Family Ties, Residential Location Choice, and Labor Supply*

Hadar Avivi, Santiago Lacouture, Felipe Lobel & Winnie van Dijk†

Mach 26, 2026

PRELIMINARY AND INCOMPLETE, PLEASE DON'T CITE OR CIRCULATE

Abstract

This paper investigates the role of family proximity in shaping location and labor supply decisions among low-income households. We combine housing lotteries in Brazil's Minha Casa Minha Vida program with administrative data linking lottery participants to their extended family networks and labor market outcomes. The lotteries generate plausibly exogenous variation in whether participants relocate and, in turn, in geographic separation from relatives. We first show that about 15% of nearby non-co-resident relatives move closer to lottery recipients who relocate. We also find that non-coresident relatives of participants experience substantial reductions in formal employment and increases in social transfer receipt. Decomposing these effects by relatives' mobility decisions, we find that the aggregate labor supply decline is driven by relatives who stay behind, not by a cost of induced mobility among those who follow. These results show that proximity generates spillover effects on family members' labor market outcomes, with implications for evaluating housing policies that relocate low-income households. In ongoing work, we test whether selection on labor market gains can explain the relatives' location decisions, which has the potential to shed light on whether formal institutions can substitute for proximity-based in-kind transfers between family members.

*We are grateful to Joe Altonji, Janet Currie, Daniel da Mata, Deniz Dutz, Mayara Felix, John Eric Humphries, Camille Landais, Magne Mogstad and seminar participants for helpful comments. Davi Resende and Nick Vilay provided excellent research assistance. The authors gratefully acknowledge financial support from the Yale Tobin Center for Economic Policy and the Cowles Foundation for Research in Economics at Yale University.

†Avivi: University College London. Lacouture: The University of Chicago. Lobel: Stanford University. van Dijk: Yale University.

1 Introduction

Low-income households often remain in areas with limited economic opportunity despite potentially large gains from moving (Rosenzweig and Stark, 1989; Bryan et al., 2014; Bazzi, 2017; Lagakos et al., 2023; Bryan and Morten, 2019; Munshi, 2020). A leading explanation is the value of proximity to social networks, and to family in particular. Consistent with this view, social ties are geographically clustered (Compton and Pollak, 2014; Bailey et al., 2018, 2020), and migrants tend to settle near pre-existing social contacts (Altonji and Card, 1991; Munshi, 2003; Stuart and Taylor, 2021; Sahai and Bailey, 2022; Blumenstock et al., 2025; Koenen and Johnston, 2025). But these patterns admit competing interpretations. Relatives may sort into the same neighborhoods because they share preferences over local wages, housing, schools, or amenities, or because they face common moving costs. Little evidence exists on whether family proximity *causally* affects location decisions and economic outcomes—or whether observed clustering simply reflects correlated unobservables.

Moreover, if proximity does affect location and labor market outcomes, the underlying mechanisms carry different policy implications. Nearby relatives can provide in-kind transfers that deliver *production value*—childcare, job referrals, informal insurance, and other inputs that support labor market participation—as well as *consumption value* from time spent together. This distinction matters along two dimensions. First, it determines whether formal institutions such as public childcare or social insurance can substitute for family networks and thereby relax mobility constraints: formal institutions may substitute for production value but not for consumption value. Second, it affects how to evaluate policies that relocate beneficiaries away from their networks. If proximity matters for labor market outcomes, such programs may impose previously unrecognized consequences through spillover effects on family members left behind.

Establishing whether proximity has a causal effect on location and labor market outcomes, and distinguishing between production and consumption channels, is empirically difficult for three reasons. First, social and family networks are rarely observed in standard data. Second, when networks are observed, within-network moves are endogenous: members choose where to live based on shared preferences or constraints, making it hard to determine whether clustering reflects a causal effect of proximity or simply correlated unobserved factors. Third,

distinguishing production from consumption value requires observing economic outcomes that respond to geographic separation and tracking those outcomes as families adjust their locations over time.

This paper investigates the role of family proximity in shaping location choices and labor supply among low-income households. We address the challenges described above by combining housing lotteries in Brazil's *Minha Casa Minha Vida* (MCMV) program with linked administrative data on extended family networks and labor market outcomes. MCMV is one of the world's largest housing subsidy programs, serving households earning below approximately USD 400 per month, and allocates heavily subsidized housing units through public lotteries. The lotteries generate plausibly exogenous variation in whether and where applicants relocate. Because MCMV projects are typically located on the urban periphery, lottery winners who take up their assignments move substantial distances—17.6 kilometers on average—from their baseline locations, creating significant variation in family separation. We link lottery participants to parents, children, siblings, in-laws, and other relatives using Brazil's comprehensive social registry (CadÚnico), and we observe location and labor market outcomes in formal employment records (RAIS) and social transfer data. The resulting dataset contains approximately 1.9 billion unique family relationships across Brazil, enabling the analysis of extended family structures at a scale that survey-based approaches cannot achieve. The setting therefore turns a policy-induced move by one family member into a source of exogenous variation in proximity for the rest of the family.

We proceed in five steps. First, we study applicants' own take-up decisions. Take-up is correlated with pre-existing family geography. Winners whose relatives mostly live more than 3 kilometers away are about 15% more likely to move than winners whose relatives mostly live nearby. Because family geography is not randomly assigned, we treat this evidence as descriptive rather than causal. However, it is consistent with family proximity influencing take-up: even heavily subsidized moves are less attractive when they sever day-to-day access to family.

Second, the lottery allows us to identify how one family member's relocation affects the location choices of others. Focusing on nearby non-co-resident relatives of lottery winners who take up the subsidy, we estimate that about 15% are *followers*—they move closer

to the recipient only when the recipient relocates. This response represents a more than 200% increase relative to baseline mobility rates. We also find no comparable response for relatives of applicants who sign a contract but do not ultimately move, suggesting that the spillovers operate through changes in family proximity rather than through direct exposure to the subsidy itself. The evidence points to active coordination within families: location choices are not independent across relatives. One family member's move changes the optimal location of others.

Third, we examine the labor market consequences of lottery-induced family separation. A recipient's relocation causes nearby non-co-resident relatives to shift away from formal employment. Formal employment falls by 15.7 percentage points—a 37% decline relative to the untreated complier mean—while Bolsa Família receipt rises by 24.9 percentage points, a 150% increase. Overall employment falls more modestly and less precisely, implying that the principal margin of adjustment is a reallocation away from formal work and toward informal activity or transfer dependence rather than a wholesale exit from employment.

Fourth, we develop an extended instrumental-variables framework to interpret these labor supply effects when relatives' own location choices are endogenous. With one instrument (the lottery) and two endogenous variables (whether the candidate moves and whether the relative moves), the standard IV estimand conflates pure separation effects for relatives who stay behind with the consequences of induced mobility for relatives who follow. Building on recent work on treatment effects with multiple unordered treatments (Kirkeboen et al., 2016; Mountjoy, 2022; Dutz et al., 2024), we identify the shares of relative location types and construct sharp bounds on type-specific labor effects. Within the complier population, most nearby relatives are stayers; about 15% are followers; and the always-mover share is small. The bounds indicate that the negative labor effects are driven by relatives who stay behind when family members move away, whereas the bounds for followers are wider and do not exclude zero. The main labor market consequence of the policy therefore appears to come from family separation itself, not from the moving costs borne by those who follow.

Fifth, we use these results to ask whether family proximity provides production or consumption value. The employment reductions among stayers are consistent with family proximity providing production value—access to childcare, job information, or informal insurance

that supports work—but could also reflect reduced need to work if relatives were previously supporting the departing family member. In ongoing work, we use a Roy-style selection framework to ask whether the relatives who follow are precisely those who would suffer the largest labor-market reductions from staying behind. If so, the selection margin itself would indicate that family proximity has production value: followers move because the productive services embedded in proximity—childcare, job networks, risk-sharing—are worth the cost of relocating. If not, the remaining value of proximity is more naturally interpreted as consumption value, or as reflecting margins not captured by the labor outcomes we observe. Combined with the bounds decomposition, this test provides a framework for distinguishing production from consumption value of family proximity using quasi-experimental variation—a distinction that, to our knowledge, has not been tested in existing work.

Taken together, the evidence establishes that family proximity causally affects both location choices and labor market outcomes. Relatives actively coordinate where they live, and geographic separation has substantial economic consequences for those left behind. Whether these consequences are also what motivates coordination—the production-value hypothesis—or whether families coordinate for other reasons and bear the labor-market consequences as collateral effects is precisely the question the Roy-model analysis is designed to resolve. Either way, the results suggest that mobility is a collective family decision rather than an individual response to prices alone, and that the consequences of separation extend beyond the relocating household.

The findings have three broader implications. First, they change how to evaluate housing assistance programs, place-based policies, and other interventions that create relocation incentives. Existing evaluations of MCMV and similar programs focus on direct effects on recipients (da Mata and Mation, 2018; Machado and Rachter, 2020a,b; Belchior et al., 2023; da Mata and Mation, 2025), as do evaluations of housing mobility programs in other contexts (e.g., Jacob and Ludwig, 2012; Chetty et al., 2016; Chyn, 2018). If recipients' moves reduce the formal employment of relatives left behind, then standard cost-benefit calculations miss potentially important family spillovers. By linking lottery winners to relatives and tracing both residential and labor-market responses, we show that the consequences of subsidized relocation extend beyond the household that receives the unit.

Second, the results bear on whether formal institutions can substitute for the support that family networks provide. The magnitude of the employment reductions we document suggests that this substitution is incomplete in Brazil's context: family networks provide support that formal institutions do not fully replace. Whether formal institutions *can* serve as adequate substitutes, and under what conditions, is precisely the question that the production-versus-consumption distinction is designed to answer. If the Roy-model analysis confirms that followers select on employment consequences, the implication is that expanding access to formal childcare or job search assistance could reduce the consequences of geographic separation and thereby increase residential mobility among low-income households. If instead the value of proximity is primarily consumption, then such policies would be less effective at relaxing mobility constraints.

Third, the results speak to models of residential choice and migration. Standard spatial-equilibrium models characterize location decisions as individual trade-offs among wages, housing costs, and amenities (Rosen, 1974; Roback, 1982). Structural migration models estimate large costs of moving but do not model the mechanisms that generate those costs (Kennan and Walker, 2011; Bryan and Morten, 2019). Discrete-choice models of neighborhood selection allow rich preference heterogeneity but assume independent optimization across households (e.g., Bayer et al., 2016). Our evidence suggests that these frameworks may benefit from explicitly accounting for interdependencies in family location decisions. The 15% of relatives who relocate to maintain proximity demonstrates that one family member's move directly affects others' optimal locations, and the large employment effects among those left behind show that separation has economic consequences that extend beyond the relocating household. Models that omit these interdependencies risk understating mobility costs and overstating the gains from policies that encourage spatial reallocation.

Organization. Section 2 describes the institutional setting. Section 3 describes the data and sample construction. Section 4 studies candidates' take-up decisions, and Section 5 studies relatives' location responses. Section 6 analyzes labor-supply effects and the bounds decomposition. Section 7 presents the Roy-style selection analysis. Section 8 concludes.

2 The Minha Casa Minha Vida Program (MCMV)

Program Origins and Objectives. *"Minha Casa Minha Vida"* (MCMV) program was created by Federal Law 11.977 in 2009 with the stated goal of reducing Brazil's housing deficit for low- and middle-income families. Over its first decade (2009–2018), MCMV financed the construction of roughly 5.5 million homes at a total federal outlay of R$ 464 billion (approximately USD 92.8 billion). Although not explicitly framed as a social mobility initiative, the program includes guidelines for community infrastructure, such as nearby schools, to support beneficiary families' long-term welfare. Resources are drawn from the federal budget, channeled through Caixa Econômica Federal ("Caixa"), which underwrites mortgages, audits allocation procedures, and enforces contract compliance.

Eligibility and Subsidy. MCMV is divided into income "faixas" (tiers) with Faixa 1 serving households earning up to R$ 1600 per month (roughly USD 400). In this segment, beneficiaries receive subsidies covering up to 90% of construction costs, while the remaining 10% is amortized over 120 months at a zero interest rate, with monthly payments capped at 5% of household income (or a minimum of R$ 25, or USD 5). These highly concessional terms ensure that beneficiary payments are typically a fraction of prevailing rents in low-income neighborhoods. In contrast, higher income bands (Faixa 2 and Faixa 3) offer smaller subsidies and mortgages with annual interest rates of approximately 5–7%. No down payment is required for Faixa 1, and the properties serve as collateral held by Caixa until the loan is fully repaid.

Beneficiary Selection and Priority Rules. To allocate scarce Faixa 1 units fairly in the event of excess demand, municipalities must conduct public lotteries under Ministry of Cities' Ordinance 140 (April 2010). Each lottery produces (i) a list of winners equal to the project's unit count, (ii) a waiting list comprising 30% of unit capacity, and (iii) a list of non-selected participants. Federal law mandates three priority categories: families in risk-prone areas, female-headed households, and households with disabled members; local governments may add up to three further criteria. As a result of the preference for female-headed households, women comprise the majority of Faixa 1 beneficiaries.

Implementation in Rio de Janeiro (2011–2018). The city of Rio de Janeiro, home to 7 million residents, received 2% of total MCMV housing units. Given that demand exceeded supply of houses, the Government assigned several allocation mechanisms, including lotteries, which we study in this paper.[1] Rio de Janeiro conducted 46 general lotteries between 2011 and 2018, assigning over 40 thousand housing units among more than 30 different housing complexes.[2] The Municipal Housing Secretariat proactively identified individuals in vulnerable socioeconomic conditions to enroll them in the program. This outreach effort resulted in a database of over 750,000 participants. Selected winners were contacted by public officials, either by phone or letter, and offered a housing unit through the program.

Winners received greatly subsidized homes: 90% of construction costs covered by government payments (approximately USD 10,000 per unit) and the remaining 10% financed at zero interest, corresponding to roughly one-fifth of average local rents. Despite the large reduction in rent expenditure, individuals are not allowed to sell or rent the houses. Eligibility was restricted to Brazilians with households earning below USD 300 per month (consistent with Faixa 1 criteria) that do not own a home or have access to home financing. Figure 1 shows a map of Rio de Janeiro in which red dots mark the locations of MCMV housing developments and the shaded areas indicate the neighborhoods where lottery applicants lived before the program. Although the projects themselves are fairly concentrated, the applicants (and their relatives) relocate from a broad swath of the city, underscoring substantial geographical variation.[3] There is substantial variation in winning households' pre-lottery addresses – spanning up to 50 kilometers.

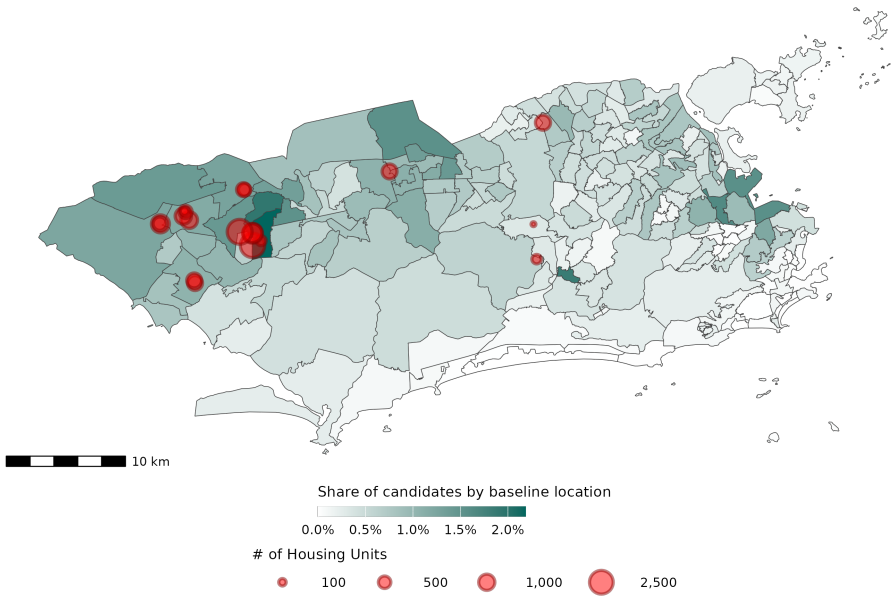
Lottery assignment mechanism. Each MCMV lottery was designed to assign a predetermined number of housing units across one or more housing projects. The lotteries were linked to Brazil's Federal Lottery (*Loteria Federal*) operated by *Caixa Econômica Federal*, a federal public bank. Candidates in each lottery were organized alphabetically and assigned

[1] Typically 40% of the housing units were assigned through lottery. The remaining units were assigned through different mechanisms, such as: (i) individuals that lost their homes in natural disasters, or that live in vulnerable conditions; (ii) elderly and individuals with disabilities.

[2] Smaller special lotteries for elderly and disabled individuals also took place in this period. We focus on general lotteries since each special lottery assigned very few units.

[3] MCMV projects were typically built in peripheral areas due to land cost constraints and federal price ceilings, which incentivized developers to choose more affordable parcels outside urban centers.

Figure 1: *Geographic Dispersion of Candidates and Projects in Rio de Janeiro*



Note: This figure shows a map of Rio de Janeiro. Red dots mark the locations of MCMV housing developments and are scaled based on the number of units in the project. The shaded areas indicate the neighborhoods where lottery applicants lived before the program. The figure is produced using our analysis sample as described in Section 3.

sequential numbers, with winners selected by matching their assigned numbers to specific digits from the federal lottery results. The specific assignment mechanism based on the lottery evolved considerably over time. Initially (2011-2014), winners were purely determined by lottery digits, deliberately generating more winners than available units. Housing units were then allocated on a first-come, first-served basis among winners. This mechanism created a significant administrative burden as it left less-desired units unassigned and resulted in lower take-up. In early 2015, the Housing Department tested a modified approach where numbers matching different federal lottery prizes were pre-assigned to specific housing projects, automatically determining which development each winner would be assigned to, though unit allocation within each project remained first-come, first-served. By mid-2015, the program shifted towards age-based prioritization: selected candidates were still chosen using lottery digits, but then ranked in descending order of age, with older winners having priority in receiving housing units.

3 Data

3.1 Data Sources

We draw on rich administrative microdata from four primary sources that we describe below. Each data source is identified at the individual-level, which allows us to build a panel following individuals and their family members over time.

MCMV Lottery data. The first source is administrative data from the housing program lotteries (2011 to 2018), covering the full set of applicants in Rio de Janeiro. Provided by the Municipal Housing Secretary, these records include individuals' names, a unique personal identifier (Cadastro de Pessoa Física – CPF), and the location of the housing projects with units allocated in each lottery. This data also contains an identification of selected winners in each lottery.

RAIS. The second source is RAIS (Relação Anual de Informações Sociais), a comprehensive matched employer-employee administrative database collected annually by Brazil's Ministry of Labor. RAIS captures all formal employment relationships in the country, as employers are legally required to report detailed information on their workforce. RAIS provides monthly records on formal employment status, including contract details, working hours, and average monthly wages for years 2007 through 2020.

CadÚnico. Third, we use restricted-access data from the Cadastro Único para Programas Sociais (CadÚnico), Brazil's unified registry for social programs. This database provides rich individual- and household-level information, including full names, family structure, housing conditions and address, income, select labor market attributes, and household expenditures. CadÚnico was created to centralize data on all potential beneficiaries of federal social assistance programs. After their initial enrollment, families are expected to update their records every two years, although compliance is uneven. Our version of the dataset spans the years 2012 to 2020.

Housing Contract data. Fourth, we draw on administrative records from Caixa Econômica Federal covering all mortgages issued to families participating in the "Minha Casa Minha

Vida" (MCMV) program over the 2010–2017 period. Each line in the raw data corresponds to a single contract, with information on the exact signing date, identification of signers, amount of subsidy and the municipality where the housing development is located.

3.2 Sample Construction

Our analysis sample is based on lottery candidates with a geocoded pre-lottery home address for which we observe at least one relative over 16 years old (also with a pre-lottery geocoded address). We briefly describe the construction of this sample; further details can be found in Appendix Section B.4.

Candidates in CadÚnico. Our sample construction starts with 737,885 unique candidates (38,085 winners) for the 32 MCMV 'general' lotteries in Rio de Janeiro between 2011 and 2016.[4] Since individuals can participate in more than one lottery, our data is built at the candidate×lottery level; thus, candidates who participated (or won) in multiple lotteries appear multiple times. This yields a total of 9,272,440 candidate×lottery (39,068 winners×lottery). To study location choices of participants, we subset to 2,077,914 candidate×lottery (200,571 unique candidates) with a geocodable residential address in CadÚnico before the lottery. We also exclude 8 lotteries with an insufficient number of linked winners to meaningfully inform any analysis (below 15). This reduces our sample to 2,002,707 candidates (3,967 winners) in 22 lotteries between years 2013 and 2016. The exclusion of the pre-2013 lotteries is a consequence of the timespan covered by CadÚnico.[5]

Family links. We identify family relationships through individual identifiers, household structure, and parental name matching across the universe of CadÚnico (see Appendix B.2). This allows us to link candidates to their direct relatives (parents, children, siblings), extended family (grandparents, grandchildren, aunts, uncles, cousins), and in-laws. Using these data, we identify (a) candidates with at least a relative over 16 years old (and with an address)

[4] We exclude post-2016 lotteries to ensure sufficient post-lottery observation periods.

[5] Although CadÚnico covers the period from 2012 to 2020, some individuals were interviewed years before. We carefully use those interview dates in identifying individuals' locations earlier in time. See Appendix B.3 for further details.

before the lottery, and (b) those relatives and their locations.[6] Since not all candidates have relatives meeting these criteria, our effective sample reduces to 560,678 candidates×lottery (1,140 winners), linked to 1,175,306 relative-candidate pairs×lottery (2,358 relatives of winners). Appendix Table A.1 presents the counts in our sample for each lottery.

3.3 Variable Construction

We construct pre- and post-lottery variables for both lottery candidates and their relatives using CadÚnico and RAIS administrative data, supplemented with housing contract information. These variables fall into three main categories: individual characteristics, housing and location outcomes, and labor market outcomes. Appendix Section B.5 provides additional construction details.

Sociodemographics. We extract individual characteristics from CadÚnico surveys completed prior to each lottery, using the survey closest to (but preceding) the lottery date. These include age, race, literacy status, disability status, and household composition. Using our family linking procedure, we construct measures of family network size and composition before the lotteries.

Locations and distances. We geocode residential addresses from CadÚnico surveys to measure locations both before and after each lottery. Pre-lottery locations use the survey closest to (but preceding) the lottery date. Post-lottery locations use the most recent available survey – up to 4 years post-lottery to maximize sample coverage. We combine these individual geocoded addresses with (a) Housing contract data, and (b) geocoded locations of the housing projects to build a set of location variables. Key location variables include: (i) program take-up indicators (contract signing and residence at housing projects), (ii) mobility indicators (moved from baseline location, moved closer to the nearest housing project allocated in the relevant lottery, distance to the projects), and (iii) family proximity measures (distance between the pair, living together, nearby within 3km, or far apart).[7] Note that

[6] Importantly, we subset to those relationships identified with information collected before each lottery.

[7] All distances are reported in kilometers, winsorized at 25km. Since unit-specific address may differ from the housing development's main address, we define residence at a housing project if living within 300 meters of the main address.

post-lottery location data are unavailable for individuals who did not complete a CadÚnico survey after a lottery; therefore, to maximize sample size, we collect post-lottery location data up to 4 years after the lottery.

Labor market and social program take-up. We measure formal employment using RAIS records both pre- and post-lottery, defining it as any registry presence during the period, along with total earnings and number of jobs. We track employment transitions including job retention and new job acquisition. We supplement formal employment with informal employment indicators from CadÚnico (individuals reporting employment but absent from RAIS) and social program participation (Programa Bolsa Família, PBF), with both defined as any participation during pre- and post-lottery periods. Post-lottery employment and program variables are subject to survey non-response, similar to location variables.

3.4 Descriptives and Balance

3.4.1 Candidates

Table 1 provides summary statistics for lottery winners and losers in our analysis sample. Winning probability is extremely low, with only 0.2% of the lottery candidates receiving an offer. Moreover, the demographic information in Table 1 suggests that the MCMV population is mostly middle-aged mothers[8], with more than half beneficiaries of PBF, which implies that they are heavily disadvantaged. Prior to the lottery, 65% had worked in the formal labor market for at least one year between 2007 and the lottery year. Given the cumulative nature of this measure over a 6-9 year period, this suggests lower annual formal employment rates among lottery candidates compared to the general Brazilian population (59% formally employed in a given year; Bouvier et al., 2022), reflecting the disadvantaged labor market position of the low-income CadÚnico population.

Column 3 reports the adjusted difference between lottery losers and winners. Since randomization is within each lottery and conditional on candidates' age, these differences are estimated from a regression with lottery×age-bins fixed effect, with standard errors clustered at the individual level (as the same individual can participate in multiple lotteries). The

[8]As described in Section 2, MCMV prioritized female headed households. This is in line with the disbursement policy of other social programs, such as Bolsa Família, which is paid to women.

Table 1: *Balance tests for MCMV housing lottery winners and losers*

Variable	Loser (1)	Winner (2)	Adj. Diff. (3)	(s.e.) (4)	# q-value< 0.1 (5)
Age	39.576	42.35	0.057	(0.068)	0
Gender: Male	0.188	0.174	-0.005	(0.011)	0
Disabled	0.039	0.039	-0.002	(0.006)	0
Indigenous	0.001	0.001	0	(0.001)	0
Illiterate	0.031	0.046	0.012	(0.006)	0
Race: Black	0.212	0.229	0.015	(0.013)	0
PBF Beneficiary	0.558	0.578	0.019	(0.014)	1
N relatives	4.113	4.155	0.114	(0.092)	0
N children	1.503	1.64	0.021	(0.036)	0
N parents	0.352	0.24	-0.024	(0.014)	1
Has kids	0.718	0.781	0.018	(0.012)	1
Has kids under 16	0.542	0.562	0.009	(0.013)	0
Lives with relative over 16	0.311	0.329	0.017	(0.014)	3
Any relative living nearby	0.509	0.477	-0.031	(0.015)	2
Any relative living far	0.338	0.361	0.025	(0.015)	0
Distance to closest HP	16.854	15.733	-0.175	(0.224)	3
Av. Distance to relatives	4.64	4.963	0.35	(0.236)	2
Any employment before lottery	0.78	0.762	0.003	(0.012)	0
Formal employment before lottery	0.658	0.612	-0.007	(0.014)	0
Total formal earnings before lottery	2009.05	2140.959	252.781	(123.93)	0
N candidate×lottery	559538	1140			
N candidate	58040	1113			

Note: This table reports the mean characteristics and differences of MCMV Rio lottery winners and losers, in 22 lotteries conducted between years 2013-2016. Demographic characteristics, informal labor market outcomes, and BFI attainment are from the CadÚnico survey completed prior to the lottery, using data from the survey closest to (but preceding) the lottery date. The difference and the standard errors are from a regression with lottery×age-bins fixed effect, with standard errors clustered at the individual level (see Appendix C.1 for details). Column 5 reports the number of lotteries with q-values (Storey, 2002) lower than 10% for testing the null of no mean differences within each lottery.

results in columns 3-4 suggest no differences in pre-lottery characteristics between lottery winners and losers.

While mean differences across lotteries are statistically insignificant, our design requires that *each* lottery's mean difference between winners and losers is zero. To assess that, we test whether the mean differences between winners and losers are zero for each lottery. To account for false discoveries arising in multiple testing, we follow Storey (2002) and estimate each lottery's *q-value*, the empirical Bayes posterior probability that the lottery is falsely classified as not random, which has a frequentist interpretation as the false discovery rate (Benjamini and Hochberg, 1995). Then in column 5, we report the number of lotteries with a q-value below 10%.[9] Analogous to standard hypothesis testing with p-values, we cannot

[9]Similar to p-value hypotheses testing, if we reject the null in all the lotteries with q-value lower than X%, then it means that we should expect at most X% of the lotteries to be actually random.

reject the null hypothesis of random assignment for most lotteries and variables.[10]

3.4.2 Candidates' relatives

Table 2 examines relatives of lottery candidates, splitting by their baseline distance from applicants. The patterns reveal sorting: non-coresident relatives living near candidates are often their children (26%), while distant relatives are more commonly siblings (33.7%) or extended family (49.2%).

Columns 3, 6, and 9 report the adjusted differences between relatives of lottery winners and losers within each distance category. Since the relative network before the lotteries is endogenous, adjusted differences also control for the number of candidates each relative is related to (Borusyak and Hull, 2020), in addition to lottery×age-bins fixed effects. The standard errors are clustered at the candidate level. The results suggest that randomization was successful across all distance categories, with most adjusted differences being small and statistically insignificant, confirming that relatives of winners and losers have similar pre-lottery characteristics regardless of their residential proximity to the candidate. We accommodate for the small (mostly insignificant) differences in socioeconomic conditions in our empirical design.

4 Candidates' Location Choices and Family Proximity

Our analysis starts by studying the importance of family proximity in the location choices of lottery candidates. Table 3 reports the effect of winning the lottery on lottery candidates' location decisions, estimated via an OLS regression with lottery×age-bins fixed effects. Column 2 in panel (i) shows that winning the housing lottery increases the probability of signing the contract by more than 50%. However, as can be seen in panel (ii), not all those who sign the contract end up moving to the housing project. Therefore, to study changes in the location of residence of lottery candidates, we use information reported in CU, collected up to 4 years post-lottery.

As can be seen in panel (i), post-lottery address data is available for 92% of the sample.

[10] Appendix Figure A.1 reports the distributions of p-values and q-values for the joint test mean zero for all covariates in each lottery.

Table 2: *Balance test for relatives' characteristics by distance from lottery candidate*

	Living together			Living nearby			Living far		
Variable	Loser (1)	Winner (2)	Adj. Diff. (3)	Loser (4)	Winner (5)	Adj. Diff. (6)	Loser (7)	Winner (8)	Adj. Diff. (9)
Age	35.505	33.67	-0.007	35.032	34.115	0.181	36.625	36.669	0.19
			(0.803)			(0.548)			(0.549)
Gender: Male	0.367	0.41	0.022	0.352	0.366	-0.007	0.343	0.347	-0.004
			(0.02)			(0.014)			(0.015)
Illiterate	0.047	0.048	0	0.044	0.045	0.001	0.046	0.055	0.009
			(0.01)			(0.007)			(0.009)
Race: Black	0.178	0.189	0.018	0.204	0.232	0.039	0.203	0.232	0.033
			(0.02)			(0.018)			(0.02)
PBF Beneficiary	0.515	0.518	-0.011	0.59	0.581	-0.017	0.618	0.627	0.008
			(0.026)			(0.022)			(0.024)
Any employment before lottery	0.592	0.59	0.008	0.602	0.567	-0.021	0.569	0.586	0.023
			(0.022)			(0.017)			(0.018)
Formal employment before lottery	0.46	0.458	-0.007	0.325	0.298	-0.033	0.159	0.177	0.015
			(0.022)			(0.016)			(0.016)
Distance to closest HP	17.076	17.338	0.616	16.687	14.722	-0.104	18.172	17.68	-0.011
			(0.377)			(0.35)			(0.517)
Distance to candidate				0.64	0.73	0.077	15.952	16.434	0.574
						(0.036)			(0.491)
Relationship: Parent	0.272	0.21	0.002	0.182	0.101	-0.025	0.075	0.065	0.001
			(0.015)			(0.01)			(0.009)
Relationship: Sibling	0.171	0.134	-0.019	0.259	0.246	0.003	0.337	0.328	-0.008
			(0.014)			(0.013)			(0.016)
Relationship: Child	0.407	0.523	0.031	0.26	0.311	-0.034	0.096	0.118	-0.002
			(0.021)			(0.018)			(0.014)
Relationship: Other	0.15	0.134	-0.014	0.299	0.342	0.056	0.492	0.489	0.008
			(0.016)			(0.017)			(0.02)
Has spouse	0.173	0.147	0	0.279	0.251	-0.009	0.409	0.424	0.019
			(0.02)			(0.019)			(0.022)
Number of children	0.934	0.857	0.065	1.324	1.248	0.075	1.644	1.65	0.052
			(0.059)			(0.056)			(0.065)
N relatives in household	1.677	1.592	-0.069	1.792	1.828	0.085	1.863	1.878	-0.003
			(0.07)			(0.064)			(0.07)
N kids under 16	0.203	0.219	0.04	0.401	0.415	0.044	0.556	0.546	0.002
			(0.033)			(0.03)			(0.037)
N relatives nearby	0.804	0.831	0.025	1.883	1.91	0.161	0.801	0.957	0.149
			(0.108)			(0.143)			(0.132)
N parents nearby	0.021	0.028	0.008	0.049	0.059	0.012	0.026	0.052	0.026
			(0.008)			(0.011)			(0.012)
N siblings nearby	0.149	0.128	-0.016	0.391	0.407	0.043	0.168	0.191	0.024
			(0.021)			(0.046)			(0.03)
N inlaws nearby	0.122	0.085	-0.03	0.346	0.373	0.044	0.178	0.25	0.072
			(0.021)			(0.041)			(0.041)
N pair×lottery	270313	539		520737	978		381898	841	
N unique pairs	28599	534		57693	972		44326	828	

Note: This table reports the mean characteristics and differences of relatives of MCMV Rio lottery winners and losers, in 22 general lotteries conducted between years 2013-2016. Relatives are grouped by distance to the applicant: relatives living with candidates in the same housing unit, those within walking distance (up to 3 km), and those living more than 3 km away. See Table 1 for variable sources and definitions. The difference and the standard errors (in parentheses) are from a regression with lottery×age-bins fixed effect and controlling for the number of candidates each relative is related to, with standard errors clustered at the candidate level.

Therefore, studying changes in the address of individuals raises the concern of selection bias.

To account for that, in panel (ii) we present, in addition to the estimates of the lottery win effects among respondents, the Lee bound adjusting for covariates (Lee, 2005; Semenova, 2020). While nonresponse to the CU survey is not random, with lottery winners having a higher probability of responding to report their address in CU up to 4 years after the lottery, the winning effect is relatively small (2.7 percentage points). Therefore, this small share of respondent-compliers results in tight bounds (panel (ii)).

Winning the lottery induces substantial relocation among lottery winners: it increases the probability that lottery winners actually move to the housing project by 15%. Importantly, a much larger share (54%) of winners sign contracts. Our preferred measure of program compliance captures only those who sign, relocate, and also update their address in CadÚnico surveys, providing a conservative estimate. This distinction between contract signing and verified residential change ensures that our analysis focuses on households that genuinely relocated rather than those who may have signed contracts without moving.

The distance over which those who relocate move is also sizeable. While the majority of lottery winners remain in their original locations, the unconditional average effect on distance moved is 2.6 kilometers—a figure that includes all non-movers. This implies that lottery winners who actually relocate move considerable distances from their baseline locations: about 17.3 kilometers. The substantial geographic displacement among movers underscores the significance of the location choice, making family proximity considerations even more salient for understanding residential mobility decisions.

Program compliance is stronger among lottery candidates whose relatives live far away. This is illustrated in Figure 2, which plots the estimated effect of winning the lottery on the probability of signing the contract and moving to the housing project—our preferred measure of program compliance—by the candidate's distance from their relatives. The figure distinguishes between candidates living with relatives in the same housing unit, those with relatives within walking distance (up to 3 km), and those with relatives living more than 3 km away. The effect of a lottery win on moving increases by nearly 20% when relatives do not live nearby. This result is consistent with family proximity playing a role in shaping individual location choices, although it does not provide evidence of a causal relationship. In the next section we provide evidence of such a relationship by leveraging the exogenous

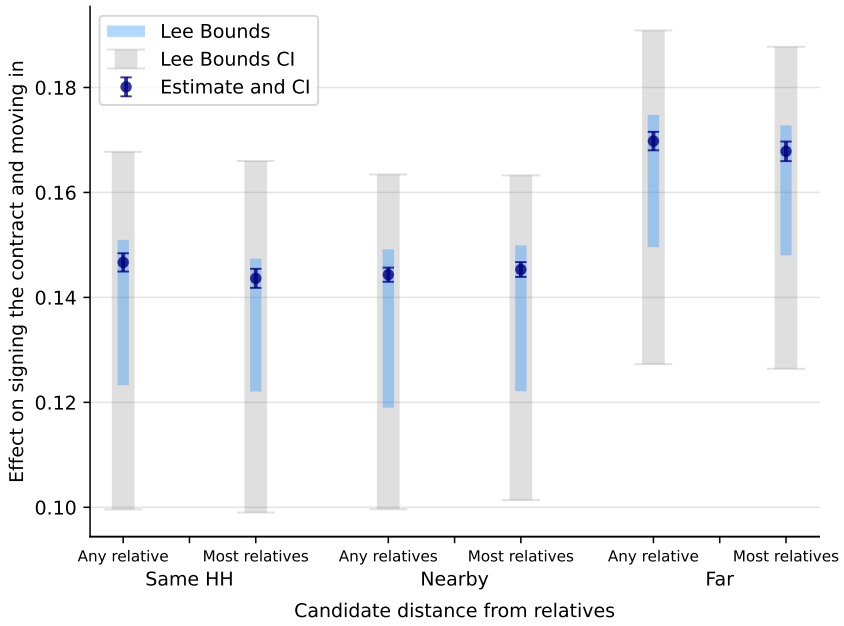
Table 3: *Effect of lottery win on candidates' location choices*

	Loser mean	Estimate	(s.e.)	[Lee Bounds]	(95% CIs)
Variable	(1)	(2)	(3)	(4)	(5)
(i) Baseline sample ($N_0 = 559538, N_1 = 1140$)					
Signs contract	0.000	0.540	(0.001)		
Address post-lottery	0.921	0.027	(0.008)		
(ii) Post-lottery CU respondents ($N_0 = 515395, N_1 = 1087$)					
Moves from baseline	0.237	0.168	(0.013)	[0.145, 0.176]	(0.134, 0.186)
Moves more than 2 kms	0.128	0.160	(0.010)	[0.134, 0.164]	(0.122, 0.175)
Moves closer to HP	0.087	0.165	(0.009)	[0.143, 0.173]	(0.131, 0.184)
Moves to HP	0.005	0.150	(0.002)	[0.129, 0.159]	(0.116, 0.168)
Signed and Moves to HP	0.000	0.153	(0.001)	[0.128, 0.158]	(0.115, 0.168)
Distance moved (uncond.)	1.813	2.641	(0.168)	[1.581, 2.541]	(0.946, 5.162)

Note: This table reports the effect of winning the lottery on candidates location choices. Column 1 is the mean outcome for lottery losers; column 2 gives the point estimate, estimated from an OLS regression with lottery×age-bin fixed effects; column 3 is the standard error, clustered at the individual level, and column 4 reports the covariate adjusted Lee bounds along with the 95% confidence intervals using 500 bootstrap iterations. N_0 and N_1 report the number of candidate×lottery losers and winners respectively.

variation in candidates' location induced by the lottery.

Figure 2: *Candidate location choice by distance from relative network*



Note: This figure shows the effect of winning the lottery on the probability that a lottery candidate signs the contract and moves to the housing project, by distance from relatives. Blue dots and error bars represent point estimates and 95% confidence intervals; light blue bars indicate the Generalized Lee bounds. See Appendices C.1 and C.2 for estimation details. The first two estimates correspond to candidates who live with relatives in the same household; the next two to candidates with relatives nearby (within 3 km); and the last two to those with relatives living farther away (more than 3 km). "Any relative" refers to candidates with at least one relative in that distance group, while "Most relatives" refers to those for whom the majority of relatives are in that group.

Since one cannot move to the housing project without an offer, individuals could belong to only two response groups: either compliers, i.e., individuals who move only if offered housing, or never-takers, i.e., individuals who would never move, regardless of the lottery result. Table 4 reports the characteristics of these groups, suggesting substantial heterogeneity in types by distance from relatives and family structure. First, never-takers tend to live much closer to their relatives compared to compliers (14 vs 31 km).

Table 4: *Lottery candidate compliers and never-takers characteristics*

	Any same household		Any nearby		Any far	
Variable	Complier (1)	Never-taker (2)	Complier (3)	Never-taker (4)	Complier (5)	Never-taker (6)
N relatives	3.829	3.689	4.836	4.497	5.607	5.552
	(0.329)	(0.162)	(0.464)	(0.152)	(0.564)	(0.211)
N children	1.867	1.669	1.815	1.718	1.808	1.617
	(0.198)	(0.076)	(0.172)	(0.061)	(0.193)	(0.075)
N parents	0.250	0.328	0.187	0.217	0.090	0.236
	(0.075)	(0.035)	(0.049)	(0.023)	(0.035)	(0.031)
Has kids	0.827	0.808	0.747	0.813	0.775	0.739
	(0.057)	(0.023)	(0.050)	(0.019)	(0.051)	(0.024)
Has kids under 16	0.578	0.483	0.481	0.605	0.586	0.607
	(0.072)	(0.029)	(0.058)	(0.023)	(0.061)	(0.027)
Distance to closest HP	15.641	17.754	13.132	15.195	13.649	15.746
	(1.338)	(0.495)	(1.123)	(0.444)	(1.224)	(0.520)
Av. Distance to relatives	0.596	1.155	2.223	2.045	12.283	13.274
	(0.193)	(0.179)	(0.411)	(0.187)	(1.073)	(0.453)
Any employment before lottery	0.857	0.615	0.800	0.750	0.499	0.839
	(0.096)	(0.079)	(0.130)	(0.061)	(0.158)	(0.067)
Formal employment before lottery	0.558	0.596	0.535	0.614	0.524	0.635
	(0.072)	(0.028)	(0.058)	(0.023)	(0.063)	(0.027)
Formal earnings before lottery	3580.816	577.797	858.48	1025.267	476.754	1550.17
	(1379.649)	(166.986)	(292.277)	(183.441)	(270.899)	(311.084)

Note: This table presents baseline characteristics of lottery participants by compliance status and proximity to relatives. Given one-sided compliance, 'Compliers' are individuals who won the lottery, signed and moved to the sorted housing unit, while "Never-takers" are winners who won but did not 'comply'. Proximity categories are defined by whether participants have relatives over 16 years old in the same household, within 3 kilometers (nearby), or farther than 3 kilometers (far). Standard errors are clustered at the candidate level and reported in parentheses.

5 Relatives' Location Choices

The previous section shows that lottery candidates who win a lottery are less likely to move to the housing project if they have family nearby and have kids below age 16. Nevertheless, pre-lottery distance between family members is not random, and therefore could reflect moving preference due to other reasons that are correlated with family proximity. For example, geographic separation from family may reflect underlying differences in mobility preferences,

education levels, or life-cycle factors that independently affect willingness to relocate, rather than pure family proximity considerations. Therefore, to study the role of family proximity on location choices, we further exploit the housing lotteries to study how the changes in lottery winners' location of residence influence their relatives' location choices. To describe our model, it is helpful to develop some notation that will be useful for the rest of the paper. For simplicity, we describe the empirical model as if there is only one lottery, while in our empirical implementation, the analysis is conducted within lotteries.

We have a sample of $\mathcal{J}$ MCMV housing lottery candidates, indexed by j. We denote $Z_j \in \{0,1\}$ the indicator which describes whether lottery candidate $j \in \{1,\ldots,\mathcal{J}\}$ won the MCMV housing lottery, and $D_j(z) \in \{0,1\}$ her potential decision to move to the housing project as a function of the lottery offer. Throughout, we sometimes refer to D_j as "location choice" to mean that the individual moves from where she is. For lottery candidates, this decision is defined as signing a contract and effectively changing the address to be at one of the housing projects selected in the lottery. By construction, $D_j(0) = 0$, as without winning the lottery, one cannot move to the housing project.

For every lottery candidate j, we identify n_j relatives who live in Rio, indexed by $i \in \{1,...,n_j\}$, where the function $J(i) = j$ gives the lottery candidate j who is i's relative. As could be seen in the previous section, using individuals' pre-lottery location of residence, we classify individuals into 3 groups: those who lived with their relatives at the same house; those who lived within walkable distance (up to 3 km) from their relatives, but not in the same household; and those who lived far from their relatives, more than 3 km apart. In the previous section, we used this classification to study the sensitivity of lottery candidates' location choices to their distance from their relatives. In the following, we will also use this classification to define our primary sample of interest.

For every individual i, we denote $R_i(z,d) \in \{0,1\}$ her location choice relative to candidate $J(i) = j$ and as a function of her relative's lottery status $Z_{J(i)} = z$ and location choice $D_{J(i)} = d$. For every pair of lottery candidate and relative $(i, J(i))$, we observe $D_{J(i)} = D_{J(i)}(Z_{J(i)})$ and $R_i = R_i(Z_{J(i)}, D_{J(i)})$, which are the potential choices, upon realized lottery status and decisions. To study the effect of lottery candidates' relocation decisions on individuals' own location choices, we exploit the MCMV lottery status as an instrumental variable, which

requires the following assumptions:

Assumption A1. *(Random assignment)* $R_i(z,d), D_{J(i)}(z) \perp\!\!\!\perp Z_{J(i)}$,

Assumption A2. *(Monotonicity in candidate's location choice)* $\Pr(D_j(1) \geq D_j(0)) = 1$.

Assumption A3. *(Relatives' location choice exclusion)* $R_i(z,d) \equiv R_i(d) \quad \forall (d,z)$,

Assumption A1 follows from the random assignment of lotteries to candidates, which we validate in Tables 1 and 2.

In contrast, Assumption A3 imposes strong restrictions on individuals' behaviour. This exclusion restriction assumption implies that non-lottery candidates' location choices are not affected directly by the housing lottery, but only through the moving decision of their relatives who won the lottery. That could be violated if, for example, those individuals lived in the same housing unit with the lottery candidates pre-lottery, so the income and wealth effect associated with the lottery win is likely to affect them directly.

To accommodate that we take the following actions. First, our analysis hereafter restricts attention to relatives who live close by, but not in the same housing unit. This restriction ensures that the exclusion restriction (Assumption A3) is more plausible, as co-resident relatives would directly benefit from the housing subsidy, confounding our ability to isolate the effects of candidates' location choices. For this group of relatives, we are assuming that no relative would move farther away from the candidate and the housing project whenever the candidate moves (Assumption A2).

Additionally, in Appendix Table A.4, we formally test this assumption following Kitagawa (2015). The results show that we cannot reject the null hypothesis of well-defined potential outcomes among compliers for both location choices (Panel A) and labor supply outcomes (Panel B). Specifically, the test examines whether the distributions of potential outcomes are consistent with our exclusion restriction and monotonicity assumptions. Across all outcomes and lotteries, the q-values remain well above conventional significance thresholds, indicating that the data cannot reject Assumptions A2 and A3.

Results. Table 5 reports the intent-to-treat (ITT) effects of winning the housing lottery on relatives' location choices, by their baseline distance from the lottery candidate. Estimates

are obtained from OLS regressions of the outcomes of interest on an indicator for relatives of candidates who won the lottery, controlling for lottery×age-bins fixed effects and for the number of candidates each relative is related to – see Appendix Section C.1 for details. Relatives living in the same household or nearby (columns 1–4) experience direct spillover effects from a lottery win, increasing their distance from the candidate by almost 2.4 km. Those who lived in the same household are 5 percentage points more likely to move closer to the housing project and 19 percentage points less likely to live in the same household post-lottery. Relatives who lived nearby are slightly less affected, with only a 1.5 percentage point increase in the probability of moving near the housing project. In contrast, we find no statistically significant behavioral responses among relatives who lived far from the candidate at baseline (columns 5-6). These patterns highlight heterogeneity in the importance of network proximity: previous location choices strongly predict the likelihood that families follow their relatives when they move.

Interpretation of these intent-to-treat impacts is attenuated by noncompliance, as the first stage estimates from Table 4 suggest that only 15% of the candidates comply and move to the housing project eventually. Therefore, next we report the instrumental variables (IV) estimates, which scale the intent-to-treat estimates by the first-stage estimates. As noted in the previous section, the exclusion restriction is unlikely to hold for relatives who lived with the lottery winners prior to the lottery. Moreover, Table 5 shows that family members who initially lived far from their network did not change their location choices when their relatives won the lottery. Therefore, in the remainder of the analysis, we focus on the sub-sample of relatives—described in columns 4—6 of Table 2—who lived within walking distance but not in the same household. Among these, 16% were related to candidates who comply and move to the housing project (see Appendix Table A.6).

Column 5 in Table 6 shows the IV estimates, which give the local average treatment effects (LATEs) for "compliers", i.e., the effects on the subset of relatives of lottery candidates that follow through and move to the housing project. As above, our TSLS regression accounts for the lottery design and the network endogeneity – see Appendix Section C.3. This table shows that winning a housing lottery increases the probability of close relatives moving closer to the housing project by 15 pp. This is an increase of more than 200% relative to the baseline

Table 5: *Reduced form estimates of candidate's lottery win on relatives' location choice by type*

Variable	Live together		Live nearby		Live far	
	Av. Loser (1)	Est. (2)	Av. Loser (3)	Est. (4)	Av. Loser (5)	Est. (6)
(i) Baseline sample						
Post-lottery address (relative)	0.788	-0.082	0.805	-0.023	0.847	-0.005
		(0.022)		(0.015)		(0.015)
Post-lottery address (both)	0.751	-0.067	0.761	-0.007	0.793	0.021
		(0.022)		(0.017)		(0.018)
# of losers		270313		520737		381898
# of winners		539		978		841
(ii) Post-lottery CU respondents (relatives)						
Lives with candidate	0.812	-0.184	0.029	0.005	0.012	-0.008
		(0.032)		(0.007)		(0.003)
		[-0.181, -0.18]		[-0.009, 0.001]		[-0.011, -0.008]
		(-0.202, -0.163)		(-0.019, 0.008)		(-0.012, -0.002)
Moves from baseline	0.136	0.080	0.209	0.036	0.266	0.038
		(0.024)		(0.018)		(0.020)
		[0.045, 0.073]		[0.02, 0.027]		[0.02, 0.028]
		(0.025, 0.102)		(0.012, 0.037)		(0.014, 0.034)
Moves closer to HP	0.048	0.065	0.074	0.021	0.088	0.017
		(0.017)		(0.012)		(0.014)
		[0.044, 0.067]		[0.015, 0.021]		[0.008, 0.015]
		(0.032, 0.083)		(0.008, 0.032)		(0.001, 0.02)
Distance moved (uncond.)	0.900	1.126	1.158	0.598	2.528	0.291
		(0.316)		(0.210)		(0.299)
		[0.763, 1.464]		[0.031, 1.238]		[-0.803, 0.176]
		(0.275, 2.25)		(-0.97, 2.595)		(-1.474, 1.894)
# of losers		213073		419232		323461
# of winners		378		757		705
(iii) Post-lottery CU respondents (both)						
Lives near candidate	0.094	0.037	0.140	-0.008	0.062	-0.011
		(0.019)		(0.014)		(0.009)
		[0.02, 0.021]		[-0.018, -0.009]		[-0.035, -0.009]
		(0.004, 0.039)		(-0.035, 0.002)		(-0.041, 0.002)
Distance to candidate	1.154	2.348	2.226	2.109	14.751	1.095
		(0.486)		(0.365)		(0.542)
		[2.28, 3.297]		[1.624, 2.518]		[0.873, 1.847]
		(1.897, 3.432)		(1.386, 2.803)		(0.444, 2.137)
# of losers		202973		396297		302933
# of winners		371		734		683

Note: This table presents the effects of winning the housing lottery on location choices of candidates' relatives. 'Av. Loser' shows the average outcome among loser relative-candidate pairs. 'Est.' shows the estimated treatment effects. Standard errors are reported in parentheses, and Lee bounds stratified by lottery are shown in brackets along with its 95% confidence intervals from 500 bootstrap iterations in parentheses. See Appendices C.1 and C.2 for estimation details. Proximity categories are defined by whether the candidate-relative pair was in the same household (live together), within 3 kilometers (live nearby), or farther than 3 kilometers (live far) before the lottery. Post-lottery outcomes are measured at the latest observed date, up to 2020. HP refers to a housing project. CU respondents refer to survey respondents after the lottery. Distance moved includes observations with zero movement.

moving likelihood (Column 4). Relatives who relocate move more than 3 km from their pre-lottery location.

Table 6: *Effect of candidate's move on nearby relatives' location choices*

				Complier	
Variable	Non-mover (1)	RF (2)	OLS (3)	Untreated Av. (4)	LATE (5)
Living together	0.029	0.005	0.063	0.052	0.035
		(0.007)	(0.028)	(0.030)	(0.048)
Relative moved	0.206	0.036	0.104	0.060	0.255
		(0.018)	(0.051)	(0.086)	(0.123)
Relative moved closer to HP	0.073	0.021	0.111	0.040	0.147
		(0.012)	(0.043)	(0.054)	(0.082)
Distance moved	1.133	0.598	2.524	-0.345	4.153
		(0.210)	(0.828)	(0.861)	(1.421)

Note: This table presents estimated effects of the candidate's move (due to winning the housing lottery) on nearby relatives' location choices. 'Non-mover' shows the average outcome for relatives of candidates who did not move. 'OLS' shows ordinary least squares estimates comparing relatives of movers to relatives of non-movers. 'RF' shows reduced-form estimates of the intent-to-treat effect of the candidate winning the lottery. 'Complier Untreated Av.' shows the average outcome for compliers in the absence of candidates' move. 'LATE' shows local average treatment effects estimated via instrumental variables, representing the causal effect of the candidate's move on nearby relatives' location choices for compliers. See Appendix C.3 for estimation details. Clustered standard errors at the candidate level are reported in parentheses. HP refers to housing project. Post-lottery outcomes are measured at the latest observed address, up to 2020.

6 Relatives' Labor Supply Responses

Thus far, our analysis provides strong evidence for preferences to reside near family members. We now turn to understanding the economic consequences of these choices. Does family separation affect relatives' labor market outcomes? For example, separation may increase childcare costs for parents or disrupt job referral networks. Our analysis reveals that family proximity indeed matters for employment outcomes, with particularly strong effects for those who remain behind when candidates relocate.

6.1 Estimated effects

Applying the framework from Section 5 to labor supply outcomes Y_i instead of location choices R_i, we can estimate the effects of candidates' moves on nearby relatives' labor supply (among relatives of compliers). These estimates, which rely on the exclusion restriction and monotonicity assumptions established in Section 5, are presented in Table 7.

Reduced-form estimates (Column 2) show that relatives of lottery winners increase informal employment by 2.2 percentage points, while decreasing formal employment by 2.5 percentage points, consistent with a small increase in PBF receipt. The contrast between relatives of candidates who did not move (Column 1) and relatives of compliers who did not move (Column 4) reveals important selection patterns. Although both groups have similar overall employment levels (around 61%), relatives of compliers are much more likely to be formally employed when the candidate doesn't move (42.8% vs. 29.8%), and are less dependent on social transfers (16.6% vs. 43.2% PBF receipt). This selection pattern is important for interpreting our results: relatives of compliers represent a specific subset of the population with stronger formal labor market attachment.

TSLS estimates of this LATE (Column 5) reveal that these modest reduced-form effects translate into economically large impacts on labor supply when candidates choose to move to a housing project. While we find no statistically significant effect on overall employment, the point estimate is negative and economically meaningful (-9.6 percentage points). This aggregate effect masks significant compositional changes: formal employment decreases by a statistically significant 15.7 percentage points, while informal employment increases by an insignificant 8.2 percentage points. Consistent with this shift toward informal work, PBF receipt surges by 24.9 percentage points, indicating increased reliance on social transfers. These represent substantial changes: the 15.7 percentage point decrease in formal employment constitutes a 37% decline from the baseline rate, while the 24.9 percentage point increase in PBF receipt represents a 150% increase in social transfer dependence. Furthermore, these TSLS estimates differ meaningfully from the OLS estimates (Column 3), with OLS substantially underestimating the negative effects of family separation. This suggests that candidates who choose to move have relatives who would experience smaller negative effects, consistent with selection based on family circumstances.

6.2 Extended Framework

The LATE estimated in the previous section provides valuable evidence that family separation affects relatives' labor supply. However, this estimate combines two distinct mechanisms: the effect on relatives who remain behind when candidates move (pure separation effects) and

Table 7: *Effect of candidates' move on nearby relatives' labor supply*

				Complier			
Variable	Non-taker (1)	RF (2)	OLS (3)	Untreated Av. (4)	LATE (5)	$LATE_0$ (6)	$LATE_1$ (7)
Any employment	0.611	-0.007	-0.09	0.612	-0.096	[-0.246, -0.119]	[-0.374, 0.375]
		(0.014)	(0.030)	(0.082)	(0.084)		
Any informal employment	0.312	0.022	-0.025	0.163	0.082	[0.144, 0.261]	[-0.389, 0.298]
		(0.013)	(0.030)	(0.077)	(0.079)		
Any formal employment	0.298	-0.025	-0.063	0.428	-0.157	[-0.391, -0.263]	[-0.673, 0.076]
		(0.010)	(0.024)	(0.070)	(0.067)		
Total formal earnings (uncond.)	1672.261	-102.115	-227.187	2299.248	-545.814	[-2042.9, -1903.6]	[-2047.8, -1228.1]
		(91.261)	(250.617)	(621.595)	(594.102)		
Number of jobs (uncond.)	1.465	-0.107	-0.281	2.068	-0.68	[-2.025, -1.952]	[-0.877, -0.45]
		(0.067)	(0.176)	(0.436)	(0.441)		
Kept same formal job (uncond.)	0.047	-0.013	-0.01	0.105	-0.071	[-0.089, 0.033]	[-0.669, 0.045]
		(0.006)	(0.013)	(0.036)	(0.038)		
Has new formal job (uncond.)	0.236	-0.018	-0.048	0.349	-0.118	[-0.326, -0.199]	[-0.674, 0.075]
		(0.011)	(0.025)	(0.071)	(0.072)		
Receives PBF	0.432	0.041	-0.007	0.166	0.249	[0.166, 0.293]	[0.027, 0.776]
		(0.016)	(0.045)	(0.103)	(0.111)		

Note: This table presents estimated effects of the candidate's move to subsidized housing on nearby relatives' labor supply outcomes. 'Non-mover' shows the average outcome for relatives of candidates who did not move. 'OLS' shows ordinary least squares estimates comparing relatives of movers to relatives of non-movers. 'RF' shows reduced-form estimates of the intent-to-treat effect of the candidate winning the lottery. 'Complier Untreated Av.' shows the average outcome for compliers in the absence of candidates' move. 'LATE' shows local average treatment effects estimated via instrumental variables, representing the causal effect of the candidate's move on nearby relatives' outcomes. $LATE_0$ and $LATE_1$ show estimated sharp bounds for the sub-LATEs (see main body). See Appendices C.3 and C.4 for estimation details. Standard errors clustered at the candidate level are reported in parentheses, and bounds are shown in brackets.

the effect on relatives who relocate to follow candidates (which combines moving costs with potential proximity benefits). Understanding which mechanism dominates has important policy implications, as housing programs could be designed to facilitate family coordination or provide additional support for separated family members.

To distinguish between these mechanisms, we extend our framework in Section 5 to allow potential outcomes to depend on the location choices of both candidates and their relatives. Let R_i be the location choice to move near the housing project of the lottery. Let $Y_i(d, r)$ represent the potential labor market outcome for relative i, where $d \in \{0, 1\}$ indicates whether the candidate moves to the housing project and $r \in \{0, 1\}$ indicates whether the relative moves closer to the housing project. The observed outcome is then $Y_i = Y_i(D_{J(i)}, R_i)$, emphasizing that family proximity–and its economic consequences–depends on the decisions of both parties. For example, if family proximity provides positive labor market benefits (through childcare, job referrals, etc.), we would expect $Y_i(0, 0) \geq Y_i(1, 0)$ – that is, a relative experiences worse labor market outcomes when the candidate moves away while the relative stays behind, compared to when both remain in their original locations.

We make the following assumptions:

Assumption Y1. *(Random assignment)* $Y_i(d, r) \perp\!\!\!\perp Z_{J(i)}$,

Assumption Y2. *(Potential outcome exclusion)* $Y_i(d, r, z) \equiv Y_i(d, r) \quad \forall (d, r, z)$,

Assumption Y3. *(Monotonicity in relative's location choice)* $\Pr(R_i(1) \geq R_i(0)) = 1$,

Assumption Y1 follows from the random assignment of lotteries. Assumption Y2 requires that lottery wins affect relatives' labor supply only through their effects on location choices of candidates and relatives, ruling out direct wealth or income effects from the lottery win itself – similar to Assumption A2. Assumption Y3 requires that relatives are weakly more likely to move closer to the housing project when their candidate relocates there.

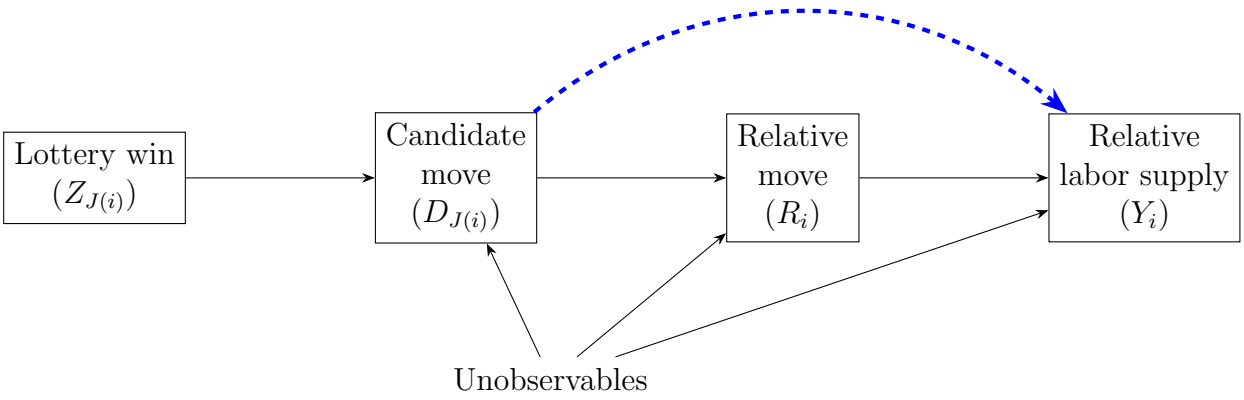
Focusing on relatives living nearby (but not cohabitating) before the lottery makes these assumptions more plausible. Assumption Y2 is credible because these relatives do not directly benefit from the housing subsidy, making direct wealth effects unlikely. Assumption Y3 is supported by our findings in Section 5, which show that relatives are indeed more likely to move closer to housing projects when candidates relocate there.

Additionally, we assess the testable implications of Assumptions Y2 and Y3 on this sample: the distributions of the potential labor supply outcomes among relatives should be well defined (Kitagawa, 2015). Panel B of Appendix Table A.4 provides support for these assumptions, as we cannot reject the null hypothesis that potential outcomes are well-defined random variables. Similarly, Appendix Table A.5 shows that we cannot reject the null hypothesis that the potential outcomes among relative groups defined by their location choice types (see Section 6.3) are well-defined random variables.

Figure 3 intuitively explains the challenge of this setting, plotting the implied Directed Acyclic Graph of our 2-level instrumental variable model given our assumptions. In the context of mediation analysis, as discussed by Pearl (2022), $D_{J(i)}$ and R_i can be thought of as mediators, where our model allows the lottery win to influence outcomes only through its effect on them. In this context, the mediators correspond to the location choices of candidates and their relatives following the lottery outcome. However, the candidate's move $D_{J(i)}$ can affect relative's labor supply Y_i through two distinct pathways: directly (as depicted by the blue line), capturing pure separation effects when relatives remain behind, and indirectly

through its effect on the relative's own location choice R_i, capturing the labor market consequences of induced mobility. Disentangling the relative importance and magnitude of these direct and indirect effects–which have different policy implications–constitutes the remainder of the paper.

Figure 3: *DAG*



Note: This directed acyclic graph illustrates the 2-level instrumental variable model used in the analysis. The lottery win ($Z_{J(i)}$) serves as an instrumental variable that influences relatives' labor supply outcomes (Y_i) only through its effects on the candidate's move to subsidized housing ($D_{J(i)}$) and the relative's subsequent location choice (R_i). Solid arrows represent the assumed causal pathways, while the dashed blue arrow represents a potential direct effect from the candidate's move to the relative's labor supply that bypasses the relative's location choice. Unobservables represent unmeasured factors that may jointly influence candidate moves, relative moves, and relative labor supply.

Identification. With two treatment variables ($D_{J(i)}$ and R_i) and only one instrument ($Z_{J(i)}$), this model is underidentified. As illustrated elsewhere in the literature, the conventional IV of $D_{J(i)}$ on Y_i using $Z_{J(i)}$ as an instrument identifies (Kirkeboen et al., 2016; Mountjoy, 2022; Dutz et al., 2024):

$$\begin{aligned}\beta_{IV} :=& \frac{\mathbb{E}[Y_i|Z_{J(i)}=1]-\mathbb{E}[Y|Z_{J(i)}=0]}{\mathbb{E}[D_{J(i)}|Z_{J(i)}=1]-\mathbb{E}[D_{J(i)}|Z_{J(i)}=0]} = \mathbb{E}[Y_i(1,R)-Y_i(0,R)|T_{J(i)}=C] \qquad (1)\\ &= \mathbb{E}[Y_i(1,0)-Y_i(0,0)|R_i(1)=R_i(0)=0, T_{J(i)}=C]\,\mathbb{P}[R_i(1)=R_i(0)=0|T_{J(i)}=C]\\ &+ \mathbb{E}[Y_i(1,1)-Y_i(0,1)|R_i(1)=R_i(0)=1, T_{J(i)}=C]\,\mathbb{P}[R_i(1)=R_i(0)=1|T_{J(i)}=C]\\ &+ \mathbb{E}[Y_i(1,1)-Y_i(0,0)|R_i(1)>R_i(0), T_{J(i)}=C]\,\mathbb{P}[R_i(1)>R_i(0)|T_{J(i)}=C]\end{aligned}$$

where $T_{J(i)} = C := \mathbb{1}[D_{J(i)}(1) > D_{J(i)}(0)]$.

In other words, the IV estimand is a weighted average of *different* treatment effects on *dif-*

ferent groups of relatives (among relatives of complier candidates). The groups are defined by the potential choices relatives would make under different candidate location choices. In particular, relatives either never move regardless of the candidate's choice ($R_i(0) = R_i(1) = 0$), always move regardless of the candidate's choice ($R_i(0) = R_i(1) = 1$), or move only when the candidate moves($R_i(1) > R_i(0)$). We accordingly refer to these 'location choice types' as 'stayers', 'always-movers', and 'followers', respectively. Relatives experience different treatment effects based on their type: stayers face pure separation effects when candidates relocate, always-movers experience moving costs regardless of candidate behavior, and followers face the combined effect of induced mobility and potential proximity benefits. Although these effects are not point-identified, the population share of each type and the types' average characteristics are (Dutz et al., 2024). We next use our data to describe the type shares and their characteristics.

6.3 Characterizing relatives by their location choice types

As shown in Section 5, 16% of relatives living nearby candidates before the lottery were related to compliers. Appendix Table A.6 further shows that, among them, the majority of relatives (85.5%) are classified as 'stayers', while 10.5% are 'followers' and only 4% are 'always-movers' (and this share is statistically insignificant).[11] This pattern suggests that most relatives have strong location attachment or face significant mobility constraints, with only a small fraction able or willing to relocate in response to candidates' moves.

Given the small share of always-movers, we pool them with followers into a single group of 'movers' for the subsequent analyses. Table 8 compares stayers and movers across several observable dimensions.[12] The first pair of columns reports the characteristics of relatives of lottery compliers. Compared to the complier candidates with relatives living nearby (recall columns (5) and (6) of Table 4), relatives are younger, more likely to be men, have fewer kids, and were less likely to be employed before the lottery.

[11] We estimate these shares via TSLS regressions with judiciously chosen outcomes and endogenous variables, with Z as instrument and including the relevant controls (See Appendix C.3 for details). Some of the estimates reported in Appendix Table A.6 are reported in Table 6. Under our assumptions, the TSLS estimate of the candidate's move on nearby relatives' moving closer to HP (third row) characterizes the share of followers, while the untreated average among compliers characterizes the share of always-movers.

[12] These characteristics are estimated using TSLS regressions with interactions of characteristics, location choices, and candidate moves as outcomes. See Appendix C.3 for technical details.

Stayers and movers exhibit systematic differences in their demographic and economic characteristics. On average, stayers are older, more likely to be women, and more educated than movers. Paradoxically, despite higher education levels, stayers are less likely to be formally employed before the lottery than movers. This pattern suggests that stayers may face different labor market constraints or have different work preferences, possibly related to caregiving responsibilities or local employment networks.

The two groups also differ significantly in their family relationships and household structure. Movers are more likely to be parents or children of the candidate, suggesting that the closest family relationships drive location coordination. In contrast, stayers are more likely to be siblings or extended family members. Additionally, movers are more likely to be married and to have children in their household, indicating that entire family units tend to move together when candidates relocate.

These patterns suggest that mobility decisions reflect a complex interplay of economic constraints, family obligations, and relationship proximity, with implications for how family separation affects different types of relatives.

6.4 Bounding effects by relative type

We have demonstrated that most relatives of candidates living nearby before the lottery decide to 'stay behind' after the candidates move, and that these relatives differ in several dimensions from those who do move toward the housing project. We now investigate how candidates' moves affect labor supply differently for relatives in these two groups.

Pooling all 'movers', we can rewrite the IV estimand as

$$\beta_{IV} = LATE_0 \, \mathbb{P}[R_i(1) = 0 | T_{J(i)} = C] + LATE_1 \, \mathbb{P}[R_i(1) = 1 | T_{J(i)} = C] \tag{2}$$

where $LATE_0 = \mathbb{E}[Y_i(1,0) - Y_i(0,0) | R_i(1) = R_i(0) = 0, T_{J(i)} = C]$ captures the effect on relatives who 'stay behind' and $LATE_1 := \mathbb{E}[Y_i(1,R) - Y_i(0,R) | R_i(1) = 1, T_{J(i)} = C]$ represents the (weighted) average effect on the groups of relatives who 'move with the candidate' given their location choice.

This decomposition is particularly informative because it isolates the effect of the candidate's location choice conditional on the relative's own mobility decision. $LATE_0$ captures

Table 8: *Characteristics of relatives to compliers, by relative location choice*

	Relatives to complier		Stayer relative		Mover relative	
Variable	Est.	(s.e.)	Est.	(s.e.)	Est.	(s.e.)
Panel A: Characteristics						
Age	32.679	(1.401)	33.316	(1.456)	28.94	(3.663)
Gender: Male	0.399	(0.033)	0.383	(0.038)	0.479	(0.111)
Illiterate	0.054	(0.017)	0.048	(0.018)	0.097	(0.066)
Race: Black	0.244	(0.049)	0.245	(0.052)	0.145	(0.099)
PBF Beneficiary	0.594	(0.057)	0.608	(0.066)	0.621	(0.104)
Any employment before lottery	0.557	(0.041)	0.563	(0.044)	0.574	(0.111)
Formal employment before lottery	0.353	(0.044)	0.326	(0.046)	0.526	(0.113)
Distance to closest HP	12.559	(1.241)	12.615	(1.305)	13.031	(2.436)
Panel B: Relationship to candidate						
Parent	0.088	(0.023)	0.089	(0.026)	0.097	(0.063)
Sibling	0.216	(0.032)	0.227	(0.035)	0.144	(0.066)
Child	0.357	(0.052)	0.318	(0.056)	0.52	(0.114)
Other	0.338	(0.039)	0.366	(0.043)	0.24	(0.088)
Panel C: Relatives network						
Distance to candidate	0.788	(0.098)	0.78	(0.104)	0.743	(0.19)
Has spouse	0.203	(0.044)	0.211	(0.046)	0.192	(0.091)
N relatives in household	1.881	(0.149)	1.827	(0.155)	2.444	(0.394)
N kids below 16 in hh	0.265	(0.064)	0.277	(0.07)	0.24	(0.116)
N relatives nearby	1.419	(0.248)	1.476	(0.267)	1.342	(0.377)
N parents nearby	0.027	(0.016)	0.033	(0.02)	0	(0)
N siblings nearby	0.299	(0.073)	0.335	(0.086)	0.145	(0.076)
N inlaws nearby	0.279	(0.085)	0.372	(0.105)	0.337	(0.252)

Note: This table presents baseline characteristics of complier lottery participants' relatives by their endogenous post-lottery location choice. Given one-sided compliance, the column 'Compliers' describes relatives of individuals who won the lottery, signed, and moved to the sorted housing unit. See Appendix C.3 for estimation details. Standard errors are clustered at the candidate level and reported in parentheses.

pure separation effects for relatives who would stay behind regardless of the candidate's choice, while $LATE_1$ identifies the impact on relatives who would move when the candidate moves, capturing the net cost/benefit of relocation costs and potential proximity benefits.

While we cannot point-identify these sub-LATEs, informative sharp bounds under these assumptions are derived by Dutz et al. (2024), which could provide meaningful insights into the heterogeneous effects.[13] Columns 6 and 7 of Table 7 present estimated bounds of $LATE_0$

[13]By informative bounds, we mean bounds that are tight enough to possibly provide meaningful economic insights: they need not include zero, though they may still include zero for some outcomes. By sharp we

and $LATE_1$, revealing stark differences between the two groups. For relatives who stay behind, we find tightly bounded effects that exclude zero for most labor market outcomes.[14] The tightness of these bounds is expected since they represent the vast majority of relatives living nearby, in our case. For them, employment falls by 11.9 to 24.6 percentage points, formal employment decreases by 26.3 to 39.1 percentage points, while informal employment increases by 7 to 19.7 percentage points. Social transfer dependence increases substantially, with PBF receipt rising by 16.6 to 29.3 percentage points.

In contrast, for relatives who move with the candidate, the bounds are wide and include zero for all outcomes except PBF receipt. While these bounds do not preclude large negative effects, their width reflects the difficulty of separately identifying moving costs from proximity benefits for this smaller group.

These results demonstrate that family separation, rather than moving costs, drives the aggregate negative labor supply effects documented in Section 6.1. The tight, meaningful bounds for $LATE_0$ provide clear evidence that housing policies generate substantial unintended consequences for family members who remain behind, suggesting potential benefits from policies that facilitate family coordination or provide targeted support for separated families. Our results also highlight that learning about the specific moving costs faced by movers would require additional assumptions in our setting.

7 Selection Model *[work in progress]*

The reduced-form and bounds estimates show that a recipient's relocation affects both the residential choices and labor-market outcomes of nearby relatives. An important remaining question is whether relatives' location choices are explained by selection on gains—that is, whether the relatives who choose to follow are those who would face the largest labor supply reductions from separation. We plan to address this question using a Roy-style selection framework in which each relative faces a choice between two location regimes after the recipient moves: remaining in place or relocating to restore proximity. The testable implication is the sign of selection on labor market gains from proximity. If selection is positive—followers

mean they extract the maximum available information from the data, given the assumptions imposed.

[14] We detail the estimation methodology in Appendix C.4.

tend to be those with higher returns to proximity—then the relocation margin itself provides evidence that family proximity has production value, and that households anticipate the labor market consequences of separation when making location decisions. If selection is not positive, the value of proximity is more naturally interpreted as consumption value or as operating through margins not captured by our labor market measures. The estimation of this framework is in progress.

8 Conclusion

Using housing lotteries in Brazil's *Minha Casa Minha Vida* program and linked administrative data on extended family networks and labor market outcomes, we show that family proximity causally affects both location choices and labor market outcomes. Take-up is correlated with pre-existing family geography: winners whose relatives live nearby are substantially less likely to relocate into subsidized housing. While this evidence is descriptive rather than causal, it is consistent with family separation being an important margin in program take-up.

Using exogenous variation in relocation based on the MCMV lotteries, we show that when recipients move, there is a causal effect on their relatives' location choices: about 15% of nearby non-co-resident relatives subsequently relocate closer to the housing project.

We also provide evidence on the economic consequences of family separation. A recipient's move reduces formal employment and increases transfer receipt among nearby non-co-resident relatives, with the principal margin being a reallocation away from formal work rather than a wholesale exit from employment. To address the fact that relatives' moving decisions are endogenous, we develop an extended instrumental-variables framework that decomposes these effects by relatives' own location responses. The employment reductions are driven by relatives who remain behind rather than by the consequences of induced mobility for those who follow. In ongoing work, we test whether selection on labor market gains can explain relatives' location decisions—that is, whether relatives who follow are those who would face the largest labor supply reductions from staying behind. If so, the selection margin would indicate that family proximity provides production value—support for work through childcare, job information, or informal insurance—rather than consumption value

alone. This has the potential to shed light on whether formal institutions can substitute for proximity-based in-kind transfers between family members. If not, the value of proximity is more naturally interpreted as reflecting non-labor channels.

These findings have three broader implications. First, the welfare effects of housing and relocation policies may be mismeasured when evaluations focus only on direct recipients. If relocating one household reduces the formal employment of relatives left behind, then standard cost-benefit calculations miss potentially important family spillovers. Second, the results bear on whether formal institutions can substitute for the support that family networks provide. If the production-value hypothesis is confirmed, expanding access to formal childcare or job search assistance could reduce the consequences of geographic separation and increase residential mobility among low-income households. Third, the results suggest that models of residential choice and migration may benefit from accounting for interdependencies in family location decisions, as one family member's move directly affects others' optimal locations and labor market outcomes.

The selection analysis described above is the immediate next step for this project. Beyond that, several directions remain open. One is to study heterogeneity by type of family relationship—for example, whether the effects of separation differ for parents, siblings, and in-laws, which may help identify the specific channels through which proximity operates. Another is to measure more directly the services family members provide; some of the administrative data sources we are integrating may offer direct observation of in-kind transfers between relatives, which would complement the revealed-preference evidence from the selection analysis.

References

Altonji, J. G. and Card, D. (1991). The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives. In Abowd, J. and Freeman, R. B., editors, *Immigration, Trade and Labor*. University of Chicago Press. 1

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., and Wong, A. (2018). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–280. 1

Bailey, M., Farrell, P., Kuchler, T., and Stroebel, J. (2020). Social connectedness in urban areas. *Journal of Urban Economics*, 118:103264. 1

Bayer, P., McMillan, R., Murphy, A., and Timmins, C. (2016). A dynamic model of demand for houses and neighborhoods. *Econometrica*, 84(3):893–942. 5

Bazzi, S. (2017). Wealth heterogeneity and the income elasticity of migration. *American Economic Journal: Applied Economics*, 9(2):219–255. 1

Belchior, C., Gonzaga, G., and Ulyssea, G. (2023). Unpacking neighborhood effects: Experimental evidence from a large-scale housing program in brazil. *Working Paper*. 4

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300. 13

Blumenstock, J. E., Chi, G., and Tan, X. (2025). Migration and the value of social networks. *Review of Economic Studies*, 92(1):97–128. 1

Borusyak, K. and Hull, P. (2020). Non-random exposure to exogenous shocks: Theory and applications. Technical report, National Bureau of Economic Research. 14

Bouvier, M., Vanek, J., and Roubaud, F. (2022). Informal workers in brazil: A statistical profile. Statistical Brief 33, Women in Informal Employment: Globalizing and Organizing (WIEGO). 12

Bryan, G., Chowdhury, S., and Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica*, 82(5):1671–1748. 1

Bryan, G. and Morten, M. (2019). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy*, 127(5):2229–2268. 1, 5

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

Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review*, 108(10):3028–3056. 4

Compton, J. and Pollak, R. (2014). Family Proximity, Childcare, and Women's Labor Force Attachment. *Journal of Urban Economics*, 79:72–90. 1

da Mata, D. and Mation, L. (2018). Labor market effects of public housing: Evidence from large-scale lotteries. Technical report, REAP, Working Paper. 4

da Mata, D. and Mation, L. (2025). Housing policies and social mobility. 4

Dutz, D., Humphries, J. E., Lacouture, S., Stapleton, S., and van Dijk, W. (2024). Allocating short-term rental assistance by targeting temporary shocks. 3, 27, 28, 30, 22

Jacob, B. A. and Ludwig, J. (2012). The effects of housing assistance on labor supply: Evidence from a voucher lottery. *American Economic Review*, 102(1):272–304. 4

Kennan, J. and Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, 79(1):211–251. 5

Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection. *The quarterly journal of economics*, 131(3):1057–1111. 3, 27

Kitagawa, T. (2015). A test for instrument validity. *Econometrica*, 83(5):2043–2063. 20, 26, 21

Koenen, M. and Johnston, D. (2025). Social networks and residential choice: Micro evidence and equilibrium implications. Job Market Paper, Harvard University. 1

Lagakos, D., Mobarak, A. M., and Waugh, M. E. (2023). The welfare effects of encouraging rural–urban migration. *Econometrica*, 91(3):803–837. 1

Lee, D. S. (2005). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. 16, 17, 18

Machado, C. and Rachter, L. (2020a). Better neighborhoods or better houses? the effects of housing policies on poor households in brazil. *Working Paper*. 4

Machado, C. and Rachter, L. (2020b). The effects of better houses on infant health. *Working Paper*. 4

Mountjoy, J. (2022). Community colleges and upward mobility. *American Economic Review*, 112(8):2580–2630. 3, 27

Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market. *Quarterly Journal of Economics*, 118(2):549–599. 1

Munshi, K. (2020). Social networks and migration. *Annual Review of Economics*, 12(1):503–524. 1

Pearl, J. (2022). Direct and indirect effects. In Geffner, H., Dechter, R., and Halpern, J. Y., editors, *Probabilistic and causal inference: the works of Judea Pearl*, ACM Books, pages 373–392. Association for Computing Machinery, New York, NY, USA. 26

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy*, 90(6):1257–1278. 5

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55. 5

Rosenzweig, M. R. and Stark, O. (1989). Consumption smoothing, migration, and marriage: Evidence from rural india. *Journal of political Economy*, 97(4):905–926. 1

Sahai, H. and Bailey, M. (2022). Social networks and internal migration: Evidence from facebook in india. *Meta*. 1

Semenova, V. (2020). Generalized lee bounds. *arXiv preprint arXiv:2008.12720*. 16, 17, 18, 19

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(3):479–498. 13, 2, 3, 5

Stuart, B. and Taylor, E. (2021). Migration Networks and Location Decisions: Evidence from U.S. Mass Migration . *American Economic Journal: Applied Economics*, 13(3):134–175. 1

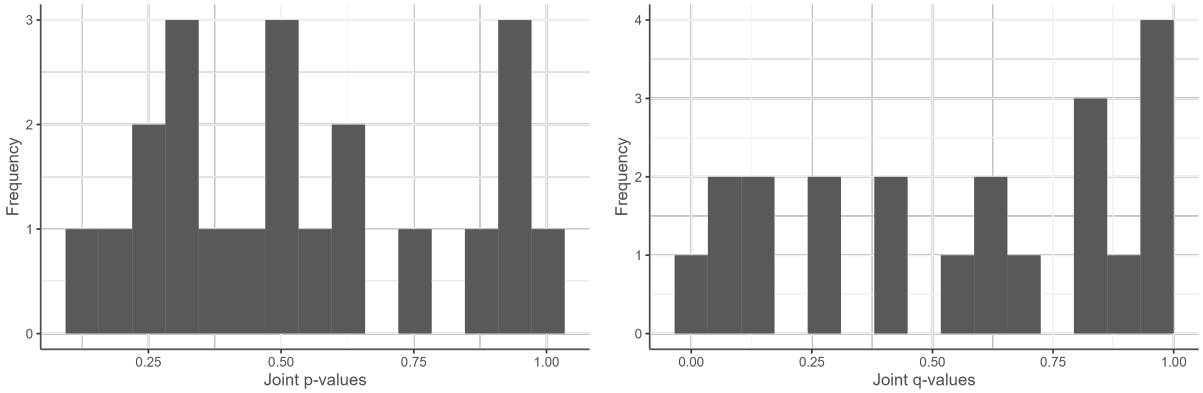
A Additional Figures and Tables

Table A.1: *Analysis sample size by lottery*

	All relationships (over 16)				Living Nearby			
	Candidates		Candidates×relatives		Candidates		Candidates×relatives	
Lottery	All	Winners	All	Winners	All	Winners	All	Winners
2013-003	19030	31	33941	53	9403	8	15507	12
2013-006	19784	110	35370	210	9761	53	16149	92
2015-001	44660	49	92379	103	22761	25	41207	42
2015-004	44586	166	92258	322	22738	68	41177	114
2015-007	40165	11	82815	25	20326	8	36775	12
2015-014	1011	10	2264	29	557	6	1132	17
2015-018	40398	82	83309	162	20442	38	36997	64
2015-019	40339	45	83207	84	20414	21	36946	40
2015-028	40967	60	84630	114	20749	24	37608	36
2015-029	2345	29	5146	64	1297	18	2419	36
2015-030	1051	30	2415	58	547	16	1082	30
2015-031	1035	6	2323	13	572	2	1165	3
2015-034	41619	23	86005	59	21087	11	38217	30
2015-036	1050	10	2362	21	574	7	1173	12
2016-003	4825	46	11044	116	2614	27	5007	49
2016-004	7041	5	15839	15	3694	4	7033	10
2016-005	50156	28	109196	60	25688	17	47974	24
2016-006	50167	61	109229	117	25698	26	47984	41
2016-007	11918	35	26968	98	6336	17	12071	47
2016-017	50225	217	109365	481	25730	101	48049	190
2016-018	4937	18	11326	40	2671	13	5120	25
2016-020	42229	47	91557	114	21351	25	39945	52
Total	560678	1140	1172948	2358	285010	535	520737	978

Note: This table shows the sample composition across the 22 housing lotteries examined in the analysis. 'Candidates' refers to the number of lottery participants, while 'Candidates×relatives' refers to the number of candidate-relative pairs where relatives are over 16 years old. 'All relationships' includes all candidate-relative pairs regardless of proximity, while 'Living Nearby' restricts to pairs where relatives live within 3 kilometers of the candidate at baseline. 'All' includes both lottery winners and losers, while 'Winners' shows only those who won the lottery.

Figure A.1: *Balance across lotteries: joint p- and q-values. Candidates*



(a) *p-values* ***(b)*** *q-values*

Note: This figure displays the distribution of p-values (panel a) and q-values (panel b) from joint tests of balance across all baseline characteristics for lottery candidates in each of the 22 lotteries examined. *p*-values test the null hypothesis of no mean differences between winners and losers within each lottery. *q*-values represent the empirical Bayes posterior probability that a lottery is falsely classified as not random, following Storey (2002), and provide a measure of the false discovery rate when accounting for multiple testing across lotteries. The analysis includes tests across all pre-lottery characteristics shown in Table 1.

Table A.2: *Balance test on candidates' characteristics by distance from relatives*

	Living together			Living nearby			Living far		
Variable	Loser (1)	Winner (2)	Diff. (3)	Loser (1)	Winner (2)	Diff. (3)	Loser (1)	Winner (2)	Diff. (3)
Age	40.046	42.591	0.028	38.377	42.584	0.101	39.088	40.677	0.025
			(0.124)			(0.107)			(0.109)
			[1]			[0]			[4]
Gender: Male	0.204	0.161	-0.027	0.196	0.187	0.01	0.208	0.177	-0.027
			(0.021)			(0.021)			(0.023)
			[2]			[1]			[6]
Disabled	0.042	0.016	-0.027	0.035	0.032	-0.007	0.038	0.044	0.005
			(0.006)			(0.008)			(0.011)
			[10]			[7]			[4]
Illiterate	0.028	0.031	0.001	0.031	0.061	0.025	0.033	0.045	0.01
			(0.01)			(0.013)			(0.012)
			[2]			[2]			[2]
Race: Black	0.196	0.212	0.011	0.223	0.277	0.052	0.215	0.24	0.023
			(0.025)			(0.024)			(0.034)
			[1]			[2]			[1]
PBF Beneficiary	0.501	0.527	0.016	0.581	0.578	-0.007	0.599	0.627	0.03
			(0.028)			(0.025)			(0.03)
			[2]			[1]			[5]
N relatives	4.234	4.318	0.058	5.765	5.754	0.142	7	7.428	0.486
			(0.215)			(0.241)			(0.373)
			[3]			[2]			[3]
N children	1.419	1.657	0.045	1.561	1.704	-0.051	1.588	1.597	-0.043
			(0.068)			(0.075)			(0.086)
			[0]			[0]			[2]
N parents	0.562	0.405	-0.027	0.47	0.269	-0.039	0.337	0.317	0.031
			(0.035)			(0.028)			(0.05)
			[10]			[2]			[4]
Has kids	0.663	0.779	0.047	0.705	0.765	-0.012	0.719	0.709	-0.027
			(0.021)			(0.023)			(0.028)
			[15]			[2]			[4]
Has kids under 16	0.419	0.456	0.003	0.537	0.563	0.008	0.604	0.601	-0.001
			(0.026)			(0.022)			(0.027)
			[1]			[0]			[1]
Lives with relative over 16	1	1	0	0.076	0.091	0.015	0.124	0.119	-0.003
			(0)			(0.015)			(0.02)
			[0]			[7]			[6]
Any relative living nearby	0.127	0.153	0.021	1	1	0	0.267	0.216	-0.053
			(0.023)			(0)			(0.027)
			[4]			[0]			[1]
Any relative living far	0.144	0.184	0.038	0.201	0.21	0.012	1	1	0
			(0.024)			(0.023)			(0)
			[3]			[1]			[0]
Distance to closest HP	17.075	17.385	0.649	16.679	14.784	-0.076	16.361	14.846	-0.317
			(0.38)			(0.357)			(0.545)
			[2]			[0]			[5]
Av. Distance to relatives	4.509	2.044	-2.515	7.655	4.316	-2.994	82.189	47.746	-31.83
			(1.032)			(1.606)			(13.627)
			[7]			[9]			[12]
Any employment before lottery	0.764	0.772	0.03	0.791	0.745	-0.017	0.793	0.788	0.006
			(0.022)			(0.023)			(0.026)
			[1]			[5]			[3]
Formal employment before lottery	0.657	0.631	0.014	0.67	0.593	-0.021	0.652	0.604	-0.02
			(0.027)			(0.025)			(0.031)
			[1]			[3]			[1]
Total formal earnings before lottery	2087.605	2353.795	435.073	2057.627	2061.984	151.971	1971.744	2204.792	319.244
			(240.301)			(240.857)			(449.167)
			[1]			[4]			[2]
N	270313	539		520737	978		381898	841	

Note: This table reports the mean characteristics and differences of MCMV Rio lottery winners and losers candidate-relative pairs, in 22 general lotteries conducted between years 2013-2016. Demographic characteristics, informal labor market outcomes, and BFI attainment are from the Cad ´Unico survey completed prior to the lottery, using data from the survey closest to (but preceding) the lottery date. Formal labor market outcomes are from RAIS employer-employee dataset. The difference and the standard errors (in parentheses) are from a regression with lottery×age-bins fixed effect, with standard errors clustered at the individual level, See Appendix C.1 for estimation details. The number of lotteries with q-values (Storey, 2002) lower than 10% for testing the null of no mean differences within each lottery is reported in squared brackets below estimates.

Table A.3: *Effect of lottery win on candidates' location choice: relative-candidate pair by distance at baseline*

	Live together		Live nearby		Live far	
Variable	Av. Loser	Est.	Av. Loser	Est.	Av. Loser	Est.
(i) Baseline sample						
		(0.001)		(0.001)		(0.001)
Signs contract	0.000	0.571	0.000	0.554	0.000	0.513
		(0.001)		(0.001)		(0.001)
Address post-lottery	0.903	0.029	0.916	0.026	0.926	0.035
		(0.012)		(0.008)		(0.009)
(ii) Post-lottery CU respondents						
Moves from baseline	0.156	0.191	0.236	0.196	0.303	0.161
		(0.016)		(0.014)		(0.016)
		[0.163, 0.200]		[0.173, 0.202]		[0.141, 0.175]
		(0.144, 0.216)		(0.16, 0.215)		(0.127, 0.187)
Moves more than 2 kms	0.075	0.190	0.102	0.181	0.202	0.163
		(0.012)		(0.010)		(0.014)
		[0.158, 0.195]		[0.157, 0.186]		[0.138, 0.172]
		(0.138, 0.21)		(0.142, 0.198)		(0.125, 0.184)
Moves closer to HP	0.058	0.166	0.088	0.173	0.110	0.150
		(0.010)		(0.009)		(0.011)
		[0.135, 0.172]		[0.156, 0.185]		[0.128, 0.162]
		(0.114, 0.186)		(0.141, 0.197)		(0.114, 0.174)
Moves to HP	0.004	0.140	0.006	0.156	0.006	0.148
		(0.003)		(0.003)		(0.003)
		[0.109, 0.146]		[0.138, 0.166]		[0.122, 0.156]
		(0.09, 0.16)		(0.123, 0.178)		(0.107, 0.167)
Signed and Moves to HP	0.000	0.139	0.000	0.160	0.000	0.154
		(0.001)		(0.001)		(0.001)
		[0.107, 0.144]		[0.136, 0.165]		[0.125, 0.159]
		(0.088, 0.158)		(0.121, 0.177)		(0.11, 0.17)
Distance moved (uncond.)	1.047	3.144	1.364	2.602	2.983	2.629
		(0.187)		(0.153)		(0.243)
		[1.683, 3.075]		[1.389, 2.503]		[1.109, 2.519]
		(0.846, 5.651)		(0.684, 4.845)		(0.491, 3.572)

Note: This table presents the effects of winning the housing lottery on candidates' location choices, by baseline proximity to relatives, at the pair level. "Av. Loser" shows the average outcome among lottery loser pairs. "Est." shows the estimated treatment effects (difference between winners and losers), its standard error, the estimated Lee bounds, and the 95% CI for the bound endpoints using 500 bootstrap iterations. See Appendices C.1 and C.2 for estimation details. Proximity categories are defined by whether participants have relatives over 16 years old in the same household (live together), within 3 kilometers (live nearby), or farther than 3 kilometers (live far). Post-lottery outcomes are measured at the latest observed date, up to 2020. CU respondents refers to survey respondents after the lottery. Clustered standard errors at the candidate level are reported in parentheses, and Lee bounds stratified by lottery are shown in brackets.

Table A.4: *Average potential outcomes for relatives nearby candidate*

	Treated Complier			Untreated Complier		
Outcome	Est. (1)	(s.e.) (2)	# $\hat{q} < .1$ (3)	Est. (1)	(s.e.) (2)	# $\hat{q} < .1$ (3)
Panel A: Location choices						
Moves from baseline	0.236	(0.042)	0	0.06	(0.086)	0
Moves towards HP	0.145	(0.033)	0	0.04	(0.054)	0
Lives with candidate	0.07	(0.022)	0	0.052	(0.03)	0
Lives nearby candidate	0.117	(0.03)	0	0.174	(0.069)	0
Panel B: Labor supply						
Any employment	0.516	(0.041)	0	0.612	(0.082)	0
Any informal employment	0.245	(0.039)	0	0.163	(0.077)	0
Any formal employment	0.271	(0.039)	0	0.428	(0.07)	0
Kept same job (uncond.)	0.034	(0.017)	0	0.105	(0.036)	0
Has new job (uncond.)	0.23	(0.034)	0	0.349	(0.071)	0
PBF Beneficiary	0.414	(0.052)	0	0.166	(0.103)	0

Note: This table presents complier average treated and untreated average location choices (panel A) and labor supply (panel B). Appendix C.3 for estimation details. Standard errors are clustered at the candidate level. The Column (3) in set reports the number of lotteries with q-values (Storey, 2002) lower than 10% for testing the null that estimates lie within 0 and 1 for each lottery.

Table A.5: *Average potential outcomes for relatives nearby the candidate, by location choice type*

	$E[Y(1,0)\|$never-mover$]$			$E[Y(1,R)\|$mover$]$		
Outcome	Est. (1)	(s.e.) (2)	# $\hat{q} < .1$ (3)	Est. (1)	(s.e.) (2)	# $\hat{q} < .1$ (3)
Any employment	0.597	(0.056)	0	0.573	(0.089)	0
Any informal employment	0.406	(0.057)	0	0.192	(0.08)	0
Any formal employment	0.236	(0.043)	0	0.381	(0.087)	0
Has new formal job (uncond.)	0.19	(0.038)	0	0.334	(0.092)	0
Receives PBF	0.485	(0.065)	0	0.67	(0.091)	0

Note: This table presents estimated average potential outcomes for relatives living nearby candidates, stratified by relatives' location choice types. '$E[Y(1,0)|$stayer$]$' shows the average outcome for relatives who never move when the candidate wins the lottery. '$E[Y(1,R)|$mover$]$' shows the average outcome for relatives who move when the candidate wins the lottery. See Appendix C.3 for estimation details. Standard errors clustered at the candidate level are reported in parentheses. The column "# $\hat{q} < .1$" indicates the number of estimates with q-values less than 0.1 after multiple testing correction.

Table A.6: *Share of relatives' types based on their location choices*

Group		Est.	(s.e.)
Relative of complier	$(D(1) > D(0))$	0.150	(0.001)
Always-mover	$(R(0) = R(1) = 1)$	0.043	(0.071)
Follower	$(R(1) > R(0))$	0.147	(0.082)
Never-mover	$(R(0) = R(1) = 0)$	0.810	(0.042)

Note: This table presents the estimated shares of relatives by their location choice types in the extended instrumental variable framework. 'Relative of complier' shows the proportion of relatives whose candidates are compliers (accept housing only when they win the lottery). Among relatives of compliers, the types are defined as: 'Always-mover' - relatives who move closer to housing projects regardless of whether their candidate wins; 'Follower' - relatives who move closer only when their candidate wins and moves; 'Stayer' - relatives who never move closer regardless of lottery outcome. See Appendix C.3 for details. Standard errors clustered at the candidate level are reported in parentheses.

B Data and Sample Construction

B.1 Data Source details

[Work in progress]

B.2 Family Links

This Appendix Section describes the procedure we follow to build family links using Cadastro Unico (CadÚnico) data. We start from a dataset containing information for the universe of individuals in CadÚnico. Our approach combines multiple identification strategies to construct family networks spanning the universe of CadÚnico beneficiaries across Brazil.

B.2.1 Data Sources and Key Variables

CadÚnico contains detailed individual records with the following key variables used in our family linkage. Individual identifiers include CPF (tax identification number) and NIS (social identification number). Personal information encompasses full name, age, and gender. Parental names consist of self-reported full names of father and mother. Household identifiers use a family unit code, linking individuals within households. Temporal information includes survey dates and data update timestamps.

Household roster. CadÚnico records each individual's relationship to the household head using standardized codes. Primary roles include "Person responsible for the family unit - RF" (household head) and "Spouse or partner". Parent-child relationships encompass "Father or mother", "Son or daughter", and "Stepson or stepdaughter". Extended family includes "Brother or sister" and "Grandson or granddaughter or great-grandchild". In-law relationships cover "Father-in-law or mother-in-law" and "Son-in-law or daughter-in-law". Other relationships use "Other relative" for relationships not captured by specific codes.

B.2.2 Data Preparation and Cleaning

String Standardization. All name variables undergo extensive cleaning to ensure consistent matching. We remove leading and trailing whitespace and convert to lowercase, remove numbers and punctuation marks using pattern matching, standardize consecutive whitespace to single spaces, and apply string squishing to eliminate formatting inconsistencies.

Sample Restrictions. We apply several filters to focus on useful observations. We drop records missing both parental names (father and mother), require at least one valid identifier (CPF or NIS), and remove non-informative parental names such as "NÃO CONSTA", "NÃO DECLARADO", and "NÃO INFORMADO".

Temporal Harmonization. To handle multiple observations per individual over time, we extract the survey year from the timestamp data and identify the minimum of the reported year and the survey year. Then, for each pair we build, we build a 'pair year' as the maximum of both individuals' link years for relationship pairs. We use this later for two purposes: to deduplicate multiple matches, retaining the earliest observed relationship when duplicates exist, to keep relationships identified before the lottery of interest.

B.2.3 Overview of the Linkage Process

Our family linkage procedure entails five sequential steps, each building on relationships identified in previous steps. First, siblings identification via parental names links individuals who report identical father and mother names. Second, siblings identification via household relationships extracts sibling relationships from CadÚnico's household roster data. Third, parents identification identifies parent-child relationships using both direct household relationships and inferred relationships through siblings. Fourth, higher-order relationships construct grandparent, uncle/aunt, cousin, spouse, and in-law relationships by chaining previously identified links. Fifth, conflict resolution and final cleaning apply hierarchical prioritization to resolve cases where the same individual pair is identified through multiple methods.

Quality Control and Deduplication. Throughout the process, we apply consistent quality filters and deduplication procedures. Household-level restrictions ensure relationship validity by requiring specific household compositions such as exactly one head, at most two parents, and exactly one spouse. All relationships are made bidirectional. When duplicate pairs exist, we retain records with the earliest 'pair year' to maximize temporal coverage. Final deduplication handles ties by retaining the first occurrence of each unique relationship pair.

B.2.4 Pair Relationship Identification

We identify family relationships through multiple complementary methods, building from direct relationships to higher-order connections. Each relationship type uses specific identification strategies designed to maximize coverage while maintaining data quality.

B.2.4.1 Siblings via Parental Names

We identify potential sibling groups by clustering individuals who report identical combinations of cleaned father and mother names. To minimize false positives from common name combinations, we restrict to groups with 2-7 members. We apply temporal logic using the earliest observed year when multiple observations exist. From each sibling group, we generate all possible pairwise combinations. We then identify the 'pair year' by capturing the maximum of the two individuals' survey dates.

B.2.4.2 Siblings via Household Relationships

We extract sibling relationships from household composition data using three mechanisms.

First, head-sibling relationships link household heads ('Person responsible for the family unit - RF') with individuals coded as 'Brother or sister'. Second, children relationships identify multiple individuals coded as 'Son or daughter' within the same household-year and create pairwise combinations among them. Third, step-relationships link multiple individuals coded as 'Stepson or stepdaughter' within the same household-year. Each mechanism creates bidirectional relationships by generating symmetric pairs.

B.2.4.3 Parents Identification

Parent-child relationships are identified through eight distinct direct pathways and an indirect pathway.

First, household heads to co-resident parents link heads with individuals coded as 'Father or mother'. Second, household heads to children link heads with 'Son or daughter' members. Third, spouse to children link individuals coded as 'Spouse or partner' with 'Son or daughter' members. Fourth, spouse to step-children link spouses with 'Stepson or stepdaughter' members. Fifth, siblings' parents link 'Brother or sister' members with 'Father or mother' members within households. Sixth, Grandchildren via children link 'Grandson or granddaughter or great-grandchild' members with 'Son or daughter' members. Seventh, in-law relationships link individuals coded as 'Spouse or partner' with 'Father-in-law or mother-in-law' members. Eighth, grandchildren via in-laws link 'Grandson or granddaughter or great-grandchild' members with 'Son-in-law or daughter-in-law' members.

Beyond direct household relationships, we infer additional parent-child links using previously identified sibling relationships. We merge the parent dataset with the sibling dataset (which includes both siblings via names and household roster) on parental identifiers, creating new parent-child relationships where siblings of known children inherit the same parents.

B.2.4.4 Grandparents

Grandparent-grandchild relationships are constructed through three pathways.

First, direct household observation links household heads with 'Grandson or granddaughter or great-grandchild' members. Second, spouse extension links individuals coded as 'Spouse or partner' with grandchildren. Third, chain inference links parents of previously identified parents using the parent dataset. We merge the parent dataset with itself on parental identifiers to create grandparent-grandchild connections. As usual, the 'pair year' becomes the maximum of the two parent identification years.

B.2.4.5 Uncles/Aunts

Uncle/aunt relationships are identified as siblings of previously identified parents. We merge the parent dataset with the sibling dataset on parental identifiers, creating uncle/aunt-niece/nephew relationships. As usual, the 'pair year' captures the maximum of the parent identification year and sibling identification year.

B.2.4.6 Nieces/Nephews

Niece/nephew relationships are identified as the reciprocal of previously identified uncles/aunts. We merge the uncle/aunt dataset with the parent dataset on uncle/aunt identifiers, creating symmetric niece/nephew-uncle/aunt relationships. The 'pair year' captures the maximum of the uncle/aunt identification year and parent identification year.

B.2.4.7 Cousins

Cousin relationships are identified as children of previously identified uncles/aunts. We merge the uncle/aunt dataset with the parent dataset on uncle/aunt identifiers, creating cousin relationships. We create bidirectional cousin relationships by generating symmetric pairs.

B.2.4.8 In-Laws

Spouses. We first identify spouses to build in-law relationships Spouse relationships are identified through four mechanisms. First, direct observation links household heads with 'Spouse or partner' members. Second, parent pairs identify pairs of individuals both coded as 'Father or mother' within the same household-year. Third, in-law parent pairs identify pairs of individuals both coded as 'Father-in-law or mother-in-law' within the same household-year. Fourth, child-in-law pairs link individuals coded as 'Son or daughter' with 'Son-in-law or daughter-in-law' members. All spouse relationships are made bidirectional.

In-laws. In-law relationships are constructed through three pathways. First, direct household observation links household heads with individuals coded as 'Son-in-law or daughter-in-law' or 'Father-in-law or mother-in-law' and creates bidirectional relationships between these categories. Second, spouse family extension merges spouse relationships with all previously identified family relationships (parents, siblings), where relatives of one spouse become in-laws of the other spouse, using consistent naming conventions. Third, parent in-law connections link parents of children with parents of children's spouses through a chain of relationships: parent to child to child's spouse to spouse's parent.

B.2.4.9 Other Relatives

Finally, we create relationships for individuals coded as 'Other relative' by linking them to various household members: heads, siblings, parents, and children. All other relative relationships are made bidirectional.

B.2.5 Descriptive Statistics by Relationship Type

The family relationship identification process generates substantial numbers of relationships across all categories. Table B.1 presents the counts of relationships identified through each pathway before deduplication and conflict resolution.

As shown in Table B.1, cousin relationships represent the largest category, comprising over 862 million pairs. This substantial number reflects the combinatorial expansion that occurs when linking children of siblings across multiple family lines through our chain inference methodology.

Sibling relationships constitute the second largest category with over 206 million pairs, identified through both parental name matching and household structure analysis. The combination of name-based and household-based identification methods captures sibling relationships across different data collection periods and household arrangements.

Parent-child relationships, when combined as directional pairs, total approximately 291 million relationships. Our eight identification pathways generate substantial coverage of parent-child links, with the majority identified through direct household relationship codes and supplemented by inference through sibling relationships.

Uncle/aunt and nieces/nephews relationships each total approximately 144 million pairs, demonstrating the importance of chain inference methods for capturing extended family networks. These relationships link parents with siblings of their own parents and their children, creating crucial connections across generational lines.

Grandparent relationships total approximately 89 million pairs (combined grandparents and grandchildren), with the majority identified through indirect chain inference rather than direct household observation. This distribution demonstrates the effectiveness of our relationship chaining approach for capturing multi-generational family networks beyond what is directly observable in household composition data.

Spouse relationships number over 32 million pairs, with household head-to-spouse links comprising the vast majority. In-law relationships contribute additional marriage-based family extensions, while other relatives provide a residual category for relationships not captured by specific codes.

The raw identification process generates approximately 1.92 billion relationship pairs before deduplication, highlighting both the comprehensiveness of our methodology and the necessity of systematic conflict resolution procedures.

B.2.6 Conflict Resolution and Network Deduplication

The comprehensive identification process inevitably generates multiple relationship classifications for the same individual pairs. Our conflict resolution procedure applies a systematic hierarchical prioritization followed by temporal deduplication to produce a final unified family network.

Hierarchical conflict resolution. When the same individual pair appears with multiple relationship types, we apply a ten-level hierarchy that prioritizes closer family relationships and more reliable identification methods. The hierarchy proceeds as follows: children (highest priority), parents, siblings, grandchildren, grandparents, uncles/aunts, nieces/nephews, cousins, spouses, in-laws, and other relatives (lowest priority).

At each hierarchy level, we identify pairs with multiple relationship classifications and retain only the higher-priority relationship type. This process systematically reduces the dataset size while preserving the most appropriate relationship classification for each pair.

The conflict resolution demonstrates substantial overlap between identification methods:

Table B.1: *Family Relationships Identified Before Deduplication*

Relationship Type	Pairs Identified	Primary Method
Cousins	865,979,741	Chain inference
Siblings	206,558,876	Names & household
Parents	145,623,456	Household & inference
Children	145,623,456	Reverse of parents
Uncle/Aunt	145,558,205	Chain inference
Nieces/Nephews	145,558,205	Chain inference
Grandparents	45,013,100	Chain inference
Grandchildren	45,013,100	Reverse of grandparents
Spouses	32,675,856	Household structure
In-laws	37,168,170	Spouse extension
Other relatives	14,522,612	Household codes
Total	**≈1.92 billion**	

Notes: Counts represent relationship pairs before conflict resolution and deduplication. Parent and child relationships are directional pairs; grandparent and grandchild relationships are similarly directional. Primary method indicates the most common identification pathway: "Names" refers to parental name matching, "Household" refers to household relationship codes, "Chain inference" refers to relationships constructed by linking previously identified relationships, and "Spouse extension" refers to relationships created through marriage connections.

- **Children prioritization** reduces the dataset by approximately 2.27 million relationships (0.12%)
- **Parent prioritization** removes an additional 2.30 million relationships (0.12%)
- **Sibling prioritization** removes 4.03 million relationships (0.21%)
- **Grandchildren prioritization** removes 206,344 relationships (0.01%)
- **Grandparents prioritization** removes 189,911 relationships (0.01%)
- **Uncles/aunts prioritization** removes 548,520 relationships (0.03%)
- **Nieces/nephews prioritization** removes 400,989 relationships (0.02%)

Higher-order relationships show more substantial consolidation during conflict resolution. Cousin relationships, being the most numerous, experience significant reduction during the hierarchy application, as many cousin pairs are reclassified to closer relationship types when conflicts arise.

Final tie-breaking and temporal deduplication For relationship pairs that remain conflicted after hierarchical prioritization, we apply temporal deduplication rules. We retain

the earliest observed relationship year ('pair year') for each unique pair, maximizing temporal coverage while ensuring that our analysis captures relationships identified before policy interventions of interest.

The final step applies within-pair deduplication by retaining only the first occurrence when multiple identical relationships exist, ensuring each unique family relationship appears exactly once in the final network.

Final network characteristics. The unified family network contains approximately 1.91 billion unique relationships spanning the universe of CadÚnico beneficiaries. Table B.2 presents the final composition of the family network after conflict resolution and deduplication.

As shown in Table B.2, cousin relationships comprise the largest category (45.03% of all relationships), followed by siblings (10.79%) and parent-child relationships (20.16% combined). Extended family relationships (uncles/aunts and nieces/nephews combined with grandparents/grandchildren) account for 17.68% of connections, while marriage-based relationships (spouses and in-laws) represent 3.63% of the final network.

Network density varies substantially across relationship types, with immediate family relationships (parents, children, siblings, spouses) showing higher connection rates per individual compared to extended family relationships. The average individual maintains connections across multiple relationship types in the final network, though this distribution exhibits substantial right-skew with some individuals connected to over 100 family members.

The temporal span of relationship identification ranges from 2011 to 2020, with relationship discovery concentrated in years corresponding to major CadÚnico registration campaigns. This temporal variation reflects both genuine relationship formation and administrative data collection patterns rather than systematic bias in relationship identification methods.

B.3 Geocoded Address and Distance Histories

B.3.1 Overview and Data Sources

This section describes the construction of distance panels that track proximity between lottery candidates, their relatives, and housing projects over time from 2010 to 2020. The ultimate goal is to measure how family proximity and housing project distances evolve, enabling our analysis of family coordination in location choices.

Our construction process uses candidate and relative rosters, raw address data from Brazil's CadUnico for these individuals, and housing project locations. We geocoded all unique addresses using the ArcGIS Online API, generating coordinates and quality scores that allow us to calculate precise distances between family members and to housing projects.

B.3.2 Temporal Construction and Panel Filling

Leveraging Survey Year Information: A major challenge in studying residential mobility among low-income families is that administrative surveys like CadUnico occur irregularly,

Table B.2: *Final Family Network Composition After Deduplication*

Relationship Type	Final Count	Percentage
Cousins	861,617,987	45.03%
Siblings	206,374,478	10.79%
Children	192,818,211	10.08%
Parents	192,808,657	10.08%
Uncles/Aunts	143,775,105	7.51%
Nieces/Nephews	143,587,198	7.50%
Grandchildren	44,680,894	2.34%
Grandparents	44,670,832	2.33%
In-laws	36,918,054	1.93%
Spouses	32,592,222	1.70%
Other relatives	13,526,684	0.71%
Total	**1,913,370,322**	**100.00%**

Notes: Final counts after hierarchical conflict resolution and temporal deduplication. Parent-child and grandparent-grandchild relationships are presented as separate directional categories. The network spans the universe of CadÚnico beneficiaries across Brazil, representing one of the largest family network datasets ever constructed.

often every two years or more, creating substantial gaps in address reporting. Many families may not update their information promptly, resulting in incomplete address data for certain years.

However, CadUnico provides a crucial piece of additional information: the "survey year" indicating when each family last reported changes to their circumstances, including residential moves. This creates an opportunity to infer address stability. For instance, if a family reports living at a particular address in 2016 but their survey year indicates their last change was in 2013, we can reasonably conclude they have been residing at that address since 2013, filling the apparent gap in our data.

Creating Complete Panels: To measure distances consistently across our study period, we need complete address information for each year from 2010 to 2020. We accomplish this through temporal interpolation that makes reasonable assumptions about residential stability. When we observe someone at an address in a particular year, we assume they remain there in subsequent years until we observe them elsewhere. This forward projection reflects the reality that most people, especially low-income families, do not move frequently.

Similarly, when we first observe someone at an address, we extend this residence backward to either 2010 or their survey year, whichever is more recent. This backward extension captures the insight that many address "first observations" in our data reflect data collection timing rather than actual move dates. For example, if an individual has observed addresses in 2013, 2016, and 2020, we fill 2010-2012 with the 2013 address (backward filling), 2014-2015 with the 2013 address (forward filling), and 2017-2019 with the 2016 address (forward filling). The combination of forward and backward filling creates complete residential histories that enable consistent distance calculations across all individuals and years.

Handling Data Quality: When individuals appear at multiple addresses in the same year, we prioritize records from surveys conducted closer to that year, then select addresses with higher geocoding accuracy scores. This ensures our distance calculations rely on the most accurate location data available.

B.3.3 Distance Calculations and Final Panel Construction

Dual Panel Structure: Our process generates two complementary distance panels. The candidate panel tracks each lottery participant's distance to housing projects and proximity to their family network over time. The relative panel captures the reverse perspective, measuring how each relative's distance to the candidate and to housing projects evolves.

We calculate all distances using the Haversine formula, which provides accurate great-circle distances between geocoded coordinates. The final panels ensure that every candidate and every candidate-relative pair has complete distance information for each year from 2010 to 2020, including to their family proximity and, for each lottery they participate in, the housing projects relevant to the lottery.

B.4 Analysis Sample Construction

[Work in progress]

B.5 Final Variable Construction

[Work in progress]

C Estimation Appendix

This appendix provides comprehensive details on the estimation strategies used to recover estimates of the different parameters presented in the main body of the paper.

Notation Throughout this appendix, we use the following notation conventions. First, we remove subscripts to simplify notation. Second, we use $Z \in \{0, 1\}$ to denote the lottery win indicator, D denotes the candidate's location choice, R denotes the relative's location choice, Y denotes the labor supply outcome of interest for relatives, and X denotes an observable pre-determined characteristic. Finally, for notational convenience, define, for any variable W, $\Delta_W := \mathbb{E}[W|Z = 1] - \mathbb{E}[W|Z = 0]$.

C.1 Balance tests and Reduced form estimates

C.1.1 Target Parameters

We are interested in estimating the mean difference between winners and losers, Δ_W, adjusted for the heterogeneity in lotteries examined and the assignment mechanism used in these lotteries. For balance tests, the variables of interest are $W = X$. For reduced form estimates (ITT) on location choices, we are interested in $W = D$ for candidates and $W = R$ for their relatives.

C.1.2 Estimation approach

We estimate Δ_W via OLS regression that adjusts for the stratified randomization design and accounts for individuals participating in multiple lotteries. Specifically, we estimate β in:

$$W = \beta Z + \mathbf{X}'\theta + \varepsilon,$$

where $\mathbf{X}$ denotes the vector of controls. For candidates, $\mathbf{X}$ includes only lottery×age-bin fixed effects. For relatives, $\mathbf{X}$ includes these fixed effects plus a variable capturing the number of candidates related to each relative (to account for non-random exposure to the treatment through the endogenous network structure). Standard errors are clustered at the candidate level to account for the same individual participating in multiple lotteries.

C.2 Lee (2005) bounds

C.2.1 Target Parameters

Since lottery winners are more likely to respond to post-lottery surveys, we face potential selection bias when studying the reduced form effects on location choices. To account for this, we compute Lee (2005) bounds to address this non-random attrition, targeting the average effects among 'always-respondents'. To be precise, let $S(z)$ denote the potential survey response indicator under lottery outcome $z \in \{0, 1\}$, so that $S = S(Z)$ denotes observed

survey response. Then, we target the treatment effect on location choices among 'always-respondents', $\mathbb{E}[D(1) - D(0)|S(1) = S(0) = 1]$, for candidates and $\mathbb{E}[R(1) - R(0)|S(1) = S(0) = 1]$ for relatives.[15]

To account for the different lotteries, we implement generalized Lee bounds following Semenova (2020), which extends Lee (2005) to allow for covariate adjustment with x-specific monotonicity direction. In this case, $\mathbf{X}$ is a discrete variable for each lottery examined. Let $W \in \{D, R\}$ denote the outcome of interest and define, $s_0(x) := \mathbb{P}[S = 1|Z = 0, \mathbf{X} = x]$ (response rate for losers in lottery x), $s_1(x) := \mathbb{P}[S = 1|Z = 1, \mathbf{X} = x]$ (response rate for winners in lottery x), $p_0(x) := s_0(x)/s_1(x)$ (relative response rate) $Q^z(u, x)$ the conditional u-quantile of W in the $S = 1, Z = z, \mathbf{X} = x$ group. Semenova (2020) shows that for continuous outcomes, the lottery-specific bounds are given by:

$$\beta_U(x) = \begin{cases} \mathbb{E}[W|W \geq Q^1(1 - p_0(x), x), Z = 1, S = 1, \mathbf{X} = x] - \mathbb{E}^0 & \text{if } s_1(x) > s_0(x) \\ \mathbb{E}^1 - \mathbb{E}^0 & \text{if } s_1(x) = s_0(x) \\ \mathbb{E}^1 - \mathbb{E}[W|W \leq Q^0(1/p_0(x), x), Z = 0, S = 1, \mathbf{X} = x] & \text{if } s_1(x) < s_0(x) \end{cases} \tag{C.1}$$

$$\beta_L(x) = \begin{cases} \mathbb{E}[W|W \leq Q^1(p_0(x), x), Z = 1, S = 1, \mathbf{X} = x] - \mathbb{E}^0 & \text{if } s_1(x) > s_0(x) \\ \mathbb{E}^1 - \mathbb{E}^0 & \text{if } s_1(x) = s_0(x) \\ \mathbb{E}^1 - \mathbb{E}[W|W \geq Q^0(1 - 1/p_0(x), x), Z = 0, S = 1, \mathbf{X} = x] & \text{if } s_1(x) < s_0(x) \end{cases} \tag{C.2}$$

where $\mathbb{E}^1 := \mathbb{E}[W|S = 1, Z = 1, \mathbf{X} = x]$ and $\mathbb{E}^0 := \mathbb{E}[W|S = 1, Z = 0, \mathbf{X} = x]$.

In the binary case, we modify the bounds to respect the $[0, 1]$ constraint on probabilities. More precisely

$$\beta_U(x) = \begin{cases} \min\left\{1, \frac{1}{p_0(x)}\mathbb{E}^1\right\} - \mathbb{E}^0 & \text{if } s_1(x) > s_0(x) \\ \mathbb{E}^1 - \mathbb{E}^0 & \text{if } s_1(x) = s_0(x) \\ \mathbb{E}^1 - \max\{0, p_0(x)\mathbb{E}^0 + 1 - p_0(x)\} & \text{if } s_1(x) < s_0(x) \end{cases} \tag{C.3}$$

$$\beta_L(x) = \begin{cases} \max\left\{0, \frac{1}{p_0(x)}\mathbb{E}^1 - \frac{1-p_0(x)}{p_0(x)}\right\} - \mathbb{E}^0 & \text{if } s_1(x) > s_0(x) \\ \mathbb{E}^1 - \mathbb{E}^0 & \text{if } s_1(x) = s_0(x) \\ \mathbb{E}^1 - \min\{1, p_0(x)\mathbb{E}^0\} & \text{if } s_1(x) < s_0(x) \end{cases} \tag{C.4}$$

Finally, the unconditional (on $\mathbf{X}$) bounds are constructed by integrating over the space of $\mathbf{X}$ (conditional on the 'always-respondents' group), this is

$$\mathbb{E}[W(1) - W(0)|S(1) = S(0) = 1] \in [\beta_L, \beta_U]$$

[15] To be precise, the exact definition of S for relatives depends on the outcome examined. For example, if R captures whether the relative moved, S is defined as observing a post-lottery address for each relative. However, if R captures whether the relative and candidate are closer to each other, S is defined as observing a post-lottery address for both the candidate and the relative in each pair.

where

$$\beta_L = \int_{\mathcal{X}} \beta_L(x) f_{\mathbf{X}}(x|S(1) = S(0) = 1)dx \quad \text{and} \quad \beta_U = \int_{\mathcal{X}} \beta_U(x) f_{\mathbf{X}}(x|S(1) = S(0) = 1)dx. \tag{C.5}$$

and $\mathcal{X}$ denotes the support of $\mathbf{X}$.

C.2.2 Estimation approach

We implement the bounds using sample analogs for all relevant components, following Semenova (2020) generalized Lee bounds approach. This method extends the original Lee (2005) bounds by estimating individual-specific selection probabilities and trimming ratios using covariate information, then reweighting observations according to their always-takers probability to construct the final bounds. For a given outcome W, we proceed in three steps.

Step 1: Individual-specific trimming ratio computation We estimate selection probabilities conditional on covariates rather than using overall sample rates. For each individual i with covariates $\mathbf{X}_i$, we estimate selection probabilities using probit regression:

$$\Pr(S = 1|\mathbf{X}, Z) = \Phi(\alpha + \beta Z + \mathbf{X}'\gamma).$$

We then predict individual-specific response rates

$$\hat{s}_1(\mathbf{X}_i) = \Phi(\hat{\alpha} + \hat{\beta} + \mathbf{X}_i'\hat{\gamma}),$$
$$\hat{s}_0(\mathbf{X}_i) = \Phi(\hat{\alpha} + \mathbf{X}_i'\hat{\gamma}),$$

and use these to compute the individual relative response rates

$$\hat{p}_0(\mathbf{X}_i) = \hat{s}_0(\mathbf{X}_i)/\hat{s}_1(\mathbf{X}_i)$$

Step 2: Individual-specific bound computation based on outcome type. Using each individual's trimming ratio $\hat{p}_0(\mathbf{X}_i)$, we apply the bounds formulas to construct individual-level bounds $\hat{\beta}_L(\mathbf{X}_i)$ and $\hat{\beta}_U(\mathbf{X}_i)$.
Binary outcomes ($W \in \{0, 1\}$): We apply the analytical bounds directly at the individual level following Equations C.4 and C.3. We estimate individual-specific conditional outcome means using probit regression:

$$\Pr(W = 1|\mathbf{X}, Z) = \Phi(\delta + \theta Z + \mathbf{X}'\eta)$$

Then we apply the bounds for $\hat{s}_1(\mathbf{X}_i) > \hat{s}_0(\mathbf{X}_i)$:

$$\hat{\beta}_U(\mathbf{X}_i) = \min\left\{1, \frac{\hat{\mathbb{E}}^1}{\hat{p}_0(\mathbf{X}_i)}\right\} - \hat{\mathbb{E}}^0 \tag{C.6}$$

$$\hat{\beta}_L(\mathbf{X}_i) = \max\left\{0, \frac{\hat{\mathbb{E}}^1 - (1 - \hat{p}_0(\mathbf{X}_i))}{\hat{p}_0(\mathbf{X}_i)}\right\} - \hat{\mathbb{E}}^0 \tag{C.7}$$

Continuous outcomes:We estimate conditional quantile functions and compute individual-specific trimmed means following Equations C.2 and C.1. For $\hat{s}_1(\mathbf{X}_i) > \hat{s}_0(\mathbf{X}_i)$: We estimate quantile functions $\hat{Q}^z(\tau|\mathbf{X})$ using quantile regression on a grid $\tau \in \{0.05, 0.10, \ldots, 0.95\}$. For each individual i, we interpolate to find: $\hat{Q}^1(\hat{p}_0(\mathbf{X}_i)|\mathbf{X}_i)$ and $\hat{Q}^1(1 - \hat{p}_0(\mathbf{X}_i)|\mathbf{X}_i)$. We then compute individual-specific bounds based on trimmed means:

$$\hat{\beta}_L(\mathbf{X}_i) = \frac{\sum_j \mathbf{1}[Z_j = 1, S_j = 1, W_j \leq \hat{Q}^1(\hat{p}_0(\mathbf{X}_i)|\mathbf{X}_i)]W_j}{\sum_j \mathbf{1}[Z_j = 1, S_j = 1, W_j \leq \hat{Q}^1(\hat{p}_0(\mathbf{X}_i)|\mathbf{X}_i)]} - \hat{\mathbb{E}}^0(\mathbf{X}_i)$$

and $\hat{\beta}_U(\mathbf{X}_i)$ analogously, where $\hat{\mathbb{E}}^0(\mathbf{X}_i)$ is computed using the same quantile regression approach.

The $\hat{s}_1(\mathbf{X}_i) < \hat{s}_0(\mathbf{X}_i)$ cases follow analogously with trimming applied to the control group.

Step 3: Always-takers reweighting and aggregation A key feature of Semenova's approach is the reweighting scheme that targets the always-respondents population. For each individual i, we compute the weight representing the individual's contribution to the always-respondent population:

$$w_i = \min\{\hat{s}_0(\mathbf{X}_i), \hat{s}_1(\mathbf{X}_i)\}$$

Finally, we obtain unconditional bounds as weighted averages of the individual-specific bounds:

$$\hat{\beta}_L = \frac{\sum_i w_i \hat{\beta}_L(\mathbf{X}_i)}{\sum_i w_i} \qquad \hat{\beta}_U = \frac{\sum_i w_i \hat{\beta}_U(\mathbf{X}_i)}{\sum_i w_i}$$

This reweighting approach naturally implements the integration in equation (C.5) by weighting each observation according to its always-takers probability, effectively integrating over the covariate distribution conditional on the always-respondents population (Semenova, 2020).

C.3 TSLS estimates

C.3.1 Target Parameters

We estimate 5 types of parameters via TSLS regressions. All of these parameters have a common feature: they reflect a Wald estimand – i.e., the ratio between two Δ_W's. In what

follows, D is a binary variable indicating a candidate signing and moving to the housing project, so that under Assumptions A1-A3 we define

$$T := \begin{cases} AT & \text{if } D(1) = D(0) = 1 \\ C & \text{if } D(1) > D(0) \\ NT & \text{if } D(1) = D(0) = 0 \end{cases}$$

Complier characterization. We characterize the average observable features of compliers. These are captured by

$$E[X|T = C] = \frac{\mathbb{E}[XD|Z = 1] - \mathbb{E}[XD|Z = 0]}{\mathbb{P}[D = 1|Z = 1] - \mathbb{P}[D = 1|Z = 0]} = \frac{\Delta_{XD}}{\Delta_D}$$

Effects of candidates' location on relatives' outcomes. We are interested in the causal effect of candidates' location choices on relatives' outcomes among compliers (relatives of candidates who move only when winning the lottery). For $W \in \{R, Y\}$, we are interested in

$$E[W(1) - W(0)|T = C] = \frac{\mathbb{E}[W|Z = 1] - \mathbb{E}[W|Z = 0]}{\mathbb{P}[D = 1|Z = 1] - \mathbb{P}[D = 1|Z = 0]} = \frac{\Delta_W}{\Delta_D}$$

Average potential outcomes among compliers. For the LATE analysis and Kitagawa (2015) tests, we are also interested in potential outcome moments among compliers:

$$\mathbb{E}[W(d)|T = C] = \frac{\mathbb{E}[W\mathbb{1}[D = d]|Z = z] - \mathbb{E}[W\mathbb{1}[D = d]|Z = z]}{\mathbb{P}[D = d|Z = 1] - \mathbb{P}[D = d|Z = 0]} = \frac{\Delta_{W\mathbb{1}[D=d]}}{\Delta_{\mathbb{1}[D=d]}}$$

Characterization of relatives' types. We identify the shares of different types of relatives based on their location responses. Here, we let R denote an indicator for the relative moving closer to the housing project. Under Assumptions A1-A3 and Y1-Y3, the shares in (1) are point-identified (among compliers):

$$\begin{aligned} \mathbb{P}[R(1) = R(0) = 0|T = C] &= \frac{\Delta_{D(1-R)}}{\Delta_D} \\ \mathbb{P}[R(1) = R(0) = 1|T = C] &= \frac{\Delta_{(D-1)R}}{\Delta_D} \\ \mathbb{P}[R(1) > R(0)|T = C] &= \frac{\Delta_R}{\Delta_D} \\ \mathbb{P}[R(1) = 1|T = C] &= \frac{\Delta_{DR}}{\Delta_D} \end{aligned}$$

Furthermore, we characterize average observables among some of these groups, which are identified via

$$\mathbb{E}[X|R(1) = R(0) = 0, T = C] = \frac{\Delta_{XD(1-R)}}{\Delta_{D(1-R)}}$$
$$\mathbb{E}[X|R(1) = 1, T = C] = \frac{\Delta_{XDR}}{\Delta_{DR}}$$

Average potential outcomes by relative type (among compliers). Under the extended framework, we are also interested in the average potential outcomes by relative type. These are useful to test Assumptions (Y2) and (Y3) using the same logic as Kitagawa (2015) tests. To be precise, for labor outcomes Y, we are interested in recovering

$$\mathbb{E}[Y(1,0)|R(1) = R(0) = 0, T = C] = \frac{\Delta_{YD(1-R)}}{\Delta_{D(1-R)}}$$
$$\mathbb{E}[Y(1,1)|R(1) = 1, T = C] = \frac{\Delta_{YDR}}{\Delta_{DR}}$$

C.3.2 Estimation approach

All target parameters can be expressed as Wald estimands of the form $\frac{\Delta_{W_{num}}}{\Delta_{W_{den}}}$, where W_{num} is the outcome variable and W_{den} is the endogenous variable. We estimate these using TSLS regressions of the form:

$$W_{num} = \beta W_{den} + \mathbf{X}'\lambda + \epsilon$$
$$W_{den} = \pi Z + \mathbf{X}'\omega + \nu$$

where W_{num} represents any outcome of interest (e.g., R, Y, XD, XDR, etc.), W_{den} represents the endogenous variable (e.g., D, DR, $D(1-R)$, etc.), Z is the lottery win indicator (instrument) and $\mathbf{X}$ is a set of controls. For analyses on candidates, $\mathbf{X}$ includes only lottery×age-bin fixed effects. For relatives, $\mathbf{X}$ includes these fixed effects plus a variable capturing the number of candidates related to each relative. Furthermore, when W_{num} is a function of labor supply outcomes Y, we include the outcome at baseline as a control to improve efficiency. Standard errors are clustered at the candidate level to account for the same individual participating in multiple lotteries. Under our assumptions, the TSLS estimate of β is an unbiased estimate of $\frac{\Delta_{W_{num}}}{\Delta_{W_{den}}}$.

Examples of specific estimations:

- Complier characteristics: $W_{num} = XD$, $W_{den} = D$,
- LATE on relatives' outcomes: $W_{num} = Y$ (or R), $W_{den} = D$,

- Relative type share: $W_{num} = R$, $W_{den} = D$ (for followers)
- Characteristics by type: $W_{num} = XDR$, $W_{den} = DR$ (for movers)

C.4 Bounds endpoints

C.4.1 Target parameters

Let D denote a binary variable indicating a candidate signing and moving to the housing project, and R denote an indicator for the relative moving closer to the housing project. Under Assumptions A1-A3 and Y1-Y3, Dutz et al. (2024) show that the sharp bounds on $LATE_1$ and $LATE_0$ are given by

$$LATE_1 \in [\tau_1^{LB}, \tau_1^{UB}] \quad \text{and} \quad LATE_0 \in [\tau_0^{LB}, \tau_0^{UB}]$$

where

$$\tau_1^{LB} = \frac{\Delta_{YR} - \min\{\Delta_R, \Delta_{Y(1-D)R}\}}{\Delta_{DR}} \tag{C.8}$$

$$\tau_1^{UB} = \frac{\Delta_{YR} - \max\{0, \Delta_{Y(1-D)R} - \Delta_{(1-D)R} + \Delta_R\}}{\Delta_{DR}} \tag{C.9}$$

and

$$\tau_0^{LB} = \frac{\Delta_{Y(1-R)} + \max\{0, \Delta_{Y(1-D)(R-1)} - \Delta_{(1-D)(R-1)} + \Delta_R\}}{\Delta_{D(1-R)}} \tag{C.10}$$

$$\tau_0^{UB} = \frac{\Delta_{Y(1-R)} + \min\{\Delta_R, \Delta_{Y(1-D)(R-1)}\}}{\Delta_{D(1-R)}} \tag{C.11}$$

C.4.2 Estimation approach

Our approach consists of estimating the 'pieces' (Δ_W) with a similar regression specification that accounts for lotteries and for baseline outcomes when relevant. For any variable W we estimate Δ_W from an OLS regression analogous to those in the reduced form estimates. Namely

$$W = \Delta_W Z + X'\lambda + \varepsilon$$

where as above, X always includes lottery×age-bin fixed effects and a variable capturing the number of candidates related to each relative. When W is a function of labor supply outcomes Y, we include the outcome at baseline as a control to improve efficiency.

We then use the estimated $\hat{\Delta}_W$ to build estimates of the endpoints of the bounds following (C.8), (C.9), (C.10), and (C.11).