account for the diversity of sectors and occupations observed in the neighborhoods, as well as measures of quality of occupations and employers. These results suggest that the key neighborhood characteristic for low-skill individuals is the quantity of jobs available in a given location – captured by the labor market access measure – rather than the quality or variety of jobs available.

This paper contributes to the literature that investigates how residential neighborhoods may shape the economic outcomes of adults. The studies that do so by exploring experimental variation from housing programs – such as the Moving to Opportunity in the U.S. – find zero (if not negative) effects on economic outcomes of adults (see Topa and Zenou, 2015; Chyn and Katz, 2021, for reviews).[5] More recently, Van Djik (2019) and Pinto (2021) find

[5]This literature shows positive contemporaneous effects on health outcomes of adults (see, for example, Kling et al., 2007). On children, the experimental and quasi-experimental literature has systematically documented that longer exposure to better neighborhoods leads to improvements in long-run outcomes (e.g. Chetty et al., 2016; Chetty and Hendren, 2018; Chyn, 2018).

heterogeneous effects across beneficiaries in the Netherlands and the U.S., respectively. Our main contribution is to unpack the effects of neighborhoods into the different mechanisms that have been emphasized in the literature, and to quantify their relative importance. In our analysis, labor market access stands out as the key factor that shapes neighborhood effects on low skill individuals. Notably, our findings reveal that labor market access does not necessarily correlate with poverty rates or income levels, which are often used as proxies for neighborhood quality.

This paper also contributes to the scarcer literature that studies the effects of housing programs in low and middle-income countries. Previous studies of the *MCMV* program show similar negative effects on formal employment using lotteries from different cities in Brazil (Rocha, 2018; Pacheco, 2019). The experimental evidence from other mid- and low-income countries indicates no effects on socio-economic outcomes, but finds negative effects on beneficiaries' social networks (Barnhardt et al., 2017; Franklin, 2019).[6] Our results show that there can be substantial heterogeneity in program effects across socio-economic groups and neighborhoods. In particular, the more disadvantaged beneficiaries seem to substantially benefit from the program, even though the labor market effects are negative for the overall population of beneficiaries. We find no heterogeneity by distance to previous residence or previous job, which suggests that disruption of social networks does not play an important role in our context. Finally, our results about mechanisms shed light on which dimensions should be prioritized by policy makers when designing these programs and, in particular, when choosing where to build the housing projects. This can be particularly important in settings where governments have limited fiscal capacity and tight budget constraints.

The remainder of the paper is organized as follows. Section 2 provides a description of the *MCMV* program, the lotteries and the data sources used in the paper. Section 3 describes our empirical strategy, and discusses the program and neighborhood-specific results. Section 4 unpacks the estimated neighborhood effects into the different mechanisms. For that, we first discuss a simple model, and then present our empirical framework and results. Section 5 concludes.

2 Background and Data

This section starts by providing a general description of the *Minha Casa, Minha Vida (MCMV)* housing lottery program and its implementation in the city of Rio de Janeiro, which is our empirical setting. We then describe the data sources, the construction of the samples used, and provide some descriptive statistics of the neighborhoods where the housing projects were built.

[6] Nevertheless, the findings in Franklin (2019) suggest that housing programs can play an important role when there is unmet demand for improved housing by slum dwellers in developing countries.

2.1 The *MCMV* housing program

The program was launched in 2009, with the goal of reducing the housing deficit in Brazil. Initially, it targeted the construction of one million housing units across the country. In 2011, the second phase of the program was launched, which aimed at providing two million additional units. By 2020, around 5 million houses had been delivered, of which 2 percent were in the city of Rio de Janeiro, where we draw our data from.

The federal government subsidizes between 90 and 95 percent of the house value. Beneficiaries can pay the remaining value in up to ten years through monthly installments, which should not exceed 10 percent of their total household income. The average monthly payment was around R$ 50 (approximately US$ 15 in 2015). To be eligible to participate in the housing lottery, individuals should earn less than R$ 1,600 of monthly household income (US$484 in 2015), be Brazilian, older than 18, and should not own a home or have had access to home financing. This income threshold was close to the median household income in the city of Rio de Janeiro, so a large fraction of the population was eligible to participate in the lottery.

The federal government supplied funds for building houses according to the estimated housing deficit in each city, while the local governments provided the necessary public infrastructure for the implementation of the program. The construction was carried out by private firms, which had to submit their projects to financial intermediaries (public banks) according to the minimum criteria established at the national level. In large cities like Rio de Janeiro, the housing projects were built in the outskirts of the city in order to reduce land costs. The average cost per house in Rio de Janeiro was approximately R$ 63,000 (US$ 19,000 in 2015). Since houses were almost fully subsidized by the government, receiving a house through the *MCMV* program represented a significant wealth shock to beneficiaries. However, individuals are not allowed to sell the houses, which can limit some of the benefits of this increase in wealth (such as using the house as collateral for greater or cheaper credit access).

The federal government required that at least 44 percent of the available homes had to be allocated through lotteries. Other 6 percent of the available houses were reserved for elderly individuals and people with disabilities. Local governments should allocate the remaining 50 percent to individuals in vulnerable socio-economic conditions, without necessarily relying on lotteries.

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 registered in the Brazilian Single Registry of Social Programs who lived in the municipality, and other individuals prospected by the Secretary.[7]

[7]The municipal Housing Secretary actively searched for individuals living in vulnerable socio-economic conditions to offer them registration in the program.

In particular, there was a focus on individuals who lived in slums with high environmental risk. This program-specific registry reached more than 600,000 individuals at its peak in 2015, and it was used to define the pool of potential beneficiaries who would be drafted in the housing lottery.

As an attempt to avoid frauds, the *MCMV* lottery was linked to a well-known federal lottery run by a federal public bank, *Caixa Econômica Federal*. Between 2011 and 2014, the city government ran 6 lotteries associated to different housing projects and distributed 8,507 houses. Individuals who were drafted in these lotteries could choose their housing project (among those being offered in each batch) in a first-come-first-served basis. This allocation method represented a significant burden on the municipal bureaucracy. Hence, in the first lottery of 2015, the city changed the allocation method in an attempt to simplify it and make it more transparent. A double-randomization design was introduced, in which individuals were randomly drafted to receive a house, and then randomly allocated across different housing projects. Later in 2015, the Housing Secretary shifted away from this model and stopped randomizing across housing projects. In order to exploit the double-randomization design, we thus only use the first lottery of 2015.

The lottery proceeded as follows: all potential beneficiaries in the program registry were alphabetically ordered 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. The number of drafted individuals was equal across housing projects, but the number of available houses was not. If the number of drafted individuals exceeded the number of available houses, they were offered in alphabetical order and the remaining drafted individuals were included in a wait list. The design of the lottery therefore generated wait lists of different sizes across housing projects.

There were six different housing projects located in three different neighborhoods: *Santa Cruz*, *Campo Grande*, and *Cosmos*. A total of 2,580 houses were drafted in this lottery: 1,500 in *Santa Cruz*, 860 in *Campo Grande*, and 220 in *Cosmos*. As Figure 1 shows, all three neighborhoods are located in the west region of the city, which is very far from the city center (about 50 km). Housing projects in different neighborhoods are far from each other (around 13 km). If an individual wishes to go from one housing project to another using public transportation, it would take her one hour and a half on average and, if she wishes to go from one housing project to the city business district, this would take on average two and a half hours.

2.2 Data

We use four main data sources in this paper. First, administrative data from the program lotteries, which contain the universe 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 individual identifier (*Cadastro de Pessoa Física* – CPF), the housing project that each individual was randomized to, and its location. The second data source is *RAIS* (*Relação Anual de Informações Sociais*), a matched employer-employee, administrative dataset collected by the Ministry of Labor that contains the universe of formal establishments and their formal workers. All formal establishments in Brazil are required to annually fill in information about each of their workers. The Labor Ministry and other Brazilian ministries use this information to manage unemployment insurance and other social programs, including the *MCMV* program. We use a restricted access version of the database that contains the *CPF* identifier of each worker, which allows us to merge the *RAIS* with the housing lotteries data. *RAIS* provides monthly information on individuals' formal labor employment and contract characteristics, including hours and monthly average wages.[8] We focus on the period of 2003-2017.

Third, we use restricted access information from the Single Registry of Social Programs, the *CadUnico*. This database contains individual-level information on family composition, household characteristics, income, some labor market characteristics, and household expenditures. This registry is designed to gather information on all potential beneficiaries of Brazilian social programs in an unique dataset. After the first registration, families are required to update their information every two years (although this is not strictly enforced). Our dataset covers the period of 2012 to 2020. Crucially, the *CadUnico* also contains the *CPF* identifier, which allows us to merge it with the two previous datasets.

Fourth, we use non-identified data from the 2010 Decennial Demographic Census, which is the latest available edition in Brazil. We aggregate the individual-level information at the census tract level, the smallest geographic unit for which there is representative information in the sample version of the Census.[9] We use these data to compute average income, education, labor market outcomes, the share of single-parent households, and housing characteristics (number of rooms, materials, etc.) for all census tracts 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.[10] 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 rich characterization of Rio de Janeiro's neighborhoods. The Appendix Section A.1 describes the details of all data sources used in the paper, and how we define and construct all variables.

[8] Since *RAIS* gathers data from all formal contracts in each year, some individuals appear multiple times in the same year. When this happens, we only keep the contract with the highest total annual wages.

[9] As only very limited information is available in the full Census, we use the sample version that contains a broad set of variables. The sample corresponds to about 15% of the total population and is representative at the census tract level.

[10] Data for public facilities come from the City Hall, while the crime data come from the Rio de Janeiro State Public Security Secretary (ISP).

Samples

We work with two complementary samples obtained by merging the data sources discussed above. Our *main sample* (lottery+*RAIS*) is obtained by merging the lottery administrative data with *RAIS*. This sample includes all individuals who participated in the formal labor market in any year before 2015 (2003-2014). This corresponds to more than 80 percent of individuals in the *MCMV* registry. We use information relative to their most recent formal employment spell in the baseline period. Since the lottery happened in the first days of 2015, we use the period of 2015 to 2017 to construct post-lottery labor market outcomes. In order to maximize statistical power, our main specification stacks all three endline years. The second sample used is the *disadvantaged sample* (lotteries + *CadUnico*), which we obtain by merging information on lottery participants with the Single Registry of Social Programs (*CadUnico*). We only use individuals' most recently updated information from *CadUnico*, both for the baseline (2002-2014) and endline (2015-2020) periods.

Ideally, one would like to merge data from all sources into a single dataset. However, the Single Registry has two characteristics that result in a smaller sample size and a more selected composition compared to *RAIS*. First, the *CadUnico* is specifically designed to cover the population that is eligible for, or is beneficiary of, federal government's social programs. Consequently, only 40 percent of potential beneficiaries in the lottery data can be found in the CadUnico data before the lottery occurs. Second, only a subset of families update their information both before and after the lottery.[11] Despite these limitations, the disadvantaged sample enables us to investigate whether there exist differences in program and neighborhood effects for this subset of individuals who are less educated and more socially vulnerable. Furthermore, the Single Registry provides a more comprehensive set of outcomes that enriches the analysis based on the main sample.

Table 1 shows the main descriptive statistics for control and drafted individuals in both samples. Panel A displays the socioeconomic variables, while Panel B summarizes the labor market characteristics. We highlight three pieces of evidence. First, the randomization was successful and there are no statistically significant differences between control and drafted groups in both samples. Second, as expected, the disadvantaged sample is composed of individuals in more vulnerable socioeconomic conditions: on average they are less educated, more likely to be a woman and non-white, and they fair worse in the formal labor market. Third, take-up was around 50 percent in both samples, which is consistent with take-up rates documented in previous studies (e.g. Kling et al., 2007; Barnhardt et al., 2017; Franklin, 2019).

[11] Additionally, the Municipal Housing Secretary registered some drafted individuals just after the lottery, which creates an unbalanced sample of updates after the lottery. We exclude these individuals and all their future updates from the data. This represents less than 0.2 percent of the full sample.

Neighborhood Characteristics

The *RAIS* data contain the exact address of all formal establishments in Brazil, which allows us to geo-reference all formal jobs in the city of Rio de Janeiro and neighboring municipalities. We use these data, and detailed information on residential zip codes, to construct a measure of labor market access for all census tracts in Rio de Janeiro, which plays a central role in our analysis.

For each census tract n, we compute labor market access as follows:

$$LMA_n = \sum_{h \in n} \sum_{j \in J} \frac{d(h,j)\omega_{hc}}{H_n} \tag{1}$$

where:

$$w_{hc} = \frac{R_h + L_j}{\sum_{h \in n} \sum_{j \in J} (R_h + L_j)}$$

and $h \in n$ denotes residential locations (zip codes) in neighborhood n; j denotes the establishment's location (zip code), and J is the set of all possible locations; $d(.,.)$ is a distance function;[12] H_n is the number of residential zip codes in n; R_h is the number of residents in h; and L_j is the number of workers in j. Given the population served by the program, we restrict the analysis to individuals with up to completed high school, and only consider jobs occupied by workers with this level of schooling. For each census tract, we evaluate the distance between half a million and three million combinations of points.

As discussed above, we use several auxiliary, non-identified data sources to measure a broad set of neighborhood characteristics. We use the Demographic Census to measure a summary of different types of income, the level of education, and share of single parent households. The table also reports the Labor Market Access measure described by equation 1, as well as the total number of formal jobs located in each neighborhood. As for income, amenities and crime, we aggregate different variables using a principal component analysis. We use three variables for income: labor household income, total income per capita and housing conditions. We use five variables for crime: number of auto thefts, thefts, rapes, kidnappings, and murders. For amenities, we use four measurements: number of public parks, schools, daycare and health facilities. The first principal component explains about 70 percent of total income variance, 60 percent of the total variance in crime measurements, and about 50 percent of the total variance in the measurements of amenities. For all these variables, we report the percentile of each neighborhood in the city's distribution across all neighborhoods. Thus, all variables range from one to 99 and the closer to zero the worse the neighborhood is for that particular measure.

Table 2 shows some interesting facts, of which we highlight three that play an important

[12] To compute the distance, we use the coordinates of all zip codes' centroids and apply the Vicenty (1975) method for calculating the shortest distance for every pair of points.

role throughout our empirical analysis. First, the *MCMV* neighborhoods are at the low end of the city's neighborhood quality distribution, regardless of the measure considered. Second, there is substantial variation across neighborhoods in almost all dimensions with some sizable gaps between them. For instance, there is a 26 percentiles differential in the income rank between Neighborhoods 1 and 3. Finally, no neighborhood strictly dominates others in all dimensions. Taking again Neighborhoods 1 and 3, even though the former clearly dominates the latter in terms of average income and schooling of residents, the reverse is true for labor market access, and also by a large margin.

3 Program and Neighborhood Effects

3.1 Empirical Strategy

In this section, we leverage the double-randomization design described in Section 2.1 to estimate the overall *MCMV* program effects, as well as neighborhood-specific treatment effects. We estimate both the intent-to-treat (ITT) and treatment-on-the-treated (TOT) parameters. To estimate the program ITT, we use the following standard specification:

$$y_i = \beta_0 + \beta_1 D_i + \boldsymbol{X}_i \boldsymbol{\beta_2} + \epsilon_i \tag{2}$$

where y_i is the outcome of interest, D_i is a dummy variable for drafted individuals, and $\boldsymbol{X}_i$ is a vector of covariates that includes gender, race, and schooling; ϵ_i denotes the error term.

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

We take advantage of the double-randomization design and estimate the ITT and TOT parameters for each of the three neighborhoods in which the housing projects were constructed. The neighborhood-specific ITT and TOT are estimated using the following regressions, respectively:

$$y_i = \delta + \sum_{k=1}^{3} \xi_k D_{i,k} + \boldsymbol{X}_i \boldsymbol{\Gamma} + \epsilon_i \tag{3}$$

$$y_i = \alpha + \sum_{k=1}^{3} \gamma_k H_{i,k} + \boldsymbol{X}_i \boldsymbol{\zeta} + \epsilon_i \tag{4}$$

where $D_{i,k}$ takes value one if individual i is randomized into neighborhood k; and $H_{i,k}$ equals

one if individual i receives a house in neighborhood k, which we instrument with $D_{i,k}$.

3.2 Effects of the *MCMV* Program

We start by examining where drafted individuals were living before the lottery, and whether they took up the offer of a *MCMV* house. We do so by using the disadvantaged sample, which allows us to contrast the distribution of drafted individuals across all census tracts in Rio de Janeiro before and after the lottery. Figure 2 shows that individuals were spread throughout the city at baseline and were not particularly concentrated in neighborhoods close to where the *MCMV* housing projects were built. In contrast, we observe a large concentration of drafted individuals living in these neighborhoods after the lottery. In Appendix B, we show that individuals started moving to the housing projects six months after the lottery, and the fraction of beneficiaries who report living in their original housing project remains roughly constant until the end of 2020.

Table 3 shows ITT estimates of the effects on individuals' housing characteristics and neighborhoods. The table reveals interesting trade-offs associated to the decision to take up a *MCMV* house. On the one hand, being drafted to receive a house leads to improvements in housing quality (as measured by the number of rooms and bedrooms), and to a substantial reduction in rent expenditures, even after taking into account the modest increase in transportation costs. On the other hand, the quality of neighborhoods individuals live in decreases in most dimensions considered. The largest deterioration seems to occur along neighborhoods' income levels and crime rates, with a reduction of 9.5 and 6.1 percentiles relative to their previous neighborhood of residence, respectively.

Turning to labor market outcomes, Table 4 shows that the program has negative effects on the extensive margin of formal employment for the main sample: the TOT estimates show that treated individuals have a 1.7 percentage point lower probability of being formally employed and spend fewer months formally employed in a given year.[13] These results are consistent with previous experimental results from developed countries (e.g Van Djik, 2019; Chyn and Katz, 2021). In the Appendix, we show that these negative effects are concentrated on white and more educated individuals (Table C.3).

Conditional on being employed, we find no impacts on wages. To further investigate the effects on the quality of jobs held, we construct rankings of occupations and employers in the city of Rio de Janeiro. We compute the average wage for all occupations in the city using the baseline period, and rank them based on their percentile in the distribution of occupation-level average wages in the city. For employers, we estimate establishment fixed effects in a log-wage regression controlling for workers' gender, race, age, and schooling, and rank them based on their percentiles in the distribution of establishment fixed effects (see also Appendix

[13] Appendix Table C.1 shows the year-by-year treatment effects. The employment results are robust to using ANCOVA or difference-in-difference specifications (Table C.2).

A). For either measure, we find no effects. Finally, Table 4 provides evidence that individuals try to adapt by switching jobs to reduce commuting time. We find that those in the treated group are more likely to switch jobs (4 percentage points less likely to keep the same job), tend to start working in firms that are closer to the housing projects (1.6 kilometer closer), and are more likely to work in firms in neighboring municipalities (the housing projects are located near municipality borders). However, we do not find any heterogeneous effects with respect to the distance to individuals' previous formal job (Figure 3).[14] Thus, even though treated individuals are more likely to switch jobs to cut on commuting time, distance to their previous job does not seem to be a key determinant of the labor market effects we find.

The results for the disadvantaged sample in Table 5 provide a different picture. For these individuals, there is an increase in the probability of being formally employed of 2.3 percentage points, and a small increase in the number of months spent formally employed, but no effect on wages conditional on being employed.[15] Consistently, there is a 6 percentage points decrease in the probability of being a beneficiary of the *Bolsa Família* program, the largest conditional cash transfer program in Brazil, in the three years after receiving a house. In Appendix Table C.4, we show that there are no statistically significant effects on informal employment. Hence, the improvements in economic self-sufficiency that lead to a lower participation in the *Bolsa Família* program seem to be driven by the positive effects on formal employment. These results are consistent with the heterogeneous effects found in the main sample, which show that the negative effects of the program are concentrated on more educated and less vulnerable individuals.

We also investigate whether the disruption of previously existing social networks could be driving these results, as discussed in other settings (e.g. Turney et al., 2006; Barnhardt et al., 2017; Harding et al., 2021). We find no evidence of heterogeneous effects on the probability of formal employment by distance to beneficiaries' previous place of residence (Figure 4).[16] To the extent that greater distances imply a greater disruption from moving, these results suggest that this is not an important factor behind the labor market effects we estimate.

3.3 Neighborhood effects

We now investigate weather the main average effects discussed in the previous section vary across the three neighborhoods where the housing projects are situated.[17] Table 6 shows the

[14] Figure 3 compares drafted and control individuals within each bin of baseline distance controlling for gender, race, schooling, and baseline formal employment.

[15] For conciseness, we do not report the other outcomes relative to quality of occupation and employers reported in Table 4, for which we also find null effects in the disadvantaged sample. The results are available upon request.

[16] As in Figure 3, we contrast drafted and control individuals within each bin of baseline distance controlling for gender, race, schooling, and baseline formal employment.

[17] For conciseness, we restrict attention to formal employment variables, wages and whether individuals are in the *Bolsa Família* program. Results by neighborhoods for the other outcomes discussed in Table 4 are available upon request.

results of estimating regression 3 in the main and disadvantaged samples separately (Panels A and B, respectively). As the table shows, the overall program effects discussed so far mask substantial heterogeneity. In the main sample (Panel A), the treatment effects on formal employment range from no effect (Neighborhood 3) to a decrease in the employment probability of 6.9 percentage points (Neighborhood 1). The effects of being allocated to Neighborhoods 1 and 2 are quite similar, while Neighborhood 3 clearly stands out.[18] The effects estimated in the disadvantaged sample follow the same pattern across neighborhoods, but are everywhere more positive (or less negative) relative to the main sample. Consistently, the effects of reduced welfare dependency are concentrated in Neighborhood 3, where beneficiaries experience a reduction of 7.5 percentage points in the probability of being in the *Bolsa Família* program.

Even though assignment to the different neighborhoods was random, it is possible that differences in the characteristics of compliers across neighborhoods could explain the differences in treatment effects. We initially examine this hypothesis in Tables 7 and 8 for the main and disadvantaged samples, respectively. First, we note that once we exclude the wait list, take-up rates are very similar across neighborhoods, albeit slightly lower for Neighborhood 1 in the main sample. Second, socioeconomic and labor market characteristics of compliers at baseline are very similar and statistically equal across neighborhoods. In Appendix D, we formally show that the differences in compliers' characteristics cannot explain the heterogeneity in treatment effects across the three neighborhoods. We implement Angrist and Fernandez-Val (2010) methodology to 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 be even slightly higher than shown in Table 6.

Finally, even though the units provided in different housing projects are identical, their market value could vary across neighborhoods. In this case, the heterogeneity in treatment effects could be reflecting differences in the magnitude of the wealth shock received. To examine this hypothesis, we scraped data on prices and characteristics of houses in the three *MCMV* neighborhoods from the three most important websites used to sell real state in Brazil (we provide details on the scraping procedure in Appendix A). Interestingly, although individuals were not allowed to sell their houses, we were able to find a small number of *MCMV* units for sale (around 80).[19] We complement this analysis with individual-level data from the Demographic Census, which contains information on rent values. The advantage of using Census data is that it provides a representative sample of individuals paying rent in those neighborhoods, as well as a wide array of housing characteristics not available in our

[18] We test whether the neighborhood-specific coefficients are jointly equal, and we reject the null for all dependent variables ($p < 0.1$) in the main sample.

[19] We identify *MCMV* units either by the combination of addresses and housing characteristics, or by the description of the houses.

scraped data.[20]

We use these data to estimate the following simple regression:

$$log(y_h) = \alpha_0 + \alpha_1 D_{1,h} + \alpha_2 D_{2,h} + \boldsymbol{\gamma} \boldsymbol{X}_h + \epsilon_h \tag{5}$$

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

We run this regression for three different samples: (i) all properties; (ii) only those with similar characteristics to the *MCMV* units; and (iii) *MCMV* units being sold online. Table 9 shows the results. 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 1 percent higher (Column 1). If we focus on units with similar characteristics to the *MCMV*, the price differences become much smaller, and they completely vanish when we use the sample of *MCMV* houses. The results using rents paid show no statistically significant differences across neighborhoods. In sum, the difference in wealth shocks received by beneficiaries across neighborhoods appears to be very small, if not zero. It is therefore unlikely that this factor can explain the large differences in neighborhood-specific treatment effects shown in Table 6.

4 Unpacking neighborhood effects

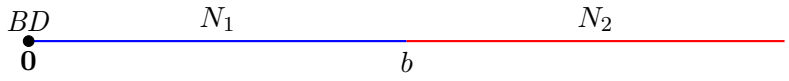
The results from the previous section show that the housing lottery program had quite heterogeneous effects across neighborhoods. Moreover, these differences cannot be explained by heterogeneity in the set of compliers, or by differences in the magnitude of the wealth shock received by beneficiaries across the three neighborhoods.

Thus, we proceed to investigate the potential mechanisms behind the heterogeneous neighborhood effects discussed in the previous section. We do so in two steps. First, we discuss a simple model based on Picard and Zenou (2018), which illustrates how different neighborhood characteristics can affect employment outcomes. Second, we propose an empirical framework based on Dix-Carneiro et al. (2018) to estimate bounds on the relative importance of the different mechanisms through which neighborhoods can shape individuals' outcomes. Guided by the model and existing literature, we focus on four main mechanisms: (i) average quality of peers in the neighborhood, which we refer to as network quality; (ii) labor market access (or proximity to jobs); (iii) amenities and infrastructure; and (iv) crime rates.

[20]The Demographic Census includes information on the number of rooms and bathrooms, dummies for the supply of potable water and energy, the type of residence (house, apartment, etc.) and the external material of the houses.

4.1 Model

Consider a unit width linear city with workers residing along its extension. The city locations are indexed by x. Workers might live in two different neighborhoods, $N_i \in [0,1]$, $i = 1, 2$, and are uniformly allocated along the city. The population sizes of each neighborhood 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 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, $w \in (0,1)$, and have the same preferences:

$$U_i(x) = e_i(x)(w - tx) - C_i(x) \tag{6}$$

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, which we discuss below. Employed individuals may lose their jobs with probability γ. When individuals are unemployed, they search for a job with success probability $\pi(x)$. For each neighborhood, total employment is defined by $E_i = \int_{N_i} e_i(x)dx$.

In this model, social interactions play a central role in finding jobs. Individuals interact only with those who live in their neighborhood. Workers choose how many individuals to interact with, $n_i(x)$, and randomly meet individuals living in the same neighborhood as them. Let neighborhood average income be denoted by $m_i = w\frac{E_i}{P_i}$. We define job search success probability as:

$$\pi_i(x) = \alpha_i n_i(x) m_i \tag{7}$$

that is, the probability of finding a job increases with the number of individuals a worker decides to meet, the average income in her neighborhood and exogenous neighborhood characteristics, represented by α_i.

The presence of neighborhood average income in expression 7 captures the idea that higher quality peers – i.e. individuals 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. α_i is a fixed parameter that represents a broad array of neighborhood characteristics, such as crime rates, amenities, public good provision, etc. These are therefore treated as exogenous in the model.

Individuals incur in a constant cost c of interacting with others, and the total cost of social interaction is simply $C_i(x) = cn_i(x)$. Workers choose the number of social interactions

that maximizes their total utility:

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

Equilibrium and characterization of neighborhood effects

In equilibrium, the flow of individuals into employment must be the same as the flow out to unemployment, so we have that:

$$e_i(x) = \frac{\pi(x)}{\gamma + \pi(x)} \tag{8}$$

and we can combine equations (7) and (8) to obtain the employment probability, which is given by:

$$e_i(x) = \frac{\alpha_i n_i(x) m_i}{\gamma + \alpha_i n_i(x) m_i} \tag{9}$$

Thus, the first-order condition for the individual problem is given by:

$$e_i^*(x) = \left(1 - \sqrt{\frac{\gamma c_i(x)}{\alpha_i (w - tx) m_i}}\right) \tag{10}$$

which implies the following expected unemployment probability:

$$u_i^*(x) = \sqrt{\frac{\gamma c_i(x)}{\alpha_i (w - tx) m_i}} \tag{11}$$

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. To illustrate the mechanisms of the model, we focus on the equilibrium where individuals in the distant neighborhood are more likely to be employed, that is $\frac{E_2}{P_2} > \frac{E_1}{P_1}$. As discussed in Picard and Zenou (2018), this corresponds to a city like New York, where low-skill workers live close to the business center while high-skill, high-income individuals live in the suburbs. Thus, in this example we assume that $\alpha_2 > \alpha_1$, which corresponds to assuming that Neighborhood 2 has better amenities or lower crime, for example.[21]

Now, consider moving one individual from x_1 to x_2, with $x_1 < b < x_2$. In words, consider moving an individual from Neighborhood 1 to Neighborhood 2, and thus further away from the central business district. The unemployment ratio before and after the residential change

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

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

is given by:

$$\frac{u_2^*(x)}{u_1^*(x)} = \underbrace{\sqrt{\frac{(w-tx_1)}{(w-tx_2)}}}_{(a):>1} \times \underbrace{\sqrt{\frac{m_1}{m_2}}}_{(b):<1} \times \underbrace{\sqrt{\frac{\alpha_1}{\alpha_2}}}_{(c):<1} \tag{12}$$

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

(a): Spatial mismatch mechanism: this term is greater than one and reflects that moving further away from the business center implies higher commuting costs, which increase the likelihood of unemployment.

(b): Network quality mechanism: it is smaller than one, as individuals are moving to a neighborhood with a higher share of employed peers and, therefore, higher income.

(c): The direct effect of neighborhoods' exogenous characteristics on employment.

4.2 Empirical Framework

The net effect of the mechanisms discussed above on employment outcomes is *a priori* ambiguous, and depends on their relative strength. This section extends the methodology proposed by Dix-Carneiro et al. (2018) to develop an empirical framework to estimate bounds on the relative importance of these different channels. We focus on four mechanisms. The first two – network quality and labor market access – come directly from the model, and have been extensively analyzed in the literature (Topa and Zenou, 2015; Chyn and Katz, 2021). The other two determinants come from separating exogenous neighborhood characteristics (parameter α_i in the model) into two dimensions: amenities (Roback, 1988; Krupka and Donaldson, 2007; Moretti, 2010), and crime rates (Grogger, 1997, 1998; Huang et al., 2004; Freedman et al., 2018). In the robustness analysis, we investigate whether expanding this set of mechanisms affects our results.

We start by assuming that there is an equilibrium relationship between neighborhood characteristics and labor market outcomes, which is not affected by the *MCMV* program.[22] One potential challenge in our context is that receiving a house can directly affect employment outcomes due to the wealth shock implied, or because beneficiaries have better and more stable housing conditions. These effects would not be mediated by neighborhood characteristics. However, given that we find no evidence of heterogeneous wealth shocks across neighborhoods (Section 3), and that all houses in the program are exactly the same, we assume that the direct effect of receiving a house is constant across neighborhoods.

Concretely, we assume that the relationship between receiving a *MCMV* house located

[22]The number of *MCMV* beneficiaries is small relative to the existing population in these neighborhoods, so it is unlikely that the program generated important general equilibrium effects in the neighborhoods.

in neighborhood n and the employment outcome of interest can be described as follows:

$$y_{in} = \alpha D_{in} + \mathcal{M}_n \beta + \varepsilon_{in} \tag{13}$$

where y_{in} is the outcome of interest for individual i living in neighborhood n, D_{in} is a dummy variable for receiving a house in neighborhood n, and $\mathcal{M}_n$ is the vector of mechanisms described above given by:

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

where LMA_n is labor market access, NQ_n denotes network quality, Am_n is amenities, and Cr_n denotes crime.

Thus, receiving a house can have a direct effect on labor market outcomes that is constant across neighborhoods, captured by α, and indirect effects that are mediated through neighborhood characteristics, $\mathcal{M}_n$. To parse-out the direct and indirect effects, we use the main sample and focus on the subsample of drafted individuals to explore variation in treatment effects across neighborhoods. Given that the direct effect of receiving a house is assumed to be constant across neighborhoods, any differences in labor market outcomes can be attributed to different neighborhood characteristics. Taking the difference between individuals drafted to Neighborhoods 1 and 2 relative to those drafted to Neighborhood 3 (which we use as a reference group):

$$\Delta E(y_n) = \Delta \mathcal{M}_n \beta + E(\Delta \varepsilon_n) \tag{14}$$

This mechanism analysis requires that $E(\Delta \varepsilon_n | \Delta \mathcal{M}_n) = 0$, which requires two assumptions to hold. First, neighborhood choice must be exogenous, which is the fundamental empirical challenge in the neighborhood effects literature (e.g. Kling et al., 2007). As we argue below, the double randomization introduces additional exogenous variation to aid identification. Second, total mediation must hold, which 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 it is not possible to empirically test this assumption, we show in the Appendix G that our results are robust to the inclusion of several other potential mechanisms.

Using expression (13), the difference in average outcomes between Neighborhoods 1 and 2 relative to Neighborhood 3 can be expressed as follows:

$$\Delta y_n = \beta^m \begin{bmatrix} \Delta LMA_1 \\ \Delta LMA_2 \end{bmatrix} + \beta^n \begin{bmatrix} \Delta NQ_1 \\ \Delta NQ_2 \end{bmatrix} + \beta^a \begin{bmatrix} \Delta Am_1 \\ \Delta Am_2 \end{bmatrix} + \beta^c \begin{bmatrix} \Delta Cr_1 \\ \Delta Cr_2 \end{bmatrix} + \Delta \varepsilon_n \tag{15}$$

To point identify the effects of all different mechanisms in a vector of size K, we would need to have randomization into $K+1$ neighborhoods. Nevertheless, we show that it is possible

to estimate bounds on the relative importance of these mechanisms by imposing mild sign restrictions on the β coefficients in equation (15). We assume the following: improved labor market access and network quality cannot make individuals strictly worse off, $\beta^m, \beta^n \geq 0$; more amenities cannot harm employment outcomes, $\beta^a \geq 0$; and higher crime rates cannot improve labor market outcomes, $\beta^c \leq 0$.

Figure 5 illustrates the intuition of our argument by plotting the differences in estimated neighborhood-specific treatment effects and changes in mechanisms across neighborhoods imposing these sign restrictions. We plot the differences between Neighborhoods 1 and 3 in the horizontal axis, and Neighborhoods 2 and 3 in the vertical axis. We start by noting that, mathematically, no positive linear combination of mechanisms that excludes labor market access can generate the differences in employment effects, as the latter does not belong to the cone generated by the former. In other words, differences in labor market access are necessary to explain the differences in employment effects across neighborhoods.

Furthermore, it is possible to impose bounds on β^m by considering all potential pairs of mechanisms that include labor market access. The upper (lower) bound is given by the combination that gives the most (least) emphasis to labor market access. Simple visual inspection of Figure 5 shows that the lower bound is given by the positive linear combination of β^m and β^c (the crime coefficient), while the upper bound is given by the combination of β^m and β^a (the coefficient on amenities). We show in the Appendix F that the expressions for the lower and upper bounds are given by:

$$\frac{-\theta_1 \Delta Cr_2 + \theta_2 \Delta Cr_1}{\Delta LMA_1 \Delta Cr_2 - \Delta LMA_2 \Delta Cr_1} \leq \beta_m \leq \frac{-\theta_1 \Delta Am_2 + \theta_2 \Delta Am_1}{\Delta LMA_1 \Delta Am_2 - \Delta LMA_2 \Delta Am_1} \tag{16}$$

where θ denotes the vector of differences in neighborhood specific treatment effects relative to Neighborhood 3.

Estimation and Results

To estimate the bounds, we use the variables discussed in Section 2 as the empirical measures of the four mechanisms discussed above. For Labor Market Access, we use the measure described by equation 1. As for income, amenities and crime, we use the respective principal components of different relevant measurements, and we standardize those as percentiles of the distribution across neighborhoods in Rio de Janeiro (see Section 2).[23] In addition to this income variable, in Appendix G we consider several alternative proxies for neighborhood's network quality: poverty rate; average years of schooling; and a principal component index of different socioeconomic measures that include income, schooling, poverty, average formal

[23] As a reminder, we summarize three income measures: labor household income, total income per capita and housing conditions. We use five crime variables: number of auto thefts, thefts, rapes, kidnappings, and murders. For amenities, we use four measurements: number of public parks, schools, daycare and health facilities.

wages, and average baseline formal wages of peers drafted to the same housing project. Our results are robust to using all of these measures.

Even though one can obtain the bounds described in expression 16 by replacing the population parameters with their respective estimates, inference is not straightforward. The neighborhood specific treatment effects are estimated from the main sample, while the variables used to measure the four mechanisms are taken from different datasets with different number of observations. Nevertheless, we show that it is possible to recover these bounds and to correctly estimate standard errors using a Multi-Sample Two-Stage Least Squares (MS2SLS) estimator. This estimator is an extension of the Two-Sample Two-Stage Least Squares estimator (Angrist and Krueger, 1992; Inoue and Solon, 2010; Pacini and Windmeijer, 2016) for the case of endogenous variables that come from different samples. In Appendix E, we discuss this estimator and derive its properties and asymptotic distribution.

We estimate the MS2SLS using the formal employment dummy as dependent variable, and Labor Market Access plus each one of the alternative mechanisms at a time as endogenous regressors. As discussed above, we use crime and amenities as the alternative mechanisms to estimate the lower and upper bounds, respectively. We instrument the two mechanisms by dummies of being drafted to Neighborhoods 1 and 2.

Table 10 shows the MS2SLS estimates of the lower and upper bounds for β^m, and their robust standard-errors. We present the results as the fraction of total variation in treatment effects that is explained by labor market access, which is given by the following expression:

$$f^m = \frac{\beta^m(\Delta LMA_1 + \Delta LMA_2)}{\theta_1 + \theta_2}$$

The results displayed in Table 10 show that labor market access explains most of the variation in labor market effects across neighborhoods – between 82 and 93 percent of total effects. These bounds are precisely estimated. In Appendix G, we consider different alternative measures of network quality: average education, average household income, average formal wages in the neighborhood and in the housing projects that individuals were drafted to, as well as a PCA index of all these variables. Our results on the bounds of Labor Market Access remain largely unchanged. This is reassuring, as the literature puts great emphasis on quality of peers as a key mechanism through which neighborhoods could affect individuals' economic outcomes.

Table F.2 shows the results of expanding the set of potential mechanisms. We consider the following additional factors: (i) total population in the neighborhood, which could capture the scale, density and potentially the diversity of the network; (ii) occupational diversity in the neighborhood, which could imply broader employment opportunities; (iii) industry diversity, which could also capture more varied economic opportunities for individuals through their network in the neighborhood; and (iv) quality of employers (establishments), which is the

measure discussed in Section 3 averaged at the neighborhood level. As Table 10 shows, the results remain largely the same. Hence, these results suggest that the key neighborhood dimension for low-skill individuals is the quantity of jobs available to them in a given location – captured by the labor market access measure – rather than the quality or variety of jobs.

5 Final remarks

This paper investigates the mechanisms through which neighborhoods can affect the economic outcome of adults. For that, we use one of the largest housing lottery programs in the world, the *Minha Casa Minha Vida* in Brazil. We leverage unique administrative data sources linking lottery registration, formal employment outcomes, and access to social programs, combined with a double-randomization design to allocate houses within the program introduced in 2015 in Rio de Janeiro. In this lottery, not only individuals were randomly selected to receive a house, but they were also randomly allocated across six housing projects located in three different neighborhoods.

We find that the program has negative average effects on formal employment, decreasing the probability of being formally employed and the number of months workers spend formally employed in a given year. Conditional on employment, we find no effects on wages or the quality of jobs held. However, once we focus on the sample of more disadvantaged beneficiaries the effects turn positive, with a 2.3 percentage point increase in the probability of being formally employed. These individuals also show a lower probability of being a beneficiary of the main conditional cash transfer program in Brazil, the *Bolsa Família*. We find no effects on informal employment, which indicates that this lower dependency on welfare is driven by the positive effects on formal employment. These average effects hide substantial heterogeneity across neighborhoods. Given the random assignment across neighborhoods, these differences cannot be explained by self-selection. Additionally, we show that they cannot be explained by differences in houses' market values across neighborhoods (i.e. differential wealth effects), or differences in observable characteristics of compliers.

We propose an empirical framework to estimate the relative importance of potential underlying mechanisms behind these heterogeneous neighborhood effects. Our results show that average quality of peers in the neighborhood, crime and amenities play a very limited role. Indeed, Labor Market Access stands out as the primary factor determining the differences in impacts across neighborhoods, accounting for 82–93 percent of the total observed variation in treatment effects on employment outcomes across the three neighborhoods. These results are very robust to different measures of quality of peers or to expanding the set of mechanisms considered. Thus, our results provide strong support in favor of the "Spatial Mismatch Hypothesis".

In sum, similarly to the previous literature, we show that the housing program has small

negative effects on the probability of formal employment. Conditional on employment, we find no effects on the quality of jobs held. However, the effects turn positive when we focus on the more disadvantaged individuals, who show improved labor market outcomes and lower reliance on social welfare programs. We find no heterogeneous effects with respect to the distance to beneficiaries' previous jobs or previous homes, which suggests that the potential disruption of previously existing social networks is not a major determinant of these effects. Our results show that neighborhoods matter, as there are substantial differences in effects across neighborhoods. For example, in the overall population of beneficiaries, the effects range from zero to a decline of 6.9 percentage points in the probability of being formally employed. Finally, our mechanism analysis shows that proximity to formal jobs is the main determinant for economic outcomes of adults. Moreover, our results indicate that it is the quantity, and not the quality or variety of formal jobs, that matters for low-skill individuals.

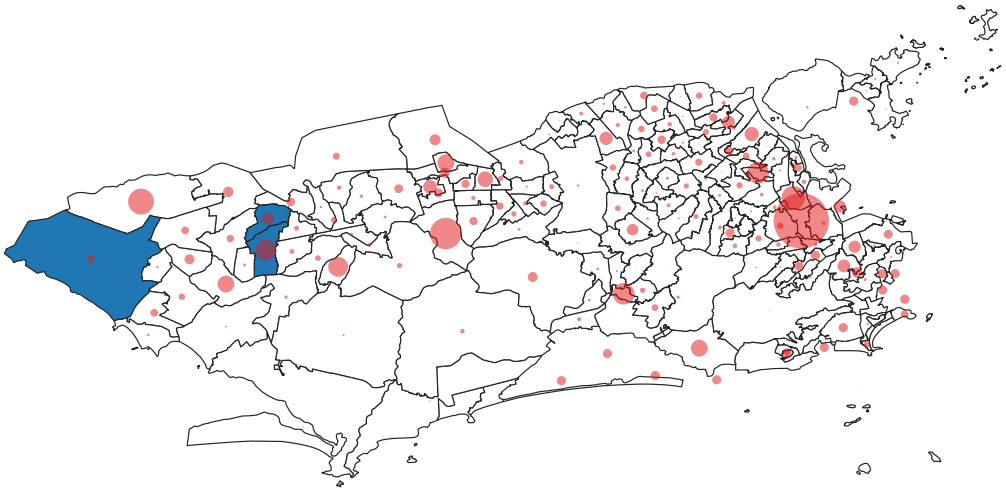
These results have important implications for the design of housing policies in developing countries. In particular, they suggest that the provision of housing units located far from employment opportunities is unlikely to result in enhanced economic prospects for low-income households, despite the highly subsidize home ownership and improvements in housing conditions.

Tables and Figures

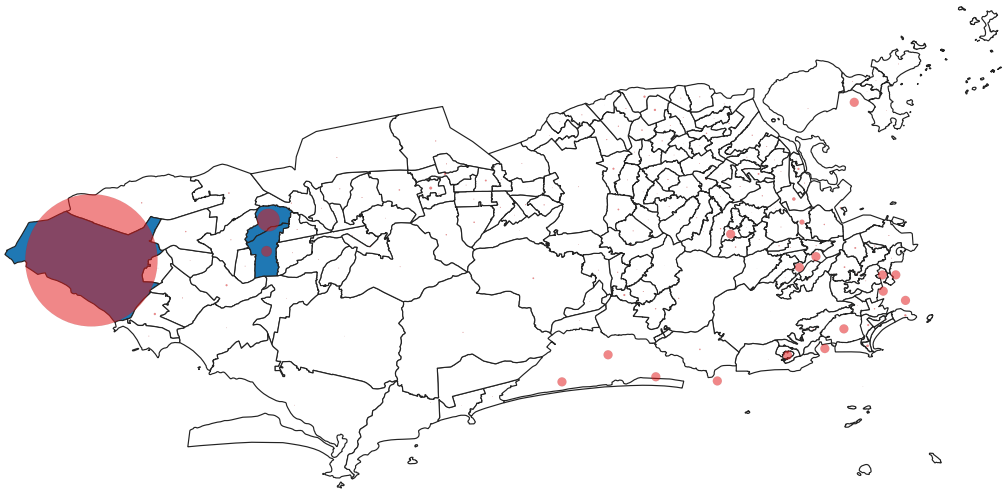
Figure 1: MCMV neighborhoods (hashed areas) and the center of Rio de Janeiro city (red area)

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 2: Relative number of drafted individuals by neighborhood



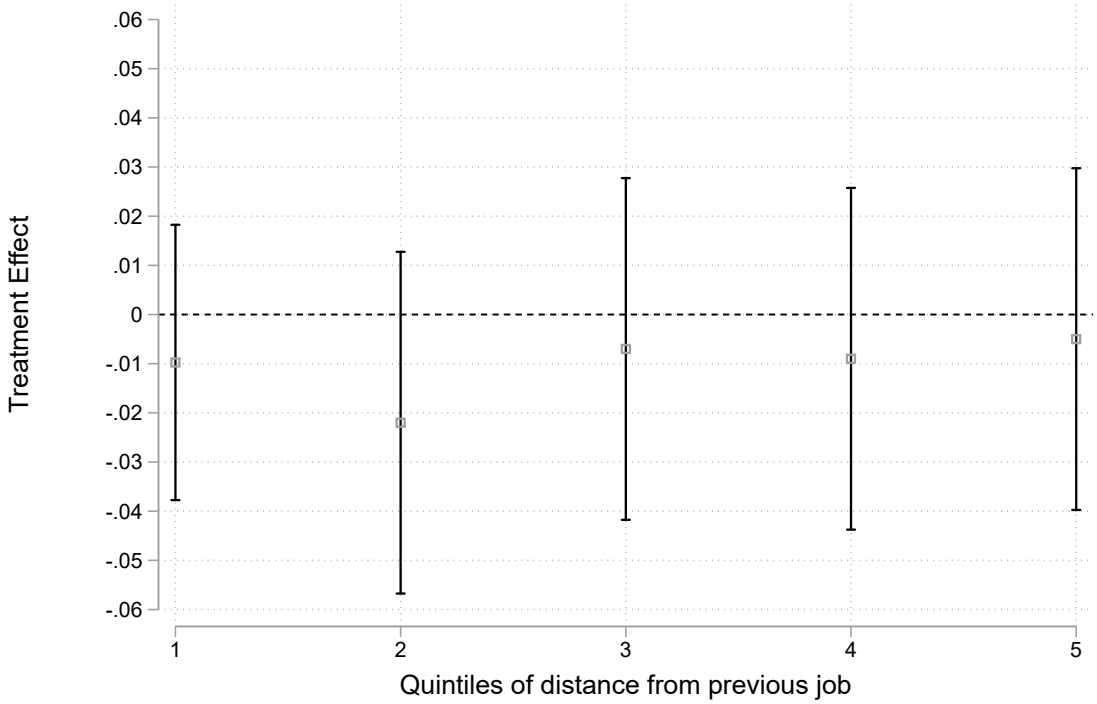
(a) Baseline



(b) Endline

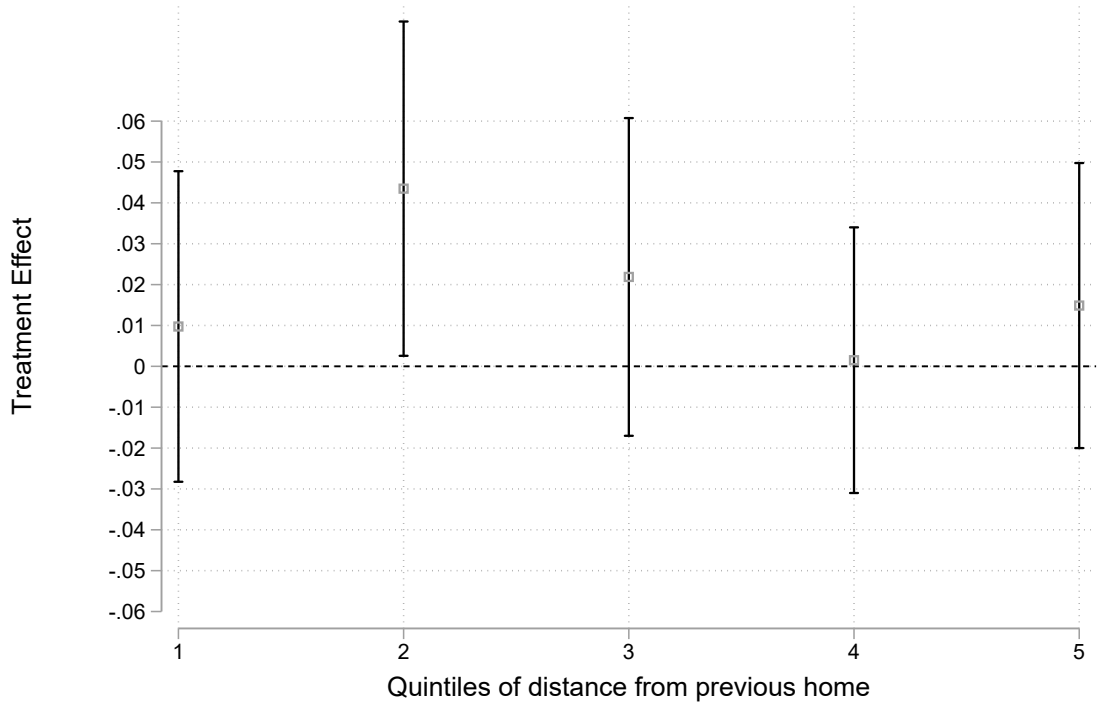
Note: This Figure shows the Rio de Janeiro map and its census tracts. The census tracts where the *MCMV* housing projects are located are highlighted in blue. In both panels, we represent the fraction of individuals that were drafted living in each census tract as the size of the red bubbles. In Panel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

Figure 3: Heterogeneous effects on employment by distance from previous job



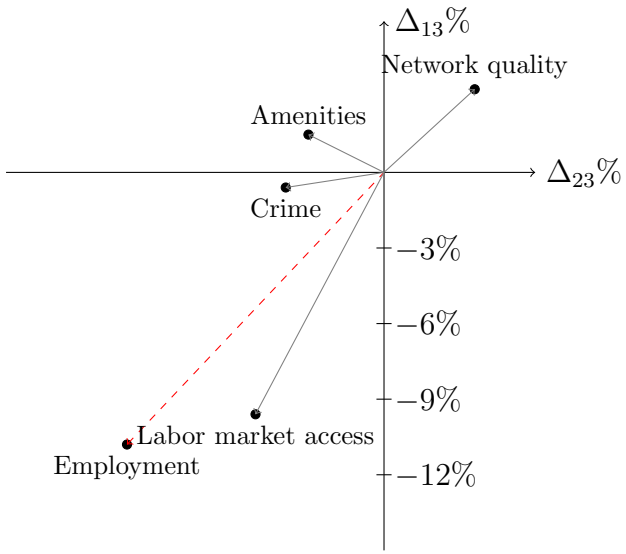
Note: This figure shows estimates and 90% confidence intervals based on the main sample. The dependent variable is a dummy for being formally employed at any point between 2015 and 2017. Bins are defined according to the median baseline distance (measured in kilometers) between the last baseline formal job and the *MCMV* housing projects. We include controls for gender, race, schooling, and baseline formal employment. Standard-errors are clustered at the individual level.

Figure 4: Heterogeneous effects on employment by distance from previous home



Note: This figure shows estimates and 90% confidence intervals based on the main sample. The dependent variable is a dummy for being formally employed at any point between 2015 and 2017. Bins are defined according to the median baseline distance (measured in kilometers) between the individuals' home and *MCMV* housing projects. We include controls for gender, race, schooling, and baseline formal employment. Standard-errors are clustered at the individual level.

Figure 5: Relative treatment effects vs. differences in mechanisms



Notes: Percent changes in employment outcomes and potential mechanisms for individuals drafted to Neighborhood 1 vs. 3 in the vertical axis. Difference between Neighborhoods 2 and 3 depicted in the horizontal axis.

Table 1: Descriptive statistics and balancing tests

	Main sample		Disadvantaged sample	
	Control Mean	Drafted Diff.	Control Mean	Drafted Diff.
Take-up	0	0.553***	0	0.506***
		(0.010)		(0.019)
Panel A: Socioeconomic characteristics				
White	0.38	0.004	0.27	0.007
		(0.010)		(0.017)
Male	0.45	-0.010	0.200	0.005
		(0.010)		(0.015)
Schooling	3.82	0.001	3.34	0.005
		(0.015)		(0.030)
Panel B: Labor market characteristics				
Ever employed	1	0	0.661	-0.019
				(0.018)
Hours	41.84	0.015	42.38	0.039
		(0.100)		(0.118)
Wages	1341.12	-43.391	656.68	15.497
		(36.378)		(39.328)
Tenure	48.00	1.259	24.25	-0.430
		(1.571)		(1.787)
Distance to previous job (km)	47.25	0.170	45.28	0.513
		(0.831)		(1.356)
Joint test (p-value)	0.70 (0.65)		0.94 (0.46)	
Observations	503,884		264,537	

Note: *White* and *Male* are dummy variables; *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* denotes the number of monthly hours and *wages* represent the average monthly compensation registered in contract. *Tenure* is measured in number of months. *Distance to previous job* is the median distance between the address of the last employment of the individual in the baseline period and the housing projects. Robust standard-errors for the main sample and clustered standard-errors at the individual level for the disadvantaged sample. We report the F statistic and its respective p-value for the joint test that all differences in means are equal to zero. * p<0.1, ** p<0.05, and *** p<0.01.

Table 2: Neighborhoods Characteristics

	Neighborhood 1	Neighborhood 2	Neighborhood 3
Income	32	18	6
Education	30	15	6
Single parent households	76	83	39
Labor Market Access	15	17	33
Formal jobs (RAIS)	29	27	26
Amenities	20	32	25
Crime rates	10	13	14

Notes: Variables are represented in percentiles of their distribution across all Rio de Janeiro's neighborhoods. *Income* is a summary of different measures of income. *schooling* is a five-level index indicating the highest level of instruction that ranges from no schooling to post-graduation. *Single parent households* is the fraction of households with children and only one parent (either father or mother). *Labor Market Access* is the average of the distance of all postal codes in the census tract to the formal jobs in the baseline period, weighted by the number of formal jobs at each address and the number of individuals living at each postal code (see Expression 1). *Formal jobs* is the number of formal jobs at each Census tract. *Amenities* is a summary measure of public parks, schools, and day care centers located in each census tract. *Crime rates* is a summary measure of robbery, burglary, and homicides. All variables are reported as percentiles of the city's distribution across neighborhoods, such that lower values imply worse outcomes.

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

	TOT	Control mean
Panel A: Housing characteristics		
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 (percentiles)		
Income	-9.537***	37.39
	(0.826)	
Education	-1.563	21.65
	(2.35)	
Single parent households	0.626	50.08
	(2.287)	
Market access	-3.333***	25.23
	(1.123)	
Formal jobs (RAIS)	-5.328***	36.765
	(0.777)	
Amenities	-1.068	28.32
	(2.222)	
Crimes	-6.068***	25.32
	(2.234)	

Notes: Data from the disadvantaged sample. In Panel A, variables are defined at the individual-level, rent and transportation expenditures are reported as nominal values in R$ at the moment of the Single Registry update. We deflate nominal prices using the Consumer Price Index (IPCA) for January of 2021. In Panel B, variables are defined at the neighborhood level and are reported as percentiles of Rio de Janeiro's neighborhood distribution for the specific variables. *Income* is a summary of different measures of income. *schooling* is a five-level index indicating the highest level of instruction that ranges from no schooling to post-graduation. *Single parent households* is the fraction of households with children and only one parent (either father or mother). *Market access* is the average of the distance of all postal codes in the census tract to the formal jobs in the baseline period, weighted by the number of formal jobs at each adress and the number of individuals living at each postal code. *Formal jobs* is the number of formal jobs at each Census tract. *Amenities* is a summary measure of public parks, schools, and day care centers located in each census tract. *Crime rates* is a summary measure of robbery, burglary, and homicides. All variables are normalized such that lower values imply worst outcomes. Clustered standard-errors at the individual level are shown in parenthesis. * p<0.1, ** p<0.05, and *** p<0.01.

Table 4: *MCMV* Program Effects – Main sample

	Control mean	ITT	TOT	Number of observations
Panel A: Employment outcomes				
Formal employment	0.63	-0.009*	-0.017*	1,511,631
		(0.005)	(0.009)	
Number of months employed	5.91	-0.108*	-0.195*	1,511,631
		(0.572)	(0.114)	
Panel B: Job quality and wages				
Log(wages)	7.59	-0.009	-0.017	749,633
		(0.010)	(0.018)	
Employer rank	51.19	0.395	0.699	749,633
		(0.625)	(1.106)	
Occupation rank	49.74	0.439	0.776	749,632
		(0.664)	(1.175)	
Panel C: Employment adjustments				
Kept the same job	0.73	-0.024**	-0.041**	344,307
		(0.011)	(0.020)	
Distance from job to housing projects (km)	45.01	-0.851*	-1.570*	344,307
		(0.447)	(0.825)	
Worked in a neighboring municipality	0.13	0.012**	0.022**	344,307
		(0.006)	(0.012)	
First-stage F test			3083.36	

Notes: Estimates from the main sample. Observations are stacked for the endline years (2015 to 2017). Panel A includes all individuals that have been formally employed. Panel B includes individuals employed at the endline period. Panel C includes only individuals employed in 2016 and 2017. *Formal employment* is a dummy for being formally employed at any point in 2015-2017. *Number of months employed* is the total number of months in each year individuals were employed, zero if not employed. *Wages* denotes real average wages individuals received during the year while employed, conditional on being employed. *Real wages* are obtained using the Consumer Price Index (IPCA) for January of 2021. See text for *Firm rank* and *Occupation rank*. *Kept the same job* is a dummy for individuals that stay linked to the same establishment as in the baseline period. *Distance from job to housing project* is the median distance, measured in kilometers, from the adress reported by the firm and the *MCMV* housing projects. *Worked in a neighboring municipality* is a dummy for individuals linked to firms with addresses that are not contained in Rio de Janeiro municipality. We include controls for gender, 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: *MCMV* Program Effects – Disadvantaged sample

	Control mean	ITT	TOT	Number of observations
Panel A: Employment outcomes				
Formal employment	0.14	0.011*	0.023*	669,317
		(0.006)	(0.012)	
Number of months employed	2.77	-0.068*	-0.135*	669,317
		(0.037)	(0.075)	
Panel B: Wages and welfare dependency				
Log(wages)	6.69	0.037	0.070	94,769
		(0.025)	(0.48)	
Bolsa Família recipient	0.48	-0.030**	-0.060**	669,137
		(0.013)	(0.027)	
First-stage F test			717.77	

Note: Estimates from the disadvantaged sample. Observations are stacked for the endline years (2015 to 2020). *Formal employment* is a dummy for being formally employed at any point in 2015-2017. The variable *monthly hours* measures the number of contractual hours formally employed, and zero if not employed. *Number of months employed* is the total number of months in each year individuals were employed, zero if not employed. *Wages* denotes real average wages individuals received during the year while employed, conditional on being employed. *Real wages* are obtained using the Consumer Price Index (IPCA) for January of 2021. *Bolsa Família receipient* is a dummy for households that report receiving any transfer from the Bolsa Família program. We include controls for gender, 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 6: Treatment effects by neighborhood

	Employment	# of months employed	Log(wages)	Bolsa Família beneficiary	First-stage F test
Panel A: Main sample					
Neighborhood 1	-0.069*	-0.435	-0.033	–	350.26
	(0.042)	(0.532)	(0.072)	–	
Neighborhood 2	-0.057*	-0.732**	-0.019	–	937.80
	(0.029)	(1.029)	(0.045)	–	
Neighborhood 3	0.011	0.072	0.004	–	1947.38
	(0.019)	(0.222)	(0.033)	–	
Equality coeff. (p-value)	0.07	0.12	0.86	–	
Equality coeff N1 and N2 vs N3 (p-value)	0.02	0.06	0.59	–	
Observations	1,511,652		749,632		
Panel B: Disadvantaged sample					
Neighborhood 1	0.010	-0.299	0.047	-0.099	148.64
	(0.024)	(0.610)	(0.123)	(0.070)	
Neighborhood 2	-0.040*	-0.805**	0.001	-0.008	653.23
	(0.026)	(0.334)	(0.080)	(0.047)	
Neighborhood 3	0.032**	0.847**	0.095	-0.075**	1116.85
	(0.016)	(0.347)	(0.075)	(0.034)	
Equality coeff. (p-value)	0.06	0.01	0.20	0.37	
Equality coeff N1 and N2 vs N3 (p-value)	0.04	0.02	0.35	0.92	
Observations			669,137		

Notes: Estimates from the main sample (Panel A) and disadvantaged sample (Panel B). *Formal employment* is a dummy for being formally employed at any point in 2015-2017. *Number of months employed* is the total number of months in each year individuals were employed, zero if not employed. *Wages* denotes real average wages individuals received during the year while employed, conditional on being employed. Real wages are obtained using the Consumer Price Index (IPCA) for January of 2021. *Bolsa Família recipient* is a dummy for households that report receiving any transfer from the Bolsa Família program. We include controls for gender, race, and schooling. We also show two Wald tests for the joint equality of treatment effects across neighborhoods, and and for the joint equality of pooled treatment effects for neighborhoods 1 and 2 vs neighborhood 3. Clustered standard errors at the individual level are shown in parenthesis: * p<0.1, ** p<0.05, and *** p<0.01.

Table 7: Balancing tests for compliers – Main Sample

	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
Distance from job to housing project	46.99	46.33	47.01	0.87
Joint Hypothesis (p-value)		0.31		

Notes: Data from the main sample. *Received house* is a dummy for treated individuals that were and excluded from the set of *MCMV* potential beneficiaries. *Received house (excluding the wait-list)* is the same as above, but restricting the sample to individuals that were offered the house directly. *White* and *Male* are dummy variables; *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* indicates the number of monthly hours registered in contract in the last year individuals' held a formal job; *wages* are the average monthly compensation registered in the contract; *tenure* is measured in months. *Distance to previous job* is the median distance between the address of the last employment of the individual in the baseline period and the housing projects The joint test shows the 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 8: Balancing tests for compliers – Disadvantaged Sample

	Neighborhood 1	Neighborhood 2	Neighborhood 3	Joint p-value for differences
Panel A: Take-up				
Received house	0.42	0.56	0.62	0.00
Received house (excluding wait list)	0.62	0.56	0.60	0.76
Panel B: Socioeconomic characteristics				
White	0.34	0.29	0.26	0.42
Male	0.23	0.27	0.24	0.67
Schooling	3.25	3.51	3.14	0.14
Distance from home to housing project	45.92	44.31	42.86	0.45
Panel C: Labor market characteristics (conditional on employment)				
Ever employed	0.66	0.75	0.68	0.80
Hours	43.01	42.02	42.27	0.28
Wages	829.24	662.89	668.31	0.78
Tenure	26.15	24.93	21.83	0.78
Distance from job to housing project	38.51	36.12	37.01	0.66

Note: Notes: Data from the disadvantaged sample. *Received house* is a dummy for treated individuals that were and excluded from the set of *MCMV* potential beneficiaries. *Received house (excluding the wait-list)* is the same as above, but restricting the sample to individuals that were offered the house directly. *White* and *Male* are dummy variables; *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* indicates the number of monthly hours registered in contract in the last year individuals' held a formal job; *wages* are the average monthly compensation registered in the contract; *tenure* is measured in months. textitDistance to previous job is the median distance between the address of the last employment of the individual in the baseline period and the housing projects The joint test shows the 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 9: Differences in house prices and rents across neighborhoods

	All (1)	*MCMV* Similar (2)	*MCMV* houses (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	–

Notes: Panel A shows estimates using data scraped from housing sites for home' selling prices. Panel B shows estimates using rent data from the 2010 Census. Results for the full sample of houses (Column 1), houses with similar characteristics to *MCMV* units (Column 2), and restricted to *MCMV* houses (Column 3). Regressions include controls for square footage, number of rooms, number of bedrooms, and dummies for the presence of garage and lobby in columns (1) and (2). Robust standard-errors are shown in parenthesis: * p<0.1, ** p<0.05, and *** p<0.01.

Table 10: Estimated Bounds

	Lower bound LB^{β_m}	Upper bound UB^{β_m}
Fraction explained	0.821	0.934
by labor market access	(0.051)	(0.046)

Notes: Fraction of total employment effects explained by labor market access is calculated as $f^m = \frac{\beta^m * (\Delta LMA_1 + \Delta LMA_2)}{\theta_1 + \theta_2}$. Both bounds and robust standard-errors are estimated using the MS2SLS (see text).

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. Fernandez-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 (2017, 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.

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.

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.

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

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

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.

Hansen, B. (2022). *Econometrics*. Princeton University Press.

Harding, D. J., L. Sanbonmatsu, G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, M. Sciandra, and J. Ludwig (2021). Evaluating contradictory experimental and nonexperimental estimates of neighborhood effects on economic outcomes for adults. *Housing Policy Debate*, 1–34.

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

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.

Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu (2013). Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American economic review 103*(3), 226–31.

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

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

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.

Pinto, R. (2021). Beyond intention to treat: Using the incentives in moving to opportunity to identify neighborhood effects. Technical report, Working paper.

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.

Turney, K., S. Clampet-Lundquist, K. Edin, J. R. Kling, G. J. Duncan, J. Ludwig, and J. K. Scholz (2006). Neighborhood effects on barriers to employment: Results from a randomized housing mobility experiment in baltimore [with comments]. *Brookings-Wharton papers on urban affairs*, 137–187.

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

Vicenty, T. (1975). Geodetic internal solution between antipodal points. *Working paper*.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association 57*(298), 348–368.

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

APPENDIX

A Data Appendix

A.1 Additional data description

This section provides additional details about the three additional data sources used to characterize neighborhoods: city hall records, the Rio de Janeiro Public Security Institute (ISP-RJ) statistics, and scraped data from real state websites.

First, the City Hall maintains registries of public facilities and their respective addresses. We collected data for daycare, parks, schools, and health facilities including hospitals and emergency care units. Not all of them are run by the municipal government, as some fall under the state government's responsibility. We georeference all facilities to construct our measures of amenities at the census tract level.

The ISP-RJ is an institute linked to the State's Public Security Secretary. It is responsible for collecting data and performing statistical analyses to inform public policies to prevent crime. It provides monthly crime statistics, which are produced by collecting information on crimes that are provided by all police stations, including the location of criminal activities. We aggregate statistics at the census tract level 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.

Finally, we also collect data for houses and apartments being sold in Rio de Janeiro. We scrape data from the most common websites used to sell real state in Brazil: https://www.zapimoveis.com.br/, https://www.loft.com.br, and https://www.vivareal.com.br/. We collect the data for house prices in February of 2022. We keep in the sample only properties in the same neighborhoods as *MCMV* housing projects and drop duplicate ads in more than one website. We also collect information on houses' characteristics, such as total area, number of rooms, number of bathrooms, among others. Despite not being allowed to do so, some individuals would attempt to sell their houses in *MCMV* housing projects. We can identify those because in some cases the homeowner explicitly advertises the house as being from the *MCMV* program, or by using a combination of address and house characteristics.

A.2 Variables Description

Table A.1 lists all variables used in the paper and provides a brief description. The exceptions are take-up, which we describe here, income, crime and amenities that are discussed in Section 2.2. There is no variable in the administrative data that indicates whether individuals take up the offer of a house or not. However, according to the rules of the program individuals who won the lottery but have not received a house should be automatically added to the set of potential beneficiaries in future lotteries. We use this information to measure take-up, which is defined as individuals who are drafted in a given lottery, and do not appear in subsequent ones.

A.3 Validation of Single Registry data

We use data from the Single Registry to construct the outcomes for the disadvantaged sample. One potential concern is that this is self-reported information and might not be as

accurate as the administrative data used in the main sample.

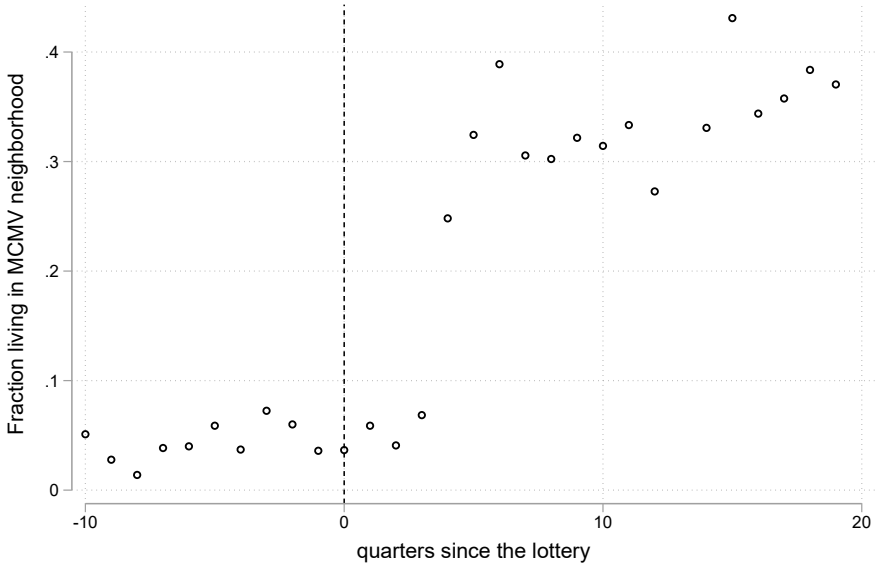
In order to validate the Single Registry data, we reshape the RAIS data to the monthly level and merge updates reported in the Single Registry with RAIS administrative data for the years of 2015 to 2017. Then, we compare information self-reported in the Single Registry with administrative records.

Table A.2 summarizes the results. We find that the large majority of individuals correctly report their employment status: around 95 percent of individuals correctly report being formally employed or not. Only 3 percent of the sample reports being employed while not having a register in RAIS, and 2 percent reports not being employed while having a formal contract in RAIS.

B Take-up Patterns

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 presents regressions of dummies for living in a given *MCMV* neighborhood on dummies that indicate being drafted to specific neighborhoods. The results show that the allocation process was strict, and being drafted to a specific neighborhood only predicts that drafted individuals were living in that neighborhood. Table B.2 shows that, for drafted individuals that were not in the wait-list, treatment take-up positively correlates with firm size, and negative correlates with the schooling and dummies for disability and white race.

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.

C Additional Results

This section presents several additional results. Table C.1 shows the year-by-year ITT and TOT on formal employment. 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.

In Table C.2, we show alternative estimates for the main sample, focusing on ITT estimates of the effect on the probability of being formally employed. 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.

Table C.3 shows the effects of the *MCMV* program on formal employment probability separately for 1) men and women; 2) white and non-white individuals; and 3) high and low skill individuals. We find that the program's effects are concentrated on high skill and white individuals, and that there are no significant differential impacts by gender. Table C.4 examines whether the program had an effect on informal employment.

Finally, we explore whether the effects can be explained by the disruption of beneficiaries' social networks. If this is the case, we expect to find that individuals that lived or worked farther away from the housing projects to be more disrupted and to experience worst impacts of the program on their labor market outcomes. In Figures **??** and **??**, we estimate semi-parametric effects of receiving a house on formal employment probability, relative to the baseline distance to their previous job and home. In the former, we use the main sample and in the latter we use the disadvantaged sample. Since we are breaking the sample into bins, there is larger potential for idiosyncratic baseline differences in employment to affect the analysis. For this reason, we control for the baseline formal employment, similarly to the ANCOVA specification described above. We can see no discernible effects heterogeneous effects of the program on employment, based on either distance considered. In Table C.5, we impose a linear parametric specification and formally show that the program has no statistically significant heterogeneous treatment effects on either distance.

D ExtrapoLATE

Table 7 shows that the characteristics of different groups of compliers are similar. In this section, we use the model suggested by Angrist and Fernandez-Val (2010) to formally show that differences in the characteristics of the compliers cannot explain differences in neighborhood-specific treatment effects.

D.1 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 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 6, 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:

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

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:

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

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.

D.2 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 high school, 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 Fernandez-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.

D.3 Extrapolation results

In Table D.3, we show in Panel A the LATE of being drafted by to neighborhoods 1 to 3, identically, to Table 6. 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.

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 The Multi-Sample Two-Stage Least Squares (MS2SLS) estimator

E.1 Setup

Consider that we are interested in the following structural relation:

$$y_i = X_i * \beta + \epsilon_i$$

where X is composed of p different variables. We have two challenges in the estimation of β. The first one is that X is endogenous. Thus, OLS estimation of the relation of interest will yield inconsistent estimates. Regarding this problem, assume that we have $Q \geq p$ instruments available (Z).

The second challenge is that we do not jointly observe y_i, X_i, and Z_i in the same data. In one sample, which we will call main sample, we observe $\{y_{iM}, Z_{iM}\}$, $i = 1, ..., n_M$. Also, suppose that each different endogenous variable is available in a different sample jointly with the vector of instruments. That is, we observe samples $\{x_{ij}, Z_{ij}\}, j = 1, ..., p,\ i = 1, ..., n_p$, where $X_i = [x_{i1}, ..., x_{ip}]$, which we call auxiliary samples. Let

Note that this setup encompasses the framework of the Two-Sample Two-Stage Least Squares (TS2SLS) estimator framework, analyzed by Angrist et al. (1996), and Inoue and Solon (2010), Pacini and Windmeijer (2016).

E.2 Estimator

In order to tackle the problem, it is useful to define the vector of errors $\varepsilon = [\varepsilon_1, ..., \varepsilon_p]^{'}$ and the matrix:

$$\overline{Z} = \begin{bmatrix} Z_1 & 0 & 0 & \dots & 0 \\ 0 & Z_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & Z_p \end{bmatrix}$$

with dimensions n_a x $p * Q$, where $n_a = n_1 + ... + n_p$.

By taking advantage of this notation, we can write the first-stage of the problem as linear system of equations (as in Hansen (2022), chapter 11). Then, we have that:

$$X = \overline{Z} * \gamma + \varepsilon$$

where γ is a Q x p vector of coefficients.

The first-stage coefficients can be estimated by least squares. We can write:

$$\hat{\gamma} = (\overline{Z}^{'}\overline{Z})^{-1} * \overline{Z}^{'}X \tag{17}$$

It also can be shown that the variance of the first-stage coefficients is:

$$V_\gamma = (\overline{Z}^{'}\overline{Z})^{-1}\Omega_\gamma(\overline{Z}^{'}\overline{Z})^{-1} \tag{18}$$

where:

$$\Omega_\gamma = \mathbb{E}[\overline{Z}_i^{'}\varepsilon_i\varepsilon_i^{'}\overline{Z}_i]$$

Note that the vector of instruments Z_{ij} is the same in all equations of the linear system. This implies that there are no efficiency gains in implementing the SUR methodology proposed by Zellner (1962)[24].

It is also useful to explicitly write the reduced-form for the model as:

$$y_{iM} = \theta * Z_{iM} + u_{iM} \tag{19}$$

where the vector of coefficients $\hat{\theta}$ can also be estimated by least squares:

$$\hat{\theta} = (Z_M^{'} Z_M)^{-1} Z_M * y_M \tag{20}$$

Also,

$$V_\theta = \left(Z_M^{'} Z_M\right)^{-1} \Omega_\theta \left(Z_M^{'} Z_M\right)^{-1}$$

where:

$$\Omega_\theta = \mathbb{E}[Z_{iM}^{'} u_{iM} u_{iM}^{'} Z_{iM}]$$

Our proposed estimator is straightforward. We first suggest estimating $\hat{\gamma}$ as above. Next, we use the vector of first-stage coefficients to generate predicted values in the main sample ($\hat{X}_m$). Finally, we can estimate the vector of coefficients of interest $\hat{\beta}$ as a regression of y_m on $\hat{X}_m$.

That is, we can write the estimator as:

$$\hat{\beta}_{MS2SLS} = (\hat{X}_M^{'} * \hat{X}_M)^{-1} \hat{X}_M^{'} * y_M$$

which is the familiar expression for the TS2SLS estimator. However, note that, in our context, the predicted values in the main sample is written as:

$$\hat{X}_M = Z_M * \hat{\gamma} = Z_M * (\overline{Z}^{'} \overline{Z})^{-1} \overline{Z}^{'} * X$$

Therefore, we can write the estimator as:

$$\hat{\beta}_{MS2SLS} = \left(\hat{\gamma}^{'} Z_M^{'} Z_M \hat{\gamma}\right)^{-1} \hat{\gamma}^{'} Z_M^{'} * y_M = \left(\hat{\gamma}^{'} Z_M^{'} Z_M \hat{\gamma}\right)^{-1} \hat{\gamma}^{'} Z_M^{'} Z_M \hat{\theta} \tag{21}$$

or, alternatively, as:

$$\hat{\beta}_{MS2SLS} = \left(X \overline{Z}^{'} (\overline{Z}^{'} \overline{Z})^{-1} Z_M^{'} Z_M (\overline{Z}^{'} \overline{Z})^{-1} * \overline{Z}^{'} X\right)^{-1} X \overline{Z}^{'} (\overline{Z}^{'} \overline{Z})^{-1} Z_M^{'} * y_M$$

E.3 Properties

We make the following assumptions:

Assumption 1: The samples $\{y_M, Z_M\}$, $\{x_j, Z_j\}, j = 1, ..., p$ are jointly independent and have defined fourth moments.

Assumption 2: $\mathbb{E}[Z_{iM}^{'} Z_{iM}] = Q_{ZZM}$ and $\mathbb{E}[\overline{Z}_i^{'} \overline{Z}_i] = Q_{\overline{ZZ}}$ are nonsingular.

Assumption 3: $E[Z_M * \epsilon] = 0$ and $E[Z_j * \varepsilon_j] = 0, j = 1, ..., p$.

[24]For a proof that the SUR model is equivalent to a system OLS estimation when covariates are all equal, see Hansen (2022)

Assumption 4: $E[u_i^2 M * Z_{iM} Z_{iM}^{'}] = \Omega_y$ and Ω_γ are finite and positive definite matrices.

Assumption 5: $\lim_{n_M \to \infty, n_a \to \infty} \frac{n_M}{n_a} = \alpha$.

Next, based on these assumptions, we develop some properties on the MS2SLS estimator. This estimator is similar to the TS2SLS estimator, so we follow the argument of Pacini and Windmeijer (2016) closely. First, we discuss the consistency of the estimator.

The assumptions described assure that:

$$\hat{\gamma} \to \mathbb{E}[\overline{Z}_i^{'} \overline{Z}_i] * \mathbb{E}[\overline{Z}_i^{'} X_i^{'}] = \gamma$$

and

$$\hat{\theta} \to \mathbb{E}[Z_{Mi}^{'} Z_{Mi}] * \mathbb{E}[Z_{Mi}^{'} y_i] = \theta$$

Therefore, the estimator is consistent:

$$\begin{aligned} \hat{\beta}_{MS2SLS} \to plim\left(\left(\frac{1}{n_M}\hat{\gamma}^{'} Z_M^{'} Z_M \hat{\gamma}\right)^{-1} \frac{1}{n_M}\hat{\gamma}^{'} Z_M^{'} Z_M \hat{\theta}\right) = \\ = (\gamma^{'} Q_{ZZM} \gamma)^{-1} \gamma^{'} Q_{ZZM} * \theta = \beta \end{aligned}$$

Now, we consider the limiting distribution of the estimator. By writing the estimator and using the structural relations between equations, we have that:

$$\begin{aligned} \hat{\beta}_{MS2SLS} = (\hat{X}_M^{'} * \hat{X}_M)^{-1} \hat{X}_M^{'} * y_M = \beta + (\hat{X}_M^{'} * \hat{X}_M)^{-1} \hat{X}_M^{'} * (u_{iM} - \theta * Z_{iM}) = \\ = \beta + (\hat{X}_M^{'} * \hat{X}_M)^{-1} \hat{X}_M^{'} * (u_{iM} - Z_{iM}(\hat{\gamma} - \gamma)\beta) \end{aligned}$$

Therefore:

$$\sqrt{n_M}(\hat{\beta}_{MS2SLS} - \beta) = \left(\frac{1}{n_M}\hat{X}_M^{'} * \hat{X}_M\right)^{-1} \underbrace{\hat{X}_M^{'} * \frac{1}{n_M}(u_{iM} - Z_{iM}(\hat{\gamma} - \gamma)\beta)}$$

Note that the term highlighted in the equation above can be written as:

$$\begin{aligned} \hat{X}_M^{'} * \frac{1}{n_M}(u_{iM} - Z_{iM}(\hat{\gamma} - \gamma)\beta) = \hat{\gamma}^{'} \overline{Z}^{'} * \frac{1}{n_M}(u_{iM} - Z_{iM}(\hat{\gamma} - \gamma)\beta) = \\ \hat{\gamma}^{'}\left(\frac{1}{n_M}\overline{Z}^{'} * \overline{Z}\right)(\sqrt{n_M}(\theta - \hat{\theta}) - \sqrt{n_M}\underbrace{(\hat{\gamma} - \gamma)\beta}) \end{aligned}$$

Once again, we can rewrite the highlighted term as:

$$(\hat{\gamma} - \gamma)\beta = (\beta^{'} \otimes I_Q) vec(\hat{\gamma} - \gamma)$$

By combining the equations above, we have that:

$$\sqrt{n_M}(\hat{\beta}_{MS2SLS} - \beta) = \left(\frac{1}{n_M}\hat{X}_M^{'} * \hat{X}_M\right)^{-1} \hat{\gamma}^{'}\left(\frac{1}{n_M}\overline{Z}^{'} * \overline{Z}\right) \left(\sqrt{n_M}(\theta - \hat{\theta}) - \sqrt{n_M}(\beta^{'} \otimes I_Q)vec(\hat{\gamma} - \gamma)\right)$$

Now, let $\Delta = (\theta^{'} \quad \gamma^{'})^{'}$, $\hat{\Delta} = (\hat{\theta}^{'} \quad \hat{\gamma}^{'})^{'}$, and $\delta = (1 - \beta^{'})^{'}$. Then:

$$\sqrt{n_M}(\hat{\beta}_{MS2SLS} - \beta) = \left(\frac{1}{n_M}\hat{X}_M^{'} * \hat{X}_M\right)^{-1} \hat{\gamma}^{'}\left(\frac{1}{n_M}\overline{Z}^{'} * \overline{Z}\right) (\delta^{'} \otimes I_Q)\sqrt{n_M}(\Delta - \hat{\Delta})$$

Note that:

$$\left(\frac{1}{n_M}\hat{X}_M^{'} * \hat{X}_M\right)^{-1} \hat{\gamma}^{'}\left(\frac{1}{n_M}\overline{Z}^{'} * \overline{Z}\right) \rightarrow (\gamma^{'} Q_{ZZM} \gamma)^{-1} * \gamma^{'} Q_{ZZM} = C$$

Then:

$$\sqrt{n_M}(\hat{\beta}_{MS2SLS} - \beta) \xrightarrow{d} N(0, V_\beta)$$

where:

$$V_\beta = C(\delta^{'} \otimes I_Q)V_\theta(\delta \otimes I_Q)C^{'}$$

Let $\tilde{\Omega}_\gamma$ be the variance-covariance matrix compatible with $vec(\gamma)$ with dimensions $Q * p$ x $Q * p$. We can write its variance as:

$$\tilde{V}_\gamma = (I_p \otimes \overline{Z}^{'}\overline{Z})\tilde{\Omega}_\gamma(I_p \otimes \overline{Z}^{'}\overline{Z})$$

Then, we can write:

$$V_\theta = \begin{bmatrix} V_\theta & 0 \\ 0 & \alpha\tilde{V}_\gamma \end{bmatrix}$$

Finally,

$$V_\beta = C(V_\theta + \alpha(\beta^{'} \otimes C)\tilde{V}_\gamma(\beta^{'} \otimes C))C^{'} = \\ = CV_\theta C^{'} + \alpha(\beta^{'} \otimes C)\tilde{V}_\gamma(\beta \otimes C^{'}))$$

Thus, if we obtain consistent estimates of $Var(\hat{\theta})$ and $Var(vec(\hat{\gamma}))$, we can consistently estimate $Var(\hat{\beta}_{MS2SLS})$ using a simple plug-in estimator as:

$$\hat{Var}(\hat{\beta}_{MS2SLS}) = \hat{C} * \hat{Var}(\hat{\theta}) * \hat{C}^{'} + (\hat{\beta}^{'}_{M2SLS} \otimes \hat{C}) * \hat{Var}(vec(\hat{\gamma})) * (\hat{\beta}_{M2SLS} \otimes \hat{C}^{'})$$

where $\hat{C}$ is the matrix of coefficients from the regression of Z_M on $\hat{X}_M$.

E.4 The just identified case

So far, we have derived the estimator for the general case that: $Q \geq p$. However, for several applications, including the one in this paper, the number of instruments is equal to the number of endogenous variables ($Q = p$). In the case where X and Z have the same

dimensions, we can simplify the general estimator to:

$$\hat{\beta}_{MS2SLS} = \left(X\overline{Z}^{'}(\overline{Z}^{'}\overline{Z})^{-1}Z_M^{'}Z_M(\overline{Z}^{'}\overline{Z})^{-1} * \overline{Z}^{'}X\right)^{-1}X\overline{Z}^{'}(\overline{Z}^{'}\overline{Z})^{-1}Z_M^{'} * y_M =$$
$$= (\overline{Z}X)^{-1}(\overline{ZZ})(Z_M^{'}Z_M)^{-1}(\overline{ZZ})(X\overline{Z})^{-1}(X\overline{Z})(\overline{Z}^{'}\overline{Z})^{-1}Z_M^{'} * y_M =$$
$$= (\overline{Z}X)^{-1}(\overline{ZZ})(Z_M^{'}Z_M)^{-1}Z_M^{'} * y_M = \gamma^{-1}\theta$$

In this case, the estimator for the variance is also simplified. Note that:

$$C = (\gamma^{'}Q_{ZZM}\gamma)^{-1}\gamma^{'}Q_{ZZM} = \gamma^{-1}Q_{ZZM}\gamma^{-1} * \gamma * Q_{ZZM} = \gamma^{-1}$$

Then, we can estimate the variance of the estimator as:

$$\hat{Var}(\hat{\beta}_{MS2SLS}) = \hat{\gamma}^{-1} * \hat{Var}(\hat{\theta}) * \hat{\gamma}^{-1} + (\hat{\beta}_{M2SLS}^{'} \otimes \hat{\gamma}^{-1}) *$$
$$\hat{Var}(vec(\hat{\gamma})) * (\hat{\beta}_{M2SLS} \otimes \hat{\gamma}^{-1})$$

F Mechanisms Analysis

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

F.1 Obtaining bounds for β^m

Let $\boldsymbol{\theta}$ be the vector of reduced-form 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 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: $\beta_n = 0$.

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 and is completely driven by the patterns in Figure 5.

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:

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

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

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

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 $\beta^a \in \mathbb{R}_{>0}$. It follows that the lower bound for β_m is such that:

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

Similarly, the upper bound fro β_m is such that:

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

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:

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

and

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

Alternatively, we show in the next section that we can estimate these bounds using a Multi-Sample Two- Stage Least squares (MS2SLS), derived in Appendix E.

F.2 Equivalence of UB^{β_m} and LB^{β_m} to the MS2SLS 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 Multi-Sample Two-Stage Least squares estimator.

As discussed in the main text, the structural model of interest can be described as:

$$\Delta y_{in} = \beta \Delta X_n + \epsilon_{in}$$

where $X_n = \{MA_n, Cr_n\}$ is a restriction of the complete set of mechanisms. However, we do not jointly observe all relevant variables. In sample main sample, we observe $\{y_{in}^M, \boldsymbol{D}_{in}^M\}$,

for $i = 1, ..., n_M$. The vector $\boldsymbol{D}_{in}^M = \begin{bmatrix} D_{i1}^M & D_{i2}^M \end{bmatrix}$ includes dummies for being drafted for neighborhoods 1 and 2.

Additionally, we observe two auxiliary samples. In the first auxiliary sample, we observe $\{MA_{in}, \boldsymbol{D}_{in}^1\}$, for $i = 1, ..., n_1$. In the second auxiliary sample, we observe $\{MA_{in}, \boldsymbol{D}_{in}^2\}$, for $i = 1, ..., n_2$. That is, in each of these auxiliary samples, we observe one potential mechanism (market access and crime rates, respectively) and the vector $\boldsymbol{D}_{in}^j = \begin{bmatrix} D_{i1}^j & D_{i2}^j \end{bmatrix}$, for $j = 1, 2$, contains dummies that measurement i of the available mechanisms was located in neighborhoods one and two. We allow a different number of measurements across samples, such that $n_a = n_1 + n_2$.

In the main sample, we can estimate reduced-form effects of the program on employment as:

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = (D^{M\prime} D^M)^{-1} D^{M\prime} y^M$$

We can stack data for auxiliary samples. Let the matrix of instruments be:

$$\overline{Z} = \begin{bmatrix} \boldsymbol{D}^1 & 0 \\ 0 & \boldsymbol{D}^2 \end{bmatrix}$$

Thus, we can write a stacked first-stage equation as:

$$X = \overline{Z} * \gamma + \epsilon$$

The matrix of instruments is entirely constituted of neighborhood dummies. Thus, first-stage coefficients represent average differences in market access and crime rates for neighborhoods 1 and 2, relative to neighborhood 3:

$$\gamma = \begin{bmatrix} \Delta MA_1 & \Delta MA_2 \\ \Delta Cr_1 & \Delta Cr_2 \end{bmatrix}$$

The model of interest has two endogenous mechanisms and two instruments, so that it is just-identified. We have shown in Appendix E that, in this case, we can consistently estimate the parameters as:

$$\beta^{MS2SLS} = \gamma^{-1}\theta$$

By inverting the matrix of first-stage coefficients γ, we can write:

$$\boldsymbol{\beta}^{MS2SLS} = \frac{1}{\Delta \hat{MA}_1 * \Delta \hat{Cr}_2 - \Delta \hat{MA}_2 * \Delta \hat{Cr}_1} * \begin{bmatrix} \Delta \hat{Cr}_2 & -\Delta \hat{Cr}_1 \\ -\Delta \hat{MA}_2 & \Delta \hat{MA}_1 \end{bmatrix} * \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}$$

Finally, note that the first element of the $\boldsymbol{\beta}^{MS2SLS}$ matrix is equivalent to the upper bound for β_m:

$$\hat{\boldsymbol{\beta}}^{MS2SLS}[1,1] = \frac{\hat{\theta}_1 * \Delta \hat{Cr}_2 - \hat{\theta}_2 * \Delta Cr_1}{\Delta \hat{MA}_2 * \Delta \hat{Cr}_1 - \Delta \hat{MA}_1 * \Delta \hat{Cr}_2} = \hat{UB}^{\beta_m}$$

Similarly, we can estimate the lower bound for β^m (LB^{β_m}) in a MS2SLS regression of Δy_{in} on MA_n and Am_n and instruments similarly as above.

This equivalence is very useful because it provides a natural way to conduct inference on the bounds of β_m. We have shown in Appendix E that, when the model is just-identified, we

can estimate the variance of $\hat{\boldsymbol{\beta}}$ as:

$$\hat{Var}(\hat{\beta}_{MS2SLS}) = \hat{\gamma}^{-1} * \hat{Var}(\hat{\theta}) * \hat{\gamma}^{-1} + (\hat{\beta}'_{M2SLS} \otimes \hat{\gamma}^{-1}) * \hat{Var}(vec(\hat{\gamma})) * (\hat{\beta}_{M2SLS} \otimes \hat{\gamma}^{-1})$$

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:

$$\hat{V}(\hat{f}^m) = \left(\frac{\Delta \hat{M} A_1 + \Delta \hat{M} A_2}{\hat{\theta}_1 + \hat{\theta}_2} \right)^2 * \hat{V}(\hat{\boldsymbol{\beta}}^{MS2SLS})[1,1]$$

F.3 Bound are invariant to the size of the compliant population

In order to estimate bounds on mechanisms, we only need to estimate the reduced-form effects of treatment. It is not necessary to estimate TOT effects. Let τ be the fraction of drafted individuals that took-up the house. Then, we can write TOT effects as:

$$\tilde{\theta} = \begin{bmatrix} \frac{\theta_1}{\tau} \\ \\ \frac{\theta_2}{\tau} \end{bmatrix}$$

Also, all mechanism effects can be written as $\tilde{\beta}^i = \frac{\beta^i}{\tau}$, for $i = m, n, a, c$. We are ultimately interested in the fraction of effects explained by labor market access. Focusing on the TOT does not alter the object of interest:

$$\tilde{f}^m = \frac{\tilde{\beta_m} * (\Delta MA_1 + \Delta MA_2)}{\tilde{\theta}_1 + \tilde{\theta}_2} = \frac{\tau * \beta^m * (\Delta MA_1 + \Delta MA_2)}{\tau * (\theta_1 + \theta_2)} = f^m$$

G Mechanism analysis: Robustness checks

Table F.1 examines whether our results are robust to considering different measures of network quality. We consider the following alternative measures: 1) average years of schooling in the neighborhood; 2) average household income, 3) a PCA index of different socioeconomic measures that include income, education, and poverty; 4) average formal wages in the neighborhood; and 5) average baseline formal wages of peers drafted to the same housing project. We compute new bounds on labor market access using these alternative measures instead of average income. As Table F.1 shows, our results remain largely unchanged.

Another important identifying assumption for our mechanism analysis is that we are able to describe the vector of potential mechanisms accurately. Even though we cannot directly test this assumption, our framework is flexible enough that is straightforward to widen the vector of potential mechanisms. In Table F.2, we add other potential mechanisms and re-estimate the bounds for labor market access. These additional mechanisms include population (which might proxy network diversity, as in (Zenou, 2013)), as well as sector and occupation diversity at the neighborhood level, which we measure as the number of 6-digit sectors and occupations in each census tract. We find that including these additional mechanisms does not significantly affect the bounds estimated in the main text.

Table A.1: Additional variables' description

Variable	Source	Reference	Description
White	RAIS and SR	Table 1, 7, 8	dummy for individuals that report being white
Male	RAIS and SR	Table 1,7,8	dummy for individuals that report being male
Schooling	RAIS and SR	Table 1, 7, 8	Categorical variable, see above
Ever employed	RAIS	Table 1, 7, 8	dummy for ever formally employed between 2002 and 2014
Hours	RAIS	Table 1, 4, 5, 7, 8	Weekly number of hours registered in formal contract
Tenure	RAIS	Table 1, 7, 8	Number of months in current job
Distance to previous job (km)	RAIS	Table 1	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
Neighborhood income	Census	Table 2, 3	See text above
Neighborhood education	Census	Table 2,3	Average years of schooling Schooling variable is categorical: 1) no education, 2) up to middle school; 3) up to high school, 4) up to college, 5) post-graduate.
Single-parent households	Census	Table 2,3	Share of single-parent households
Formal Jobs	RAIS	Table 2, 3	Total number of formal jobs in establishments located in the neighborhood
Market access	RAIS and SR	Table 2,3 and Figure 3, 4	see separate description
Amenities	city hall records	Table 2,3 and Figure 3, 4	see main text
Crime rates	ISP-RJ	Table 2 ,3 Figure 3, 4	see main text
Treated	Lottery	Table 1,7,8	see text above different housing projects
Number of neighborhood residents	SR	Figure 1	Total number of individuals who were drafted and report living one neighborhood postal-code (either before of after the lottery)
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, 6	number of months the worker was employed during each year
log(wages)	RAIS	Table 4,5,6	natural logarithm of average wages during the year only for employed workers
Firm quality	RAIS	Table 4	Percentile of firms in the Rio de Janeiro labor market. Ranking is based on an estimate of firm fixed-effects using a wage regression and controlling for gender, race, age, and schooling.
Occupation quality	RAIS	Table 4	Percentile of occupations in the Rio de Janeiro labor market. Ranking is based on average wages for workers employed in each occupation in the baseline period.
Kept the same job	RAIS	Table 4	Worker was still linked to the same firm in the 2016 and 2017 as in the baseline period.
Distance from job to housing projects	RAIS	Table 4	Distance between the centroid of the postal code reported by the firm workers were employed and the housing projects, measured in kilometers.
Worked in a neighboring municipality	RAIS	Table 4	Dummy for individuals that were employed by firms that reported addresses in neighboring municipalities
House prices	selling sites	Table 9	prices of houses being offered during February 2022 in any of the *MCMV* neighborhoods
Number of rooms	selling sites	Table 9	number of rooms being advertised alongside the house being sold
Number of bathrooms	selling sites	Table 9	number of bathrooms being advertised alongside the house being sold
Area	selling sites	Table 9	total area being advertised alongside the house being sold
Garage	selling sites	Table 9	dummy for a garage spot being advertised alongside the house being sold
Lobby	selling sites	Table 9	dummy for the presence of a lobby being advertised alongside the house being sold

Table A.2: Validation of employment status reported in the Single Registry

	Employed in SR	Not employed at the SR
Employed in RAIS	15.2%	2.1%
Not employed in RAIS	3,0%	79,7%

Note: Comparison of employment status reported in the Single Registry for individuals that provided updates between 2015 and 2017 with administrative records from RAIS.

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: Data from the disadvantaged sample. The dependent variable is a dummy for individuals that report a postal code for neighborhoods 1, 2, and 3. Standard-errors in parenthesis are clustered at the individual-level. * p<0.1, ** p<0.05, and *** p<0.01.

Table B.2: Selection into take-up

	All	Neighborhoods 1 and 2	Neighborhood 3
Male	0.006	0.036	-0.015
	(0.023)	(0.037)	(0.029)
Firm size	0.011**	0.013*	0.010*
	(0.004)	(0.007)	(0.005)
Log(wages)	0.002	0.002	0.001
	(0.004)	(0.006)	(0.005)
Tenure	-0.004*	-0.003	-0.004
	(0.002)	(0.003)	(0.003)
Hours	0.010	0.079	-0.056
	(0.104)	(0.160)	(0.139)
Schooling	-0.043***	-0.034	-0.047**
	(0.017)	(0.028)	(0.020)
White	-0.045*	-0.006	-0.071**
	(0.024)	(0.039)	(0.031)
Age	-0.001	-0.002	-0.001
	(0.001)	(0.002)	(0.002)
Observations	2412	899	1513

Note: Estimates for the main sample. The dependent variable is a dummy for drafted individuals that were excluded form the *MCMV* registry in following lotteries, which proxies being treated. *White* and *Male* are dummy variables; *schooling* is a five-level index indicating the highest level of instruction that ranges from no schooling to post- graduation. *Hours* denotes the number of monthly hours and *wages* represent the average monthly compensation registered in contract. *Tenure* is measured in number of months. *Firm size* is the number of workers employed in the same firm as the individual. Robust standard-errors are shown in parenthesis. * p¡0.1, ** p¡0.05, and *** p¡0.01.

Table C.1: Dynamic effects on formal employment

	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: Results obtained using the main sample. Formal employment is a dummy for being formally employed at any point in 2015-2017. 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.2: ITT Estimates – alternative specifications

	ANCOVA	Dif-in-Dif
	(1)	(2)
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	

Note: Results obtained using the main sample. Formal employment is a dummy for being formally employed at any point in 2015-2017. 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. Clustered standard-errors at the individual-level are shown in parenthesis.

Table C.3: Heterogeneous effects on formal employment

	Control mean (1)	ITT (2)	TOT (3)	Observations (4)
Women	0.60	-0.009	-0.017	828,489
		(0.007)	(0.014)	
Men	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: Results obtained from the main sample. Formal employment is a dummy for being formally employed at any point in 2015-2017. Each line shows results for a different group: males, females, non-whites, whites, low education, and high education. High education is defined as at least completed high school. 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.4: Effects on informal employment

	Control mean	ITT	TOT	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)	

Note: Results based on the disadvantaged sample. Informal employment is a dummy for individuals that reported working without a formal contract (*carteira assinada*). Number of months employed informally is the total number of months in each year individuals were informally employed, zero if not employed or formally employed. Clustered standard-errors at the individual-level are shown in parenthesis. * p<0.1, ** p<0.05, and *** p<0.01.

Table C.5: Heterogeneous effects by distance from previous home

	Main sample	Disadvantaged sample
Drafted	0.002	0.034
	(0.013)	(0.027)
Drafted x distance (100 km)	0.022	-0.039
	(0.018)	(0.048)
Distance (100 km)	0.035***	0.038***
	(0.003)	(0.004)
N	1,511,633	607,203

Note: Estimates based on the main sample (column 1) and disadvantaged sample (column 2). Formal employment is a dummy for being formally employed at any point in 2015-2017. The first (second) column includes controls for the distance to the baseline employment (home) and an interaction between the draft dummy and the distance variable. * p<0.1, ** p<0.05, and *** p<0.01.

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			
	-0.069*	-0.057*	0.011
	(0.042)	(0.029)	(0.019)
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	

Note: Estimates from the main sample. Te dependent variable is formal employment is a dummy for being formally employed at any point in 2015-2017. Panel A shows the original neighborhood-specific treatment effects estimates in Table 6. Panel B shows the estimates re-weighted by the complier composition of neighborhood 3. We also show two Wald tests for null hypothesis that the re-weighted coefficients are equal to the unweighted coefficient and to the estimate for neighborhood 3. Clustered standard errors at the individual level are shown in parenthesis: * p¡0.1, ** p¡0.05, and *** p¡0.01.

Table F.1: Labor Market Access Bounds – Alternative measures of network quality

	Lower bound LB^{β_m}	Upper bound UB^{β_m}
Original measure	0.821	0.934
	(0.051)	(0.046)
Average education	0.821	0.934
	(0.051)	(0.046)
Average household income	0.821	0.934
	(0.051)	(0.046)
PCA index	0.791	0.934
	(0.049)	(0.046)
Formal wages	0.821	0.907
	(0.051)	(0.066)
Wages in the housing project	0.761	0.957
	(0.047)	(0.046)

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 $f^m = \frac{\beta^m * (\Delta MA_1 + \Delta MA_2)}{\theta_1 + \theta_2}$. Bounds and robust standard-errors are estimated using the MS2SLS.

Table F.2: Labor Market Access Bounds – Additional mechanisms

	Lower bound LB^{β_m}	Upper bound UB^{β_m}
Original measure	0.821	0.934
	(0.051)	(0.046)
Population	0.821	0.934
	(0.051)	(0.046)
Occupation variety	0.821	0.934
	(0.051)	(0.046)
Sector diversity	0.799	0.934
	(0.049)	(0.046)
Firm quality	0.821	0.934
	(0.051)	(0.046)

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 $f^m = \frac{\beta^m * (\Delta MA_1 + \Delta MA_2)}{\theta_1 + \theta_2}$. Bounds and robust standard-errors are estimated using the MS2SLS.