Family Ties, Residential Location Choice, and Labor Supply

Hadar Avivi University College London
Santiago Lacouture University of Chicago
Felipe Lobel Duke University
Winnie van Dijk Yale University

NBER SI 2026

July 30, 2026

Motivation

- Low-income households often remain in areas with limited economic opportunity (e.g., Rosenzweig and Stark 1989, Bryan et al. 2014, Bazzi 2017, Bryan and Morten 2019, Munshi 2020, Lagakos et al. 2023)

- A leading explanation is the value of **proximity** to social networks

Motivation

- Low-income households often remain in areas with limited economic opportunity (e.g., Rosenzweig and Stark 1989, Bryan et al. 2014, Bazzi 2017, Bryan and Morten 2019, Munshi 2020, Lagakos et al. 2023)

- A leading explanation is the value of **proximity** to social networks
  - Family and social networks exhibit geographic clustering (e.g., Compton and Pollak 2014, Bailey et al. 2018; 2020, Green 2024, Koenen and Johnston 2025)
  - Migrants tend to settle near pre-existing social contacts (e.g., Altonji and Card 1991, Munshi 2003, Beaman 2012, Stuart and Taylor 2021, Sahai and Bailey 2022, Blumenstock et al. 2025)

Motivation

- Low-income households often remain in areas with limited economic opportunity (e.g., Rosenzweig and Stark 1989, Bryan et al. 2014, Bazzi 2017, Bryan and Morten 2019, Munshi 2020, Lagakos et al. 2023)

- A leading explanation is the value of **proximity** to social networks

- But these patterns admit competing interpretations
    - Members of the network may sort into the same neighborhoods because they share preferences over local wages, housing, schools, or amenities
    - Or because they start out there and face moving costs
    - Little evidence exists on whether network proximity *causally* affects location decisions and economic outcomes

Motivation

- Low-income households often remain in areas with limited economic opportunity (e.g., Rosenzweig and Stark 1989, Bryan et al. 2014, Bazzi 2017, Bryan and Morten 2019, Munshi 2020, Lagakos et al. 2023)
- A leading explanation is the value of **proximity** to social networks
- But these patterns admit competing interpretations
- If there is a value of proximity, understanding the underlying mechanisms is important
  - ▷ Members of the network can provide **production** value (e.g., childcare, job referrals, insurance)
  - ▷ They can also provide **amenity** value (e.g., spend time together)
  - ▷ This distinction matters: e.g.,
    1. Institutions often suggested for resolving mobility constraints, only works for production value (e.g., Fetter et al. 2024)
    2. To evaluate policies that relocate individuals, which could have previously unrecognized spillovers

Motivation

- Low-income households often remain in areas with limited economic opportunity (e.g., Rosenzweig and Stark 1989, Bryan et al. 2014, Bazzi 2017, Bryan and Morten 2019, Munshi 2020, Lagakos et al. 2023)
- A leading explanation is the value of **proximity** to social networks
- But these patterns admit competing interpretations
- If there is a value of proximity, understanding the underlying mechanisms is important
  - ▷ Members of the network can provide **production** value (e.g., childcare, job referrals, insurance)
  - ▷ They can also provide **amenity** value (e.g., spend time together)
- Establishing these causal effects and understanding mechanisms is challenging
  - ▷ Networks are rarely **observed**, let alone **linked** to location and employment
  - ▷ Location choices are **endogenous**, and moreover are **jointly determined**

Motivation

- Low-income households often remain in areas with limited economic opportunity (e.g., Rosenzweig and Stark 1989, Bryan et al. 2014, Bazzi 2017, Bryan and Morten 2019, Munshi 2020, Lagakos et al. 2023)
- A leading explanation is the value of **proximity** to social networks
- But these patterns admit competing interpretations
- If there is a value of proximity, understanding the underlying mechanisms is important
  - ▷ Members of the network can provide **production** value (e.g., childcare, job referrals, insurance)
  - ▷ They can also provide **amenity** value (e.g., spend time together)
- Establishing these causal effects and understanding mechanisms is challenging

  $\longrightarrow$ **This paper**: **housing lotteries** + linked **family network** and **employment registries** to study how family proximity shapes location choices and labor supply

What We Do

- Use housing lotteries in Brazil's *Minha Casa, Minha Vida* (MCMV) program as a source of exogenous moves for lottery candidates
  - ▷ Use lottery offer to instrument for candidate moves, in order to study impact on relatives

- Because relatives can respond by relocating, the lottery shifts two location choices at once $\longrightarrow$ develop an empirical framework to distinguish the **separation** channel (stayers) from the **joint move** channel (followers)

What We Do

- Use housing lotteries in Brazil's *Minha Casa, Minha Vida* (MCMV) program as a source of exogenous moves for lottery candidates
  - ▷ Use lottery offer to instrument for candidate moves, in order to study impact on relatives

- Because relatives can respond by relocating, the lottery shifts two location choices at once $\longrightarrow$ develop an empirical framework to distinguish the **separation** channel (stayers) from the **joint move** channel (followers)

- We then ask three questions:
  1: How does family proximity affect **location choice**?
  2: Does loss of proximity from family affect **labor supply**?
  3: Do people **sort** on labor market returns to proximity? (Roy 1951, Borjas 1987)
     - ▷ A test of whether the **productive** value of proximity drives location choices

Toy model

What We Find

1: Study how spatial separation from family affects **location choices**

- ◇ ∼17% of nearby relatives of lottery compliers move closer to the housing project
- ◇ Relatives move to lower-quality nbhds

⟶ Evidence of preference for living near family members

What We Find

1: Study how spatial separation from family affects **location choices**

⟶ Evidence of preference for living near family members

2: Ask whether separation from family affects **labor supply**

- ◇ Candidate move reduces formal employment for relatives who stay behind
- ◇ Switch toward informal work emerges as an important margin of adjustment (Berniell et al. 2021; 2023)
- ◇ Except in households with young children, where employment falls on both margins

⟶ Labor market decisions respond to family proximity

⟶ MCMV also generates labor-market spillovers through separation, not just relocation

⟶ Consistent with lost childcare

What We Find

1: Study how spatial separation from family affects **location choices**

$\longrightarrow$ Evidence of preference for living near family members

2: Ask whether separation from family affects **labor supply**

$\longrightarrow$ Labor market decisions respond to family proximity

$\longrightarrow$ MCMV also generates labor-market spillovers through separation, not just relocation

$\longrightarrow$ Consistent with lost childcare

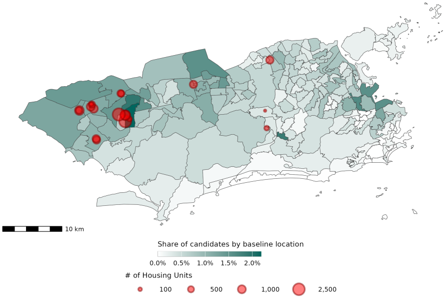
3: Test for **sorting on labor-market gains** from proximity (Borjas 1987)

◇ Evidence for Roy selection on the production value of proximity

$\longrightarrow$ Policy that substitute family support (public childcare, social insurance) could reduce mobility frictions

Minha Casa, Minha Vida (MCMV)

- One of the largest subsidized housing programs in the world: ∼5.5 million homes (cost: $92.8 bn)
- **Large subsidies**: for lowest-income tier, 90% subsidy + 10% subsidized mortgage   Details
- Allocated by **lottery**   Details
- **Project nbhds** are lowest tercile for schooling, poverty, housing quality (Belchior et al. 2023)
- Previous lit: effects of MCMV on candidates' employment, housing quality, health, and kids' education (da Mata and Mation 2018, Machado and Rachter 2020a, Belchior et al. 2023, da Mata and Mation 2025, Leite Mariante and Ribeiro 2025)



Maps for other cities

Minha Casa, Minha Vida (MCMV)

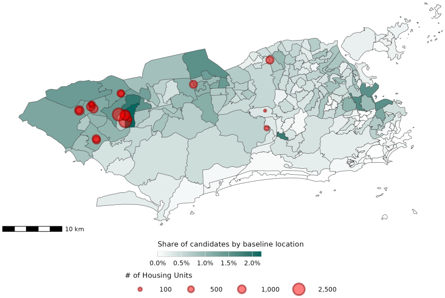
- One of the largest subsidized housing programs in the world: ∼5.5 million homes (cost: $92.8 bn)

- **Large subsidies**: for lowest-income tier, 90% subsidy + 10% subsidized mortgage Details

- Allocated by **lottery** Details

- **Project nbhds** are lowest tercile for schooling, poverty, housing quality (Belchior et al. 2023)

- Previous lit: effects of MCMV on candidates' employment, housing quality, health, and kids' education (da Mata and Mation 2018, Machado and Rachter 2020a, Belchior et al. 2023, da Mata and Mation 2025, Leite Mariante and Ribeiro 2025)



Maps for other cities

⟶ We focus on relatives outside the household

⟶ We collect data for 28 lotteries (2013-16; Rio de Janeiro, Guarulhos, Sorocaba, Sao Jose do Rio Preto)

Data

1. **Municipality Housing Departments**
    - Names and identifiers (CPF) of lottery winners and losers
    - Locations of housing projects
2. **Caixa Bank Contract Data**
    - Take-up among winners
    - Date of signature and subsidy amounts
3. **RAIS Employer-Employee Matched Data** (2008-2020)
    - Formal employment status, job spells, earnings, sectors, occupation

Outcome construction

Data

1. **Municipality Housing Departments**

2. **Caixa Bank Contract Data**

3. **RAIS Employer-Employee Matched Data** (2008-2020)

4. **Cadastro Unico (CU) Micro Data** (2012-2020)

Details: CU

   - "*Census of the poor*" in Brazil
     - Unusually broad coverage because it serves as the gateway for many social programs
     - Covers 40 percent of the national population, i.e., 80 million individuals per year
   - Addresses, sociodemographics (age, gender, race, education), household composition
   - CU is a panel but not updated annually. Pool outcomes over 4 yrs post lottery

Details

$\longrightarrow$ We use this dataset to construct **family network links**

Network construction

   - Use household composition and reported parents' full names
   - Build chains to link higher order relationships

Data

1. **Municipality Housing Departments**
2. **Caixa Bank Contract Data**
3. **RAIS Employer-Employee Matched Data** (2008-2020)
4. **Cadastro Unico (CU) Micro Data** (2012-2020)

Details: CU

Analysis Sample

Restrict to

1. candidates $\times$ lottery with
   (a) pre-lottery address in CU
   (b) at least one relative $>$16 years old with pre-lottery address in CU

2. relative $\times$ lottery with
   (a) $>$16 years old and with address in CU at pre-lottery
   (b) whose related candidate has pre-lottery address in CU

Data

1. **Municipality Housing Departments**
2. **Caixa Bank Contract Data**
3. **RAIS Employer-Employee Matched Data** (2008-2020)
4. **Cadastro Unico (CU) Micro Data** (2012-2020)

Details: CU

Analysis Sample

- **Final sample** $\approx$ 610K candidates $\times$ lottery; $\approx$ 1.4 million relative-candidate pairs $\times$ lottery

Data

1. **Municipality Housing Departments**
2. **Caixa Bank Contract Data**
3. **RAIS Employer-Employee Matched Data** (2008-2020)
4. **Cadastro Unico (CU) Micro Data** (2012-2020)

Details: CU

Analysis Sample

- **Final sample** $\approx$ 610K candidates $\times$ lottery; $\approx$ 1.4 million relative-candidate pairs $\times$ lottery

Descriptive Statistics

Candidates Relatives

- Balanced between lottery winners and losers
- Sample of candidates is predominantly women and participants in social programs
- 40% have children under 16
- Relatives: $^1/_3$ live with candidate, $^1/_3$ w/in 3km, $^1/_3$ further away. Focus on **nearby relatives**
- We observe an address for 92% in the 4-year post-period

Effect of Lottery Win on Candidates' Location Choices

	Loser mean	Estimate	(s.e.)
	(1)	(2)	(3)
Post-lottery CU respondents (up to 4 yrs)			
Moves from baseline	0.244	0.160	(0.009)
Moves more than 2 kms	0.132	0.170	(0.008)
Moves to HP	0.007	0.156	(0.006)
Signed and Moves to HP	0.000	0.160	(0.006)
Distance moved (uncond.)	1.839	1.527	(0.151)

Empirical specification | Effect on rent (Rio) | Effect on neighborhood quality | Lee bounds to account for attrition | Distance moved by city | Take-up by family proximity at baseline | Take-up by distance from HP

Effect of Lottery Win on Candidates’ Location Choices

	Loser mean	Estimate	(s.e.)
	(1)	(2)	(3)
Post-lottery CU respondents (up to 4 yrs)			
Moves from baseline	0.244	0.160	(0.009)
Moves more than 2 kms	0.132	0.170	(0.008)
Moves to HP	0.007	0.156	(0.006)
Signed and Moves to HP	0.000	0.160	(0.006)
Distance moved (uncond.)	1.839	1.527	(0.151)

→ Winning increases the probability of moving to the HP by 16 pp

Empirical specification | Effect on rent (Rio) | Effect on neighborhood quality | Lee bounds to account for attrition | Distance moved by city

Take-up by family proximity at baseline | Take-up by distance from HP

Effect of Lottery Win on Candidates' Location Choices

	Loser mean	Estimate	(s.e.)
	(1)	(2)	(3)
Post-lottery CU respondents (up to 4 yrs)			
Moves from baseline	0.244	0.160	(0.009)
Moves more than 2 kms	0.132	0.170	(0.008)
Moves to HP	0.007	0.156	(0.006)
Signed and Moves to HP	0.000	0.160	(0.006)
Distance moved (uncond.)	1.839	1.527	(0.151)

$\longrightarrow$ Distance moved (scaling by compliance) $\frac{1.52}{0.16} = 9.54$ km

Empirical specification | Effect on rent (Rio) | Effect on neighborhood quality | Lee bounds to account for attrition | Distance moved by city

Take-up by family proximity at baseline | Take-up by distance from HP

Effect of Lottery Win on Relatives' Location Choices

A candidate winning the lottery induces relatives to move closer to the HP

Scaling by takeup; live together: $\frac{0.087}{0.16} = 0.543$, live nearby: $\frac{0.029}{0.16} = 0.18$

Placebo: No effect on non-relative neighbors

Variable	Live together		Live nearby		Live far	
	Av. Loser (1)	Est. (2)	Av. Loser (3)	Est. (4)	Av. Loser (5)	Est. (6)
Distance moved (uncond.)	1.252	0.273	1.189	0.539	2.651	1.045
		(0.346)		(0.352)		(0.561)
Distance to candidate	1.410	1.029	2.294	1.338	16.258	-0.278
		(0.507)		(0.492)		(1.778)
Moves closer to HP	0.069	0.087	0.080	0.029	0.096	0.018
		(0.009)		(0.012)		(0.010)
# of losers	261,925		430,423		342,250	
# of winners	3,722		1,723		2,534	

Note: Nearby = less than 3 km, Far = more than 3 km

⟶ We establish that lottery affects location choices (for both candidates and relatives)

⟶ We are interested in understanding how proximity to family can affect economic outcomes

Effect of Lottery Win on Relatives' Location Choices

A candidate winning the lottery induces relatives to move closer to the HP

Scaling by takeup; live together: $\frac{0.087}{0.16} = 0.543$, live nearby: $\frac{0.029}{0.16} = 0.18$

Placebo: No effect on non-relative neighbors

	Live together		Live nearby		Live far	
Variable	Av. Loser (1)	Est. (2)	Av. Loser (3)	Est. (4)	Av. Loser (5)	Est. (6)
Distance moved (uncond.)	1.252	0.273	1.189	0.539	2.651	1.045
		(0.346)		(0.352)		(0.561)
Distance to candidate	1.410	1.029	2.294	1.338	16.258	-0.278
		(0.507)		(0.492)		(1.778)
Moves closer to HP	0.069	0.087	0.080	0.029	0.096	0.018
		(0.009)		(0.012)		(0.010)
# of losers	261,925		430,423		342,250	
# of winners	3,722		1,723		2,534	

Note: Nearby = less than 3 km, Far = more than 3 km

⟶ We establish that lottery affects location choices (for both candidates and relatives)

⟶ We are interested in understanding how proximity to family can affect economic outcomes

Effect of Candidate Move on Location Choices of Relatives: IV Framework

We are interested in the effect of **family proximity** on individual location choices. The assumptions needed to estimate this parameter via 2SLS:

Effect of Candidate Move on Location Choices of Relatives: IV Framework

We are interested in the effect of **family proximity** on individual location choices. The assumptions needed to estimate this parameter via 2SLS:

IV setup:

- $J(i) = j$ indicates that i is a relative of lottery candidate j
- $Z_j = \mathbb{1}[\text{lottery candidate } j \text{ wins lottery}]$
- $D_j(z) = \mathbb{1}[\text{candidate } j \text{ moves to the HP under } z]$
- $R_i(z, d) = \mathbb{1}[\text{relative } i \text{ moves closer to } J(i)\text{'s HP under } (z, d)]$

We would like to use Z_j as an instrument for $D_j(Z_j)$, to study the effect of $D_j(Z_j)$ on $R_i(D_j)$

Effect of Candidate Move on Location Choices of Relatives: IV Framework

We are interested in the effect of **family proximity** on individual location choices. The assumptions needed to estimate this parameter via 2SLS:

IV setup:

- $J(i) = j$ indicates that i is a relative of lottery candidate j
- $Z_j = \mathbb{1}[\text{lottery candidate } j \text{ wins lottery}]$
- $D_j(z) = \mathbb{1}[\text{candidate } j \text{ moves to the HP under } z]$
- $R_i(z, d) = \mathbb{1}[\text{relative } i \text{ moves closer to } J(i)\text{'s HP under } (z, d)]$

We would like to use Z_j as an instrument for $D_j(Z_j)$, to study the effect of $D_j(Z_j)$ on $R_i(D_j)$

Checking the standard IV assumptions:

A1 - Random assignment. $D_{J(i)}(z), R_i(z, d) \perp\!\!\!\perp Z_{J(i)}$ $\longrightarrow$ recall balance ✓

A2 - Monotonicity candidate's location choice. $\Pr(D_j(1) \geq D_j(0)) = 1$ $\longrightarrow$ recall take-up ✓

A3 - Exclusion from potential relative location choice. $R_i(d, z) \equiv R_i(d) \quad \forall (d, z)$

Effect of Candidate Move on Location Choices of Relatives: IV Framework

We are interested in the effect of **family proximity** on individual location choices. The assumptions needed to estimate this parameter via 2SLS:

IV setup:

- $J(i) = j$ indicates that i is a relative of lottery candidate j
- $Z_j = \mathbb{1}[\text{lottery candidate } j \text{ wins lottery}]$
- $D_j(z) = \mathbb{1}[\text{candidate } j \text{ moves to the HP under } z]$
- $R_i(z, d) = \mathbb{1}[\text{relative } i \text{ moves closer to } J(i)\text{'s HP under } (z, d)]$

We would like to use Z_j as an instrument for $D_j(Z_j)$, to study the effect of $D_j(Z_j)$ on $R_i(D_j)$

Checking the standard IV assumptions:

A1 - Random assignment. $D_{J(i)}(z), R_i(z, d) \perp\!\!\!\perp Z_{J(i)}$ $\longrightarrow$ recall balance ✓

A2 - Monotonicity candidate's location choice. $\Pr(D_j(1) \geq D_j(0)) = 1$ $\longrightarrow$ recall take-up ✓

A3 - Exclusion from potential relative location choice. $R_i(d, z) \equiv R_i(d) \quad \forall (d, z)$

$\longrightarrow$ focus on relatives living nearby ✓

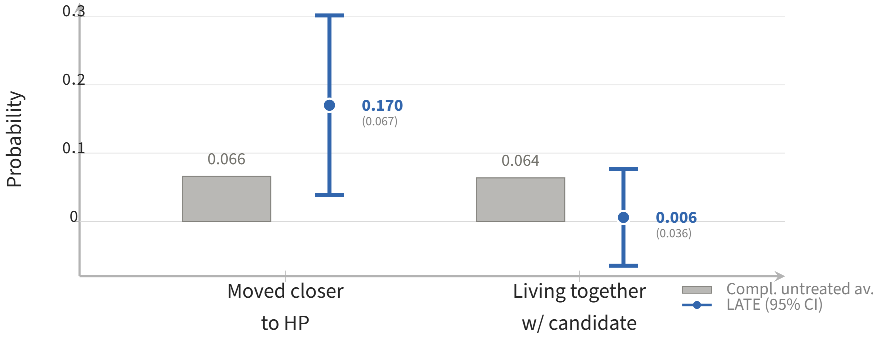
Kitagawa test | No increase in direct transfers from relatives

Effect of Candidate Move on Location Choices of Relatives: IV Estimates

We instrument for *candidate's move to the HP*, obtain effect on *their relatives' location choices*

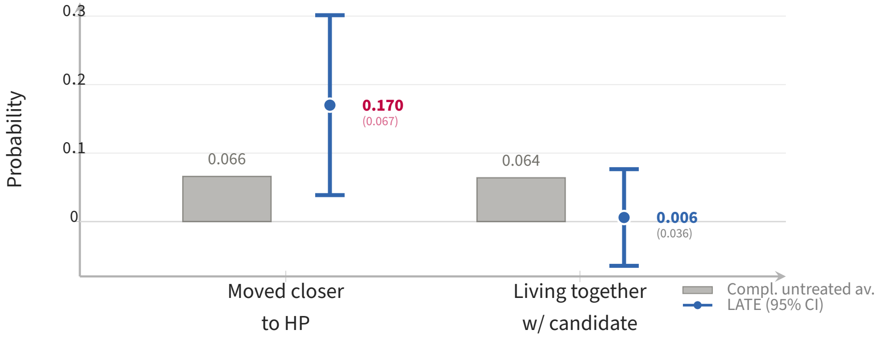
Effect of Candidate Move on Location Choices of Relatives: IV Estimates

We instrument for *candidate's move to the HP*, obtain effect on *their relatives' location choices*



Effect of Candidate Move on Location Choices of Relatives: IV Estimates

We instrument for *candidate's move to the HP*, obtain effect on *their relatives' location choices*

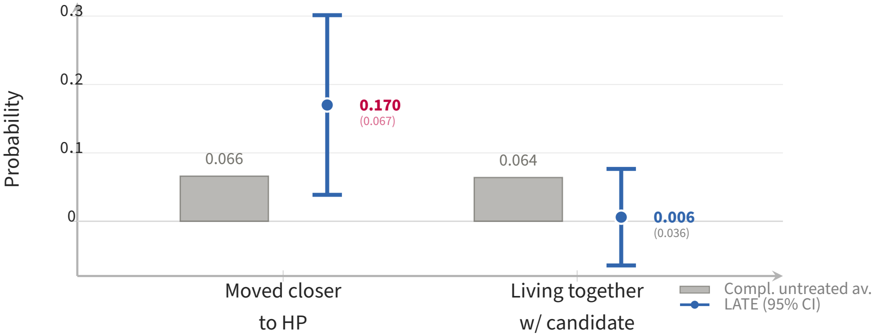


Relatives living nearby are 17 p.p. more likely to follow candidates who move away

$\Rightarrow$ Evidence of preference for living near family members, strong enough to relocate

Effect of Candidate Move on Location Choices of Relatives: IV Estimates

We instrument for *candidate's move to the HP*, obtain effect on *their relatives' location choices*



Relatives living nearby are 17 p.p. more likely to follow candidates who move away

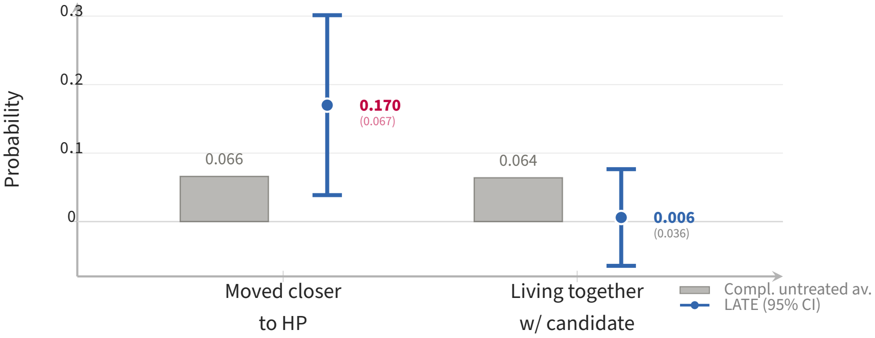
⟹ Evidence of preference for living near family members, strong enough to relocate

Followers move to higher poverty areas relative to baseline (consistent w/ Belchior et al. 2023)

Details

Effect of Candidate Move on Location Choices of Relatives: IV Estimates

We instrument for *candidate's move to the HP*, obtain effect on *their relatives' location choices*



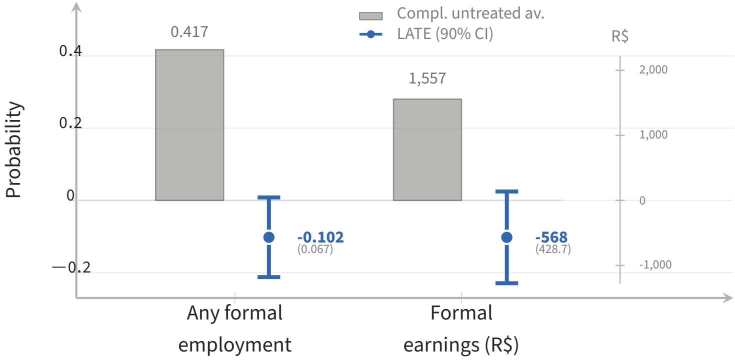
Relatives living nearby are 17 p.p. more likely to follow candidates who move away

$\Rightarrow$ Evidence of preference for living near family members, strong enough to relocate

Next: Does family proximity affect labor supply?

Effect of Candidate Move on Relatives' Labor Market Outcomes: IV Estimates

We use the same 2SLS model to study labor market outcomes

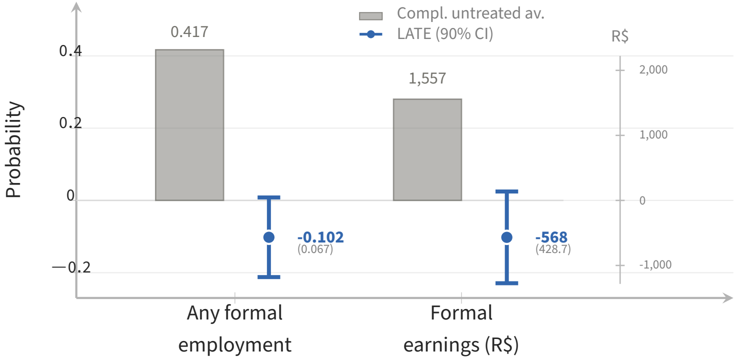


- Relatives' formal employment falls by 10 p.p.
- Earnings also fall: although insignificant, impacts are economically meaningful (30%)

More outcomes

Effect of Candidate Move on Relatives' Labor Market Outcomes: IV Estimates

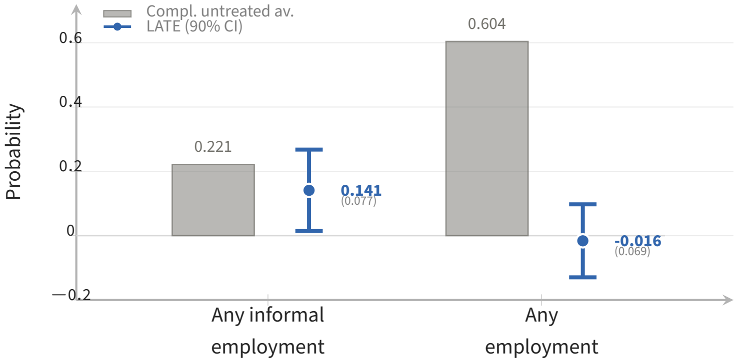
We use the same 2SLS model to study labor market outcomes



- Relatives' formal employment falls by 10 p.p.
- Earnings also fall: although insignificant, impacts are economically meaningful (30%)
- Caveat: informality is a well-documented margin of adjustment in similar contexts

More outcomes

Do Relatives Switch to Informal Work? IV Estimates

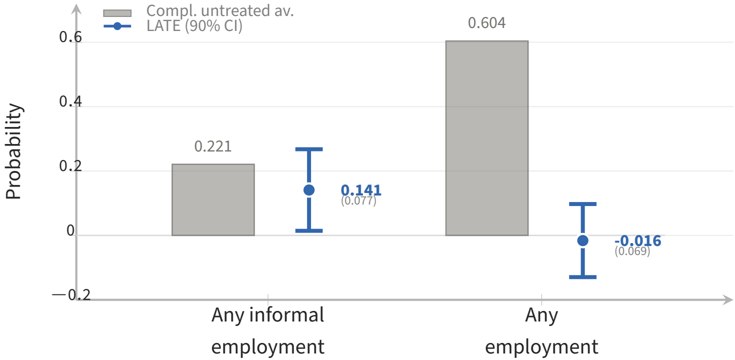


- ↑ informal emp, total emp unchanged. Formal workers switch to informal **switching**

Decomposition

- Informality as the flexible margin under caregiving shocks (Berniell et al. 2021; 2023, Clark et al. 2019) and policy incentives (Bosch and Campos-Vazquez 2014, Gerard and Gonzaga 2021, Gerard et al. 2021, Derenoncourt et al. 2025)

Do Relatives Switch to Informal Work? IV Estimates



- ↑ informal emp, total emp unchanged. Formal workers switch to informal **switching**

Decomposition

Could be loss of proximity, but relatives are also moving, which could disrupt labor market attachment

Q: effects of **separation** or **disruption due to moving** to maintain proximity?

Relative Types

Possible response types for relatives:

- Relatives who are **followers** : $R_i(1) > R(0)$
- Relatives who **stay behind**: $R_i(1) = R(0) = 0$
- Relatives who are **always-movers**: $R_i(1) = R(0) = 1$
- Relatives who are **avoiders**: $R_i(1) < R(0)$

Relative Types

Possible response types for relatives:

- Relatives who are **followers** : $R_i(1) > R(0)$
- Relatives who **stay behind**: $R_i(1) = R(0) = 0$
- Relatives who are **always-movers**: $R_i(1) = R(0) = 1$
- ~~Relatives who are avoiders $R_i(1) < R(0)$~~

We introduce an additional assumption:

A4 - Monotonicity in relative's location choice. $\Pr(R_i(1) \geq R_i(0)) = 1$

This assumption allows us to identify the share of each relative type

Who are the followers?

Relative Types

Possible response types for relatives:

- Relatives who are **followers** : $R_i(1) > R(0)$ $\frac{E[R|Z=1]-E[R|Z=0]}{E[D|Z=1]-E[D|Z=0]} = 0.170$ (0.067)
- Relatives who **stay behind**: $R_i(1) = R(0) = 0$ $\frac{E[D(1-R)|Z=1]-E[D(1-R)|Z=0]}{E[D|Z=1]-E[D|Z=0]} = 0.763$ (0.034)
- Relatives who are **always-movers**: $R_i(1) = R(0) = 1$ $\frac{E[(D-1)R|Z=1]-E[(D-1)R|Z=0]}{E[D|Z=1]-E[D|Z=0]} = 0.066$ (0.058)
- ~~Relatives who are avoiders~~ $\cancel{R_i(1) < R(0)}$

We introduce an additional assumption:

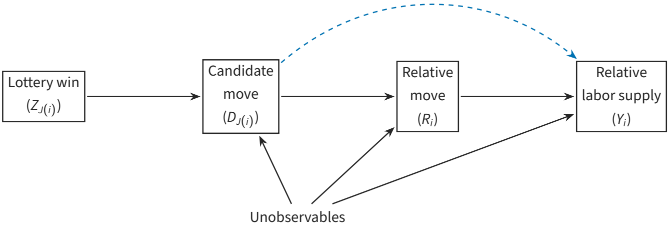
A4 - Monotonicity in relative's location choice. $\Pr(R_i(1) \geq R_i(0)) = 1$

This assumption allows us to identify the share of each relative type

Who are the followers?

No evidence of "always movers" (share not statistically different from zero)

Empirical Framework



We make the following assumptions:

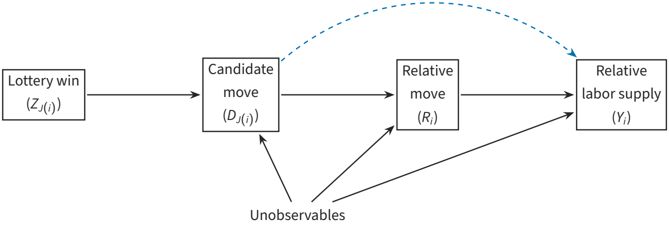
AY1 - Random assignment. $Y_i(d,r) \perp\!\!\!\perp Z_{J(i)}$

AY2 - Exclusion from potential outcome. $Y_i(d,r,z) \equiv Y_i(d,r) \quad \forall (d,r,z)$

Kitagawa test

$Y_i(d,r,z)$: the potential labor market outcome of relative i of candidate $J(i)$ under location choices (d,r) and lottery outcome z

Empirical Framework



We make the following assumptions:

AY1 - Random assignment. $Y_i(d,r) \perp\!\!\!\perp Z_{J(i)}$

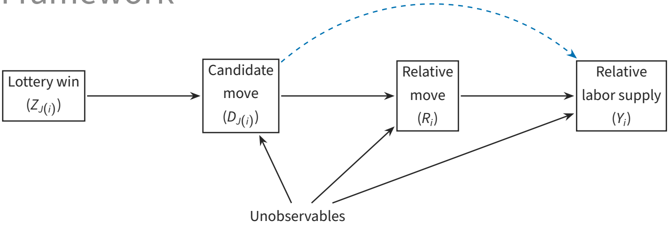
AY2 - Exclusion from potential outcome. $Y_i(d,r,z) \equiv Y_i(d,r) \quad \forall (d,r,z)$

Kitagawa test

$Y_i(d,r,z)$: the potential labor market outcome of relative i of candidate $J(i)$ under location choices (d,r) and lottery outcome z

- Under-identification: 2 endogenous variables (D and R) and only 1 instrument (Z)

Empirical Framework



We make the following assumptions:

AY1 - Random assignment. $Y_i(d, r) \perp\!\!\!\perp Z_{J(i)}$

AY2 - Exclusion from potential outcome. $Y_i(d, r, z) \equiv Y_i(d, r) \quad \forall (d, r, z)$

Kitagawa test

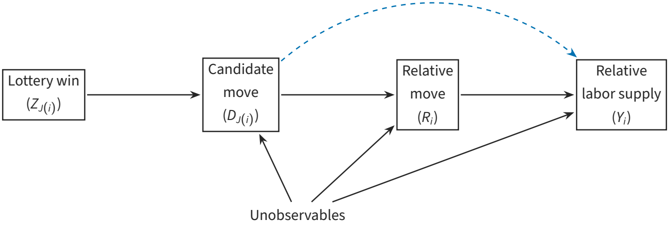
$Y_i(d, r, z)$: the potential labor market outcome of relative i of candidate $J(i)$ under location choices (d, r) and lottery outcome z

- Under-identification: 2 endogenous variables (D and R) and only 1 instrument (Z)
- Instrumenting Y_i on D_j using Z_j identifies (Mountjoy 2022, Kirkeboen et al. 2016):

$$\beta_{IV} = \pi_0 \underbrace{LATE_0}_{\text{seperation effect}} + \pi_1 \underbrace{LATE_1}_{\text{joint move effect}}$$

for Relatives who **stay behind**, Relatives who are **followers**

Empirical Framework



We make the following assumptions:

AY1 - Random assignment. $Y_i(d, r) \perp\!\!\!\perp Z_{J(i)}$

AY2 - Exclusion from potential outcome. $Y_i(d, r, z) \equiv Y_i(d, r) \quad \forall (d, r, z)$

Kitagawa test

$Y_i(d, r, z)$: the potential labor market outcome of relative i of candidate $J(i)$ under location choices (d, r) and lottery outcome z

- Under-identification: 2 endogenous variables (D and R) and only 1 instrument (Z)
- Instrumenting Y_i on D_j using Z_j identifies (Mountjoy 2022, Kirkeboen et al. 2016):

$$\beta_{IV} = \pi_0 \underbrace{LATE_0}_{\text{seperation effect}} + \pi_1 \underbrace{LATE_1}_{\text{joint move effect}}$$

No point identification of the sub-LATEs. Estimate sharp bounds (Horowitz and Manski 1995, Dutz et al. 2024)

Details

Labor Market Response

				Complier			
	Non-taker	RF	OLS	Untreated	LATE	$LATE_0$ (stay behind)	$LATE_1$ (followers)
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any formal employment	0.299	-0.017	-0.057	0.417	-0.102	[-0.215, -0.008]	[-0.476, 0.244]
	(0.002)	(0.010)	(0.024)	(0.070)	(0.067)		
Any informal employment	0.354	0.030	0.005	0.221	0.141	[0.214, 0.437]	[-0.325, 0.396]
	(0.001)	(0.013)	(0.026)	(0.071)	(0.077)		
Any employment	0.607	0.003	-0.042	0.604	-0.016	[-0.063, 0.16]	[-0.238, 0.482]
	(0.001)	(0.012)	(0.023)	(0.065)	(0.069)		

All outcomes

Labor Market Response

				Complier			
	Non-taker	RF	OLS	Untreated	LATE	$LATE_0$ (stay behind)	$LATE_1$ (followers)
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any formal employment	0.299	-0.017	-0.057	0.417	-0.102	[-0.215, -0.008]	[-0.476, 0.244]
	(0.002)	(0.010)	(0.024)	(0.070)	(0.067)		
Any informal employment	0.354	0.030	0.005	0.221	0.141	[0.214, 0.437]	[-0.325, 0.396]
	(0.001)	(0.013)	(0.026)	(0.071)	(0.077)		
Any employment	0.607	0.003	-0.042	0.604	-0.016	[-0.063, 0.16]	[-0.238, 0.482]
	(0.001)	(0.012)	(0.023)	(0.065)	(0.069)		

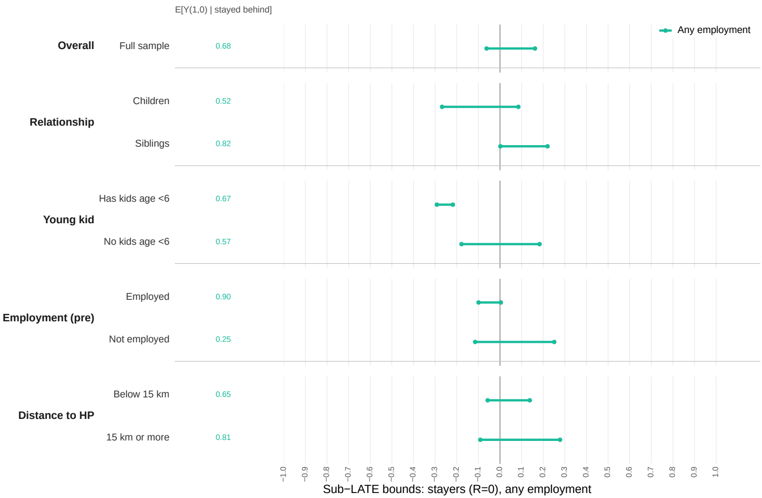
Relatives who stay behind **switch** from formal to informal jobs ($LATE_0$)

$\Longrightarrow$ proximity to family members has **production value**

Bounds for **followers** are too wide to conclude anything ($LATE_1$)

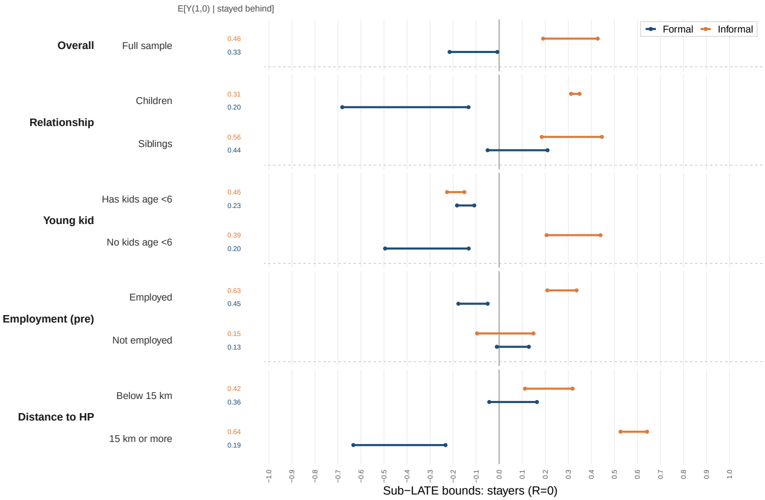
All outcomes

Heterogeneity in Effects of Separation (For Stayers)



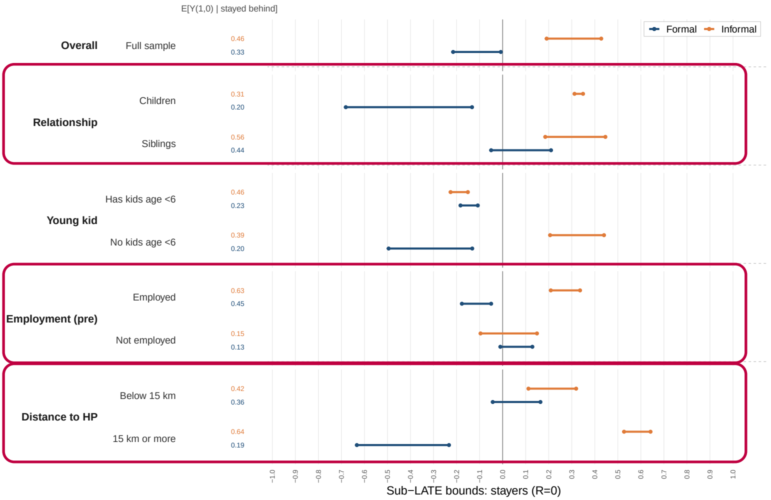
- $\downarrow$ employment among households with young kids

Heterogeneity in Effects of Separation (For Stayers)



- ↓ employment among households with young kids

Heterogeneity in Effects of Separation (For Stayers)



- ↓ employment among households with young kids
- Switching to informal mostly among working children and people living far from the HP

Test for Selection on Labor Market Gains

- So far: separation reduces formal employment $\Rightarrow$ family proximity has **production value**

Test for Selection on Labor Market Gains

- So far: separation reduces formal employment $\Rightarrow$ family proximity has **production value**
- But production value need not **drive location choices**
  - Proximity also has **amenity value** (companionship, time together)
    - Sorting on production value $\implies$ move is correlated with earnings

Toy model

Test for Selection on Labor Market Gains

Toy model

- So far: separation reduces formal employment $\Rightarrow$ family proximity has **production value**
- But production value need not **drive location choices**
  - Proximity also has **amenity value** (companionship, time together)
    - Sorting on production value $\implies$ move is correlated with earnings
- If the production channel drives choices, followers, $F \equiv \mathbb{1}\{R(1) > R(0)\}$, gain more from proximity than those who stay behind, $B \equiv \mathbb{1}\{R(1) = 0, T = C\}$:

$$\underbrace{E[Y(1,1) - Y(1,0) \mid F]}_{\text{followers' gain from proximity}} > \underbrace{E[Y(1,1) - Y(1,0) \mid B]}_{\text{stayers' gain from proximity}}$$

Test for Selection on Labor Market Gains

- So far: separation reduces formal employment $\Rightarrow$ family proximity has **production value**
- But production value need not **drive location choices**
  - Proximity also has **amenity value** (companionship, time together)

    Toy model

    - Sorting on production value $\Longrightarrow$ move is correlated with earnings
- If the production channel drives choices, followers, $F \equiv \mathbb{1}\{R(1) > R(0)\}$, gain more from proximity than those who stay behind, $B \equiv \mathbb{1}\{R(1) = 0, T = C\}$:

$$\underbrace{E[Y(1,1) - Y(1,0) \mid F]}_{\text{followers' gain from proximity}} > \underbrace{E[Y(1,1) - Y(1,0) \mid B]}_{\text{stayers' gain from proximity}}$$

- But we only (partially) identify $E[Y(1,0) - Y(0,0)|B]$ and $E[Y(1,1) - Y(0,0)|F]$: each group's counterfactual location choice is never observed

Solution: "connect the dots" with parametric selection model (Kline and Walters 2016)

Selection Model

Location choices:

$$D_{J(i)} = \mathbb{1}\{\mu_c(X_{J(i)}, Z_{J(i)}) + \nu_{cJ(i)} > 0\}$$
$$R_i = \mathbb{1}\{\mu_n(X_i, D_{J(i)}) + \nu_{ni} > 0\}$$

- $(\nu_{cJ(i)}, \nu_{ni})$ are unobserved tastes, moving costs, and other constraints to mobility

$$(\nu_{cJ(i)}, \nu_{ni})|X_i, Z_{J(i)} \sim N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

Potential labor market outcomes:

$$E[Y(d,r)|X, Z, \nu_c, \nu_n] = \psi_{(d,r)}(X) + \beta_{(d,r)c}\nu_c + \beta_{(d,r)n}\nu_n$$

- (Productive value) followers gain more from *restoring proximity* – sorting on earning gains

$$\beta_{(1,0)n} < \beta_{(1,1)n}$$

Identification: interaction of X (strata and dist from HP) with Z (Kline and Walters 2016, Hull 2015) Identification

Do Relatives Sort on Labor-Market Gains? A Roy Test

We estimate the model and test: Model Parameters

$$\Delta \equiv \underbrace{E[Y(1,1) - Y(1,0) \mid F]}_{\text{followers' gain from proximity}} - \underbrace{E[Y(1,1) - Y(1,0) \mid B]}_{\text{stayers' gain from proximity}} > 0, \quad H_0 : \Delta \leq 0 \text{ vs. } \Delta > 0.$$

$\Delta > 0$ = relatives who follow are those who gain most $\Rightarrow$ Roy sorting

Panel A. Roy test: Δ *byoutcome*

	Formal empl.	Informal empl.	Earnings (R$)
$\widehat{\Delta}$	**1.329**	−0.734	4459
	(0.605)	(0.748)	(4789)

Panel B. Loadings, formal employment Informal & earnings

	Separate $(1,0)$	Follow $(1,1)$
β_c (candidate)	−0.007	−0.059
β_n (*relative*)	**−0.770**	+**0.114**

$\Delta\beta_n = \beta_{11n} - \beta_{10n} = +\mathbf{0.88} > 0$: *followers gain more.*

Conclusion

Housing lotteries $+$ family-network data $\Rightarrow$ effects of family proximity on loc. choice & labor supply

Conclusion

Housing lotteries $+$ family-network data $\Rightarrow$ effects of family proximity on loc. choice & labor supply

1. **Family proximity shapes location choices**

    $\longrightarrow$ Revealed pref. for proximity, not inertia or shared preferences over location

Conclusion

Housing lotteries $+$ family-network data $\Rightarrow$ effects of family proximity on loc. choice & labor supply

1. **Family proximity shapes location choices**

   $\rightarrow$ Revealed pref. for proximity, not inertia or shared preferences over location

2. **Separation from family affects labor supply**

   $\rightarrow$ Proximity has production value

Conclusion

Housing lotteries $+$ family-network data $\Rightarrow$ effects of family proximity on loc. choice & labor supply

1. **Family proximity shapes location choices**

   $\longrightarrow$ Revealed pref. for proximity, not inertia or shared preferences over location

2. **Separation from family affects labor supply**

   $\longrightarrow$ Proximity has production value

3. **Relatives sort on labor-market gains from proximity**

   ▷ We find evidence that followers' formal employment gain is larger than stayers'

   $\longrightarrow$ Evidence for Roy selection on the production value of proximity

Conclusion

Housing lotteries $+$ family-network data $\Rightarrow$ effects of family proximity on loc. choice & labor supply

1. **Family proximity shapes location choices**
   - $\rightarrow$ Revealed pref. for proximity, not inertia or shared preferences over location
2. **Separation from family affects labor supply**
   - $\rightarrow$ Proximity has production value
3. **Relatives sort on labor-market gains from proximity**
   - ▷ We find evidence that followers' formal employment gain is larger than stayers'
   - $\rightarrow$ Evidence for Roy selection on the production value of proximity

$\Longrightarrow$ MCMV generates spillovers on recipients' relatives':

- ▷ location choices - send them to worse neighborhoods
- ▷ Switch formal $\rightarrow$ informal employment

Conclusion

Housing lotteries $+$ family-network data $\Rightarrow$ effects of family proximity on loc. choice & labor supply

1. **Family proximity shapes location choices**

   $\rightarrow$ Revealed pref. for proximity, not inertia or shared preferences over location

2. **Separation from family affects labor supply**

   $\rightarrow$ Proximity has production value

3. **Relatives sort on labor-market gains from proximity**

   - We find evidence that followers' formal employment gain is larger than stayers'

   $\rightarrow$ Evidence for Roy selection on the production value of proximity

$\Longrightarrow$ MCMV generates spillovers on recipients' relatives':

- location choices - send them to worse neighborhoods
- Switch formal $\rightarrow$ informal employment

General implications:

Conclusion

Housing lotteries $+$ family-network data $\Rightarrow$ effects of family proximity on loc. choice & labor supply

1. **Family proximity shapes location choices**

   $\rightarrow$ Revealed pref. for proximity, not inertia or shared preferences over location

2. **Separation from family affects labor supply**

   $\rightarrow$ Proximity has production value

3. **Relatives sort on labor-market gains from proximity**

   ▷ We find evidence that followers' formal employment gain is larger than stayers'

   $\rightarrow$ Evidence for Roy selection on the production value of proximity

$\Longrightarrow$ MCMV generates spillovers on recipients' relatives':

- ▷ location choices - send them to worse neighborhoods
- ▷ Switch formal $\rightarrow$ informal employment

General implications:

- Relocation programs have network spillovers: a **"multiplier effect" on moves to / from opportunity**

Conclusion

Housing lotteries $+$ family-network data $\Rightarrow$ effects of family proximity on loc. choice & labor supply

1. **Family proximity shapes location choices**

   $\rightarrow$ Revealed pref. for proximity, not inertia or shared preferences over location

2. **Separation from family affects labor supply**

   $\rightarrow$ Proximity has production value

3. **Relatives sort on labor-market gains from proximity**

   - ▷ We find evidence that followers' formal employment gain is larger than stayers'

   $\rightarrow$ Evidence for Roy selection on the production value of proximity

$\Longrightarrow$ MCMV generates spillovers on recipients' relatives':

- ▷ location choices - send them to worse neighborhoods
- ▷ Switch formal $\rightarrow$ informal employment

General implications:

- Relocation programs have network spillovers: a **"multiplier effect" on moves to / from opportunity**
- Location choices load on productive value $\Rightarrow$ **replacing family care can relax mobility frictions**

Thank you!

References

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association 97*(457), 284–292.

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

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

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

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

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

Beaman, L. A. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the U.S. *The Review of Economic Studies 79*(1), 128–161.

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

References (cont.)

Bergman, P., R. Chetty, S. DeLuca, N. Hendren, L. F. Katz, and C. Palmer (2024). Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice. *American Economic Review 114*(5), 1281–1337.

Berniell, I., L. Berniell, D. de la Mata, M. Edo, and M. Marchionni (2021). Gender gaps in labor informality: The motherhood effect. *Journal of Development Economics 150*, 102599.

Berniell, I., L. Berniell, D. de la Mata, M. Edo, and M. Marchionni (2023). Motherhood and flexible jobs: Evidence from latin american countries. *World Development 167*, 106225.

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

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. Technical report, National Bureau of Economic Research.

Borusyak, K. and P. Hull (2023). Nonrandom exposure to exogenous shocks. *Econometrica 91*(6), 2155–2185.

Bosch, M. and R. M. Campos-Vazquez (2014). The trade-offs of welfare policies in labor markets with informal jobs: The case of the "seguro popular" program in mexico. *American Economic Journal: Economic Policy 6*(4), 71–99.

Bosch, M. and W. F. Maloney (2010). Comparative analysis of labor market dynamics using markov processes: An application to informality. *Labour Economics 17*(4), 621–631.

References (cont.)

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

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

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

Chiquiar, D. and G. H. Hanson (2005). International migration, self-selection, and the distribution of wages: Evidence from mexico and the united states. *Journal of Political Economy 113*(2), 239–281.

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

Clark, S., C. W. Kabiru, S. Laszlo, and S. Muthuri (2019). The impact of childcare on poor urban women's economic empowerment in africa. *Demography 56*(4), 1247–1272.

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

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

References (cont.)

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

Derenoncourt, E., F. Gerard, L. Lagos, and C. Montialoux (2025). Minimum wages and informality. Technical report, National Bureau of Economic Research.

Diamond, R. and C. Gaubert (2022). Spatial sorting and inequality. *Annual Review of Economics 14*(1), 795–819.

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

Engbom, N., G. Gonzaga, C. Moser, and R. Olivieri (2022). Earnings inequality and dynamics in the presence of informality: The case of brazil. *Quantitative Economics 13*(4), 1405–1446.

Fetter, D., L. Lockwood, and P. Mohnen (2024). Long-run intergenerational effects of social security. Technical report, Working Paper.

Franklin, S. (2020). Enabled to work: The impact of government housing on slum dwellers in south africa. *Journal of Urban Economics 118*, 103265.

Gerard, F. and G. Gonzaga (2021). Informal labor and the efficiency cost of social programs: Evidence from unemployment insurance in brazil. *American Economic Journal: Economic Policy 13*(3), 167–206.

Gerard, F., J. Naritomi, and J. Silva (2021). Cash transfers and formal labor markets: Evidence from brazil.

References (cont.)

Green, A. (2024). Networks and geographic mobility: Evidence from world war ii navy ships. Technical report, Working Paper.

Günther, I. and A. Launov (2012). Informal employment in developing countries: Opportunity or last resort? *Journal of Development Economics 97*(1), 88–98.

Horowitz, J. L. and C. F. Manski (1995). Identification and robustness with contaminated and corrupted data. *Econometrica: Journal of the Econometric Society*, 281–302.

Hull, P. (2015). Isolateing: Identifying counterfactual-specific treatment effects by stratified comparisons.

Imbens, G. W. and D. B. Rubin (1997). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies 64*(4), 555–574.

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

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

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

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

References (cont.)

Kline, P. and C. R. Walters (2016). Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics 131*(4), 1795–1848.

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

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

Leite Mariante, G. and B. Ribeiro (2025). Subsidized housing and intergenerational mobility: evidence from brazil. Technical report, Working Paper.

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

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

Maloney, W. F. (1999). Does informality imply segmentation in urban labor markets? evidence from sectoral transitions in mexico. *The World Bank Economic Review 13*(2), 275–302.

Martínez Leyva, A. (2025, July). Job flexibility and informality. Job Market Paper, University of Warwick, Department of Economics.

McKenzie, D. and H. Rapoport (2010). Self-selection patterns in mexico-u.s. migration: The role of migration networks. *Review of Economics and Statistics 92*(4), 811–821.

References (cont.)

Meghir, C., R. Narita, and J.-M. Robin (2015). Wages and informality in developing countries. *American Economic Review 105*(4), 1509–1546.

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

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

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

Posadas, J. and M. Vidal-Fernández (2013). Grandparents' childcare and female labor force participation. *IZA Journal of Labor Policy 2*, 14.

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

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

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

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

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence from brazil. *American Economic Review 108*(8), 2015–2047.

References (cont.)

Ulyssea, G. (2020). Informality: Causes and consequences for development. *Annual Review of Economics 12*, 525–546.

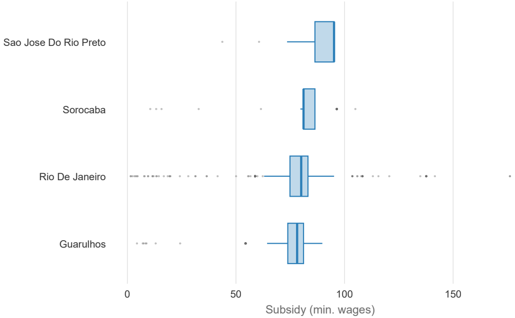
Zamarro, G. (2020). Family labor participation and child care decisions: The role of grannies. *SERIEs 11*(3), 287–312.

Appendix: Further Details on the Setting

How Large Is the MCMV Subsidy?

Back

- Median subsidy: **R$ 62,000 ≈ 80 monthly minimum wages** (CAIXA records, matched lottery takers)
- Across cities, median subsidy ranges from R$ 59,000 (Rio) to R$ 76,000 (Sorocaba); 70 to 95 monthly minimum wages



Data from CAIXA records contracts, matched on all MCMV lottery roster

Monthly minimum wage used is year specific from IPEADATA (series Salário mínimo)

The Lottery and Assignment Process

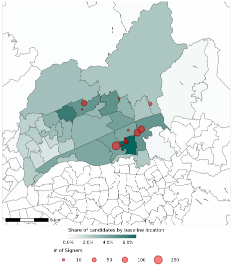
Back

Faixa-1 selection regulated by federal law and Ministry of Cities rules (Law 11,977/2009; Portarias 140/2010, 610/2011, 163/2016)

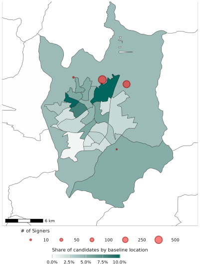
- **Roster**: applicants registered in the **municipal housing cadaster** meeting Faixa-1 eligibility (income below program ceiling; no homeownership; no prior federal housing benefit)
- **Selection**: when demand exceeds units, allocation is by **public lottery**, random *within strata*; design differs by city:
  - ▷ **Rio de Janeiro** (2015 onward): drawn individuals **ranked by age**, units allocated to the oldest, remainder to reserve $\Rightarrow$ treatment probability varies with age within a draw $\rightarrow$ strata $=$ within-lottery **age deciles**
  - ▷ **Other cities**: applicants binned by **number of prioritization criteria met** (typically 0–2, 3–4, 5–6; e.g., female-headed household, disabled member, risk-area residence), draws within bins $\rightarrow$ strata $=$ observed **criteria groups**
- **Take-up**: winners must document eligibility with CAIXA and sign the contract
- **Implication for empirical specification**: lottery $\times$ strata fixed effects in all analyses

Housing Projects in Other Cities

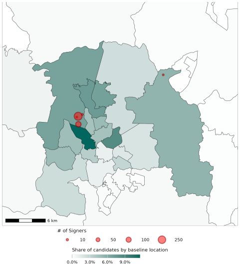
Back



Guarulhos



Sao Jose do Rio Preto



Sorocaba

Appendix: Details on CadÚnico and Variable Construction

Details on CadÚnico

Back

CadÚnico: "Census of the poor" in Brazil

- Administrative registry of (potential) beneficiaries of federal social programs
- Covers 40 percent of the bottom of the national income distribution
    - Our access includes the full universe of records $\Rightarrow$ 80 million individuals per year
- Information on education, self-reported labor and non-labor income, and location of residence
- Enrollment and updates conducted at municipal offices, or by trained enumerators at residency

Data structure:

- Families expected to update records no later than every 2 years

  $\longrightarrow$ Each entry contains individual and family identifiers

  $\longrightarrow$ We pool post-lottery data across 4 years

  $\longrightarrow$ This provides us with post-lottery address data for about 92% of our candidates' analysis sample

Descriptive Statistics for Candidates

	Loser (1)	Winner (2)	Adj. Diff. (3)	(s.e.) (4)
(i) Demographics				
Age	39.53	39.576	-0.005	(0.134)
Older than 60	0.065	0.062	0.003	(0.003)
Male	0.19	0.172	0.000	(0.006)
PBF beneficiary	0.550	0.399	0.008	(0.008)
Has children under 16	0.428	0.433	0.018	(0.013)
Is a grandparent	0.088	0.037	0.003	(0.003)
Distance to closest HP	15.859	8.866	0.059	(0.158)
(ii) Family network				
N relatives	4.11	4.134	0.053	(0.051)
(iii) Labor market				
Any employment before lottery	0.75	0.754	-0.005	(0.007)
Formal employment before lottery	0.639	0.43	-0.007	(0.008)
Total formal earnings before lottery	1341.829	1271.371	-1.419	(44.056)
N candidate $\times$ lottery	605,204	5,016		
N candidate	81,891	4,677		

Note: adjusted difference controls for lottery and strata fixed effects. Strata fixed effects are the randomization design covariates.

Empirical specification | Balance across lotteries | Stratified randomization explanation | Back to data

Descriptive Statistics for Candidates

	Loser (1)	Winner (2)	Adj. Diff. (3)	(s.e.) (4)
(i) Demographics				
Age	39.53	39.576	-0.005	(0.134)
Older than 60	0.065	0.062	0.003	(0.003)
Male	0.19	0.172	0.000	(0.006)
PBF beneficiary	0.550	0.399	0.008	(0.008)
Has children under 16	0.428	0.433	0.018	(0.013)
Is a grandparent	0.088	0.037	0.003	(0.003)
Distance to closest HP	15.859	8.866	0.059	(0.158)
(ii) Family network				
N relatives	4.11	4.134	0.053	(0.051)
(iii) Labor market				
Any employment before lottery	0.75	0.754	-0.005	(0.007)
Formal employment before lottery	0.639	0.43	-0.007	(0.008)
Total formal earnings before lottery	1341.829	1271.371	-1.419	(44.056)
N candidate × lottery	605,204	5,016		
N candidate	81,891	4,677		

Note: adjusted difference controls for lottery and strata fixed effects. Strata fixed effects are the randomization design covariates.

Winners and losers are balanced

Empirical specification | Balance across lotteries | Stratified randomization explanation | Back to data

Descriptive Statistics for Candidates

	Loser (1)	Winner (2)	Adj. Diff. (3)	(s.e.) (4)
(i) Demographics				
Age	39.53	39.576	-0.005	(0.134)
Older than 60	0.065	0.062	0.003	(0.003)
Male	0.19	0.172	0.00	(0.006)
PBF Beneficiary	0.550	0.399	0.008	(0.008)
Has children under 16	0.428	0.433	0.018	(0.013)
Is a grandparent	0.088	0.037	0.003	(0.003)
Distance to closest HP	15.859	8.866	0.059	(0.158)
(ii) Family network				
N relatives	4.11	4.134	0.053	(0.051)
(iii) Labor market				
Any employment before lottery	0.75	0.754	-0.005	(0.007)
Formal employment before lottery	0.639	0.43	-0.007	(0.008)
Total formal earnings before lottery	1341.829	1271.371	-1.419	(44.056)
N candidate $\times$ lottery	605,204	5,016		
N candidate	81,891	4,677		

Note: adjusted difference controls for lottery and strata fixed effects. Strata fixed effects are the randomization design covariates.

Predominantly women, many are Programa Bolsa Família (PBF) recipients

Empirical specification | Balance across lotteries | Stratified randomization explanation | Back to data

Descriptive Statistics for Candidates

	Loser (1)	Winner (2)	Adj. Diff. (3)	(s.e.) (4)
(i) Demographics				
Age	39.53	39.576	-0.005	(0.134)
Older than 60	0.065	0.062	0.003	(0.003)
Male	0.19	0.172	0.000	(0.006)
PBF beneficiary	0.550	0.399	0.008	(0.008)
Has children under 16	0.428	0.433	0.018	(0.013)
Is a grandparent	0.088	0.037	0.003	(0.003)
Distance to closest HP	15.859	8.866	0.059	(0.158)
(ii) Family network				
N relatives	4.11	4.134	0.053	(0.051)
(iii) Labor market				
Any employment before lottery	0.75	0.754	-0.005	(0.007)
Formal employment before lottery	0.639	0.43	-0.007	(0.008)
Total formal earnings before lottery	1341.829	1271.371	-1.419	(44.056)
N candidate $\times$ lottery	605,204	5,016		
N candidate	81,891	4,677		

Note: adjusted difference controls for lottery and strata fixed effects. Strata fixed effects are the randomization design covariates.

We identify an avg of $\sim$4 relatives per recipient

Empirical specification | Balance across lotteries | Stratified randomization explanation | Back to data

Descriptive Statistics for Relatives

	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	34.928	33.07	0.442	34.992	33.213	-0.338	36.54	35.411	0.014
			(0.28)			(0.396)			(0.349)
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)
Relationship: Parent	0.295	0.358	-0.016	0.181	0.118	-0.016	0.075	0.088	0
			(0.008)			(0.008)			(0.007)
Relationship: Sibling	0.158	0.11	-0.003	0.261	0.3	-0.012	0.336	0.322	0.001
			(0.006)			(0.011)			(0.012)
Relationship: Child	0.352	0.146	0.009	0.256	0.196	-0.017	0.095	0.096	0.005
			(0.007)			(0.012)			(0.008)
Relationship: Other	0.195	0.386	0.009	0.302	0.386	0.046	0.493	0.494	-0.007
			(0.009)			(0.014)			(0.014)
Formal employment before lottery	0.471	0.501	-0.008	0.329	0.435	0.003	0.177	0.41	0.003
			(0.004)			(0.005)			(0.005)
PBF Beneficiary	0.5	0.353	-0.009	0.588	0.504	-0.017	0.61	0.5	0.017
			(0.01)			(0.017)			(0.015)
N pair $\times$ lottery	328,452	4,697		534,427	2,208		404,078	3,049	
N unique pairs	69,696	4,644		68,521	2,195		61,670	3,015	

Note: adjusted difference controls for lottery and strata fixed effects. Nearby = less than 3 km, Far = more than 3 km.

$^1/_3$ live with candidate, $^1/_3$ w/in 3km, $^1/_3$ further away; results going forward are for **nearby relatives** unless indicated otherwise

Key Outcomes and Measurement

Back

Location choices (CadÚnico addresses; forward-filled when not updated; Geocoded using ArcGis)

$\rightarrow$ Baseline: $t = -1$; Post-lottery $t = 4$

- **Distances**: Harvesine distance between geo-locations
- **Candidate's choice** (binary): (D_j) is defined as living $>$2 km from baseline / within 500m of HP
- **Relative's choice** (binary): (R_i) is defined as living $>$500m closer to the candidate's housing project than at baseline

Labor market (year-level indicators, aggregated over the window, use CadÚnico $+$ RAIS)

$\rightarrow$ Baseline: $t \in [-3, -1]$; post-lottery: $t \in [0, 4]$

- **Formal employment**: a formal contract in RAIS in any year;
- **Informal employment**: reports employment in CadÚnico but holds no formal contract in any year (as in Belchior et al. 2023)
- **Any employment**: formal or informal employment in any year
- **Formal earnings**: sum of RAIS earnings over the window, zeros included (nominal R$)

Belchior et al. (2023) show there is high agreement between self-reported formal employment in CU and RAIS employment

Appendix: Constructing Family Networks

Constructing Family Networks

Back

Family links built from **CadÚnico**: household composition + reported **parents' full names**

- Two sources of links:
  1. **Co-residence**: relationship-to-head codes classify co-resident pairs (parents, children, siblings, spouses, grandparents, in-laws, step-relatives)
  2. **Parent names**:
     - individuals reporting the same mother *and* father are siblings,
     - a reported parent name matched to a person's own name is a parent–child link
- Higher-order relations by **chaining** (grandparent $=$ parent of parent; uncle/aunt $=$ sibling of parent; cousins; in-laws via spouses)
- Quality safeguards on name-based links:
  - ▷ **locally rare** names ($\leq$ 10 same-(full)name individuals in the city)
  - ▷ age-consistency checks (parent $\geq$ 10 yrs older; siblings within 50 yrs)
- Links **constructed from data observed before the lottery**, individuals **aged 16+**, in the **same city**

Appendix: More on Empirical Specs, Balance, and Valdity Checks

Empirical Specification: Balance, First Stage and Reduced Form

$$W_{ij} = \beta Z_j + \theta_{ls} + \gamma' X_{ij} + \varepsilon_j$$

- i indexes the relative, j indexes the candidate
- W_{ij} is the LHS variable of interest: characteristic X_{ij} for balance, location choices D_j, R_i, labor market outcomes Y_{ij}
- θ_{ls} is the lottery-strata fixed effect
- X_{ij} is included for relatives only, denotes network size (Borusyak and Hull 2023) and baseline formal employment
- Std. errors clustered at candidate j-level

Back to balance Back to results

Balance Across Lotteries

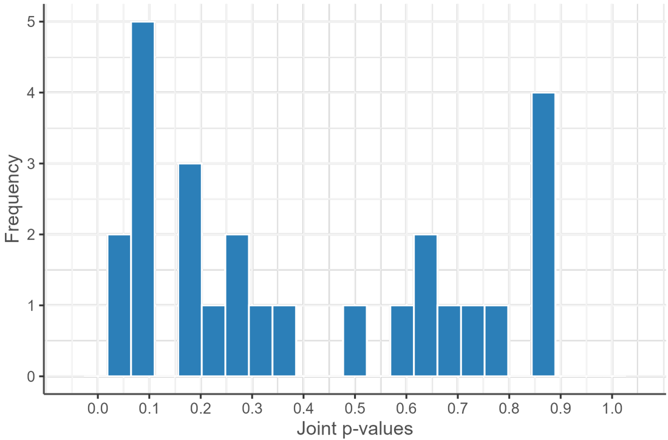


Figure 1.1: Distribution of p-values of the joint test of equality in each lottery

Back to main

Appendix: Lee Bounds on Candidate Location Choices

Effect of Lottery Win on Candidates' Location Choices

Back

	Loser mean	Estimate	(s.e.)	[Lee Bounds]	(LB 95% CIs)
	(1)	(2)	(3)	(4)	(5)
(i) Baseline sample					
Has address 4 yrs post-lottery	0.922	0.012	(0.003)		
(ii) Post-lottery CU respondents (up to 4 yrs)					

Post-lottery address data is available for 92% of the sample, mildly correlated with winning, but Lee bounds are very tight

Effect of Lottery Win on Candidates' Location Choices

Back

	Loser mean	Estimate	(s.e.)	[Lee Bounds]	(LB 95% CIs)
	(1)	(2)	(3)	(4)	(5)
(i) Baseline sample					
Has address 4 yrs post-lottery	0.922	0.012	(0.003)		
(ii) Post-lottery CU respondents (up to 4 yrs)					
Moves from baseline	0.244	0.160	(0.009)	[0.145, 0.176]	(0.134, 0.186)
Moves more than 2 kms	0.132	0.170	(0.008)	[0.154, 0.184]	(0.122, 0.175)
Moves to HP	0.007	0.156	(0.006)	[0.129, 0.159]	(0.116, 0.168)
Signed and Moves to HP	0.000	0.160	(0.006)	[0.138, 0.168]	(0.125, 0.178)
Distance moved (uncond.)	1.839	1.527	(0.151)	[1.481, 2.441]	(0.746, 5.062)

Winning increases the probability of moving to the HP by 16 pp

Effect of Lottery Win on Candidates' Location Choices

Back

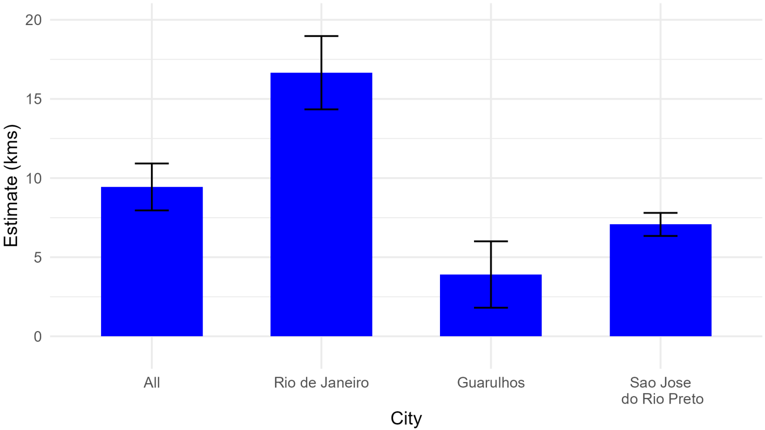
	Loser mean	Estimate	(s.e.)	[Lee Bounds]	(LB 95% CIs)
	(1)	(2)	(3)	(4)	(5)
(i) Baseline sample					
Has address 4 yrs post-lottery	0.922	0.012	(0.003)		
(ii) Post-lottery CU respondents (up to 4 yrs)					
Moves from baseline	0.244	0.160	(0.009)	[0.145, 0.176]	(0.134, 0.186)
Moves more than 2 kms	0.132	0.170	(0.008)	[0.154, 0.184]	(0.122, 0.175)
Moves to HP	0.007	0.156	(0.006)	[0.129, 0.159]	(0.116, 0.168)
Signed and Moves to HP	0.000	0.160	(0.006)	[0.138, 0.168]	(0.125, 0.178)
Distance moved (uncond.)	1.839	1.527	(0.151)	[1.481, 2.441]	(0.746, 5.062)

Distance moved by complier candidates $\frac{1.52}{0.16} = 9.54$ km

Appendix: Candidate Distance Moved by City

Distance Moved by Complier Candidates

Back

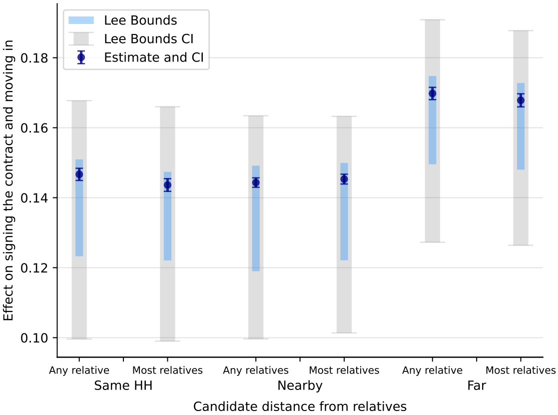


Complier candidates' moves averaged 5-17 kms across cities

Appendix: Take-up by Family Proximity at Baseline

Candidates' Relocation Choices, by Distance from Relatives

Back



Note: Nearby = less than 3 km, Far = more than 3 km

Program compliance is stronger among lottery candidates whose relatives do not live nearby

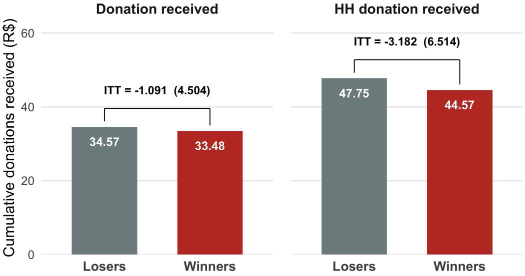
Appendix: Testing the Exclusion Restriction (A3)

Exclusion Restriction (A3): No Direct Transfer

Back

A3 requires the candidate winning to affect a relative's outcomes *only* through their move — not through a direct resource transfer (e.g., because the candidate now pays less in rent per month, they could share the additional income directly with their relative). If winning raised the donations relatives *receive*, that would be a direct income channel violating A3.

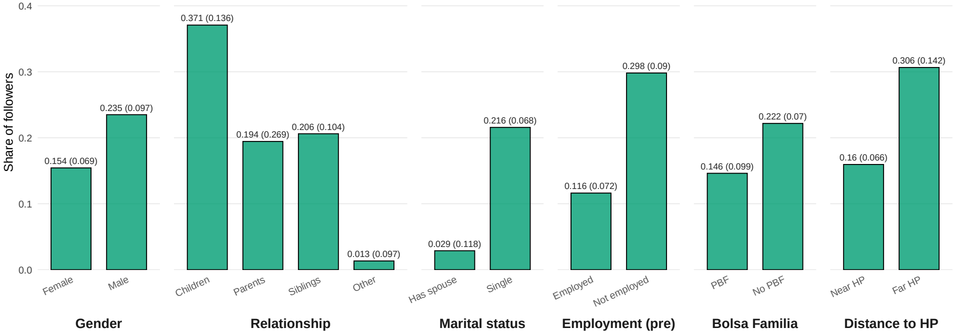
ITT on relatives' donations received, by baseline distance to candidate



No significant effect on donations received, consistent with A3 (exclusion)

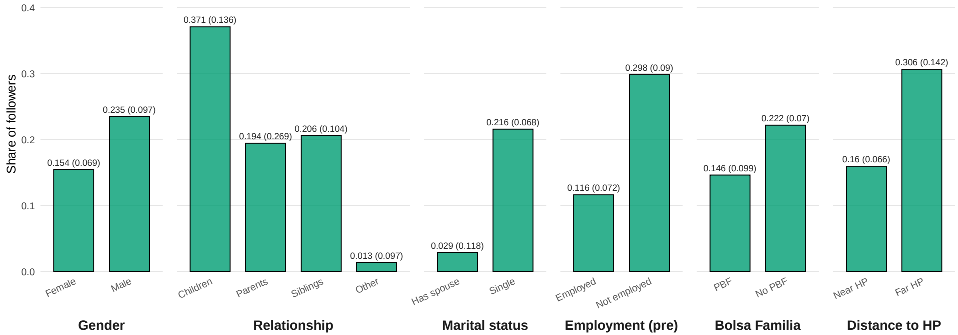
Who Are the Followers? Likelihood of Following by Observables

Back



Who Are the Followers? Likelihood of Following by Observables

Back



Unemployed, singles, children of candidates are most likely to follow

Appendix: Kitagawa Tests

Average Potential Outcomes among Compliers (Kitagawa 2015)

Back to main

- Under the LATE assumptions, the marginal distributions of $Y(1)$ and $Y(0)$ among compliers are identified (Imbens and Rubin 1997, Abadie 2002). Column (1) show estimated mean potential outcomes for compliers when treated, $\mathbb{E}[Y(1) \mid \text{complier}]$, and untreated, $\mathbb{E}[Y(0) \mid \text{complier}]$
- Instrument validity + monotonicity require the implied complier outcome probabilities to be nonnegative (Kitagawa 2015). Column (3) reports the number of violations: none, for any outcome. Estimates are sensible as probabilities (all in $[0, 1]$), consistent with the identifying assumptions.

	Treated Complier			Untreated Complier		
	Est.	(s.e.)	$\# \hat{q} < .1$	Est.	(s.e.)	$\# \hat{q} < .1$
Outcome	(1)	(2)	(3)	(1)	(2)	(3)
Panel A: Location choices						
Moves from baseline	0.365	(0.037)	0	0.139	(0.076)	0
Moves towards HP	0.237	(0.034)	0	0.066	(0.058)	0
Lives with candidate	0.07	(0.02)	0	0.064	(0.031)	0
Lives nearby candidate	0.128	(0.025)	0	0.18	(0.061)	0
Panel B: Labor supply						
Any employment	0.588	(0.029)	0	0.604	(0.065)	0
Any informal employment	0.361	(0.032)	0	0.221	(0.071)	0
Any formal employment	0.315	(0.03)	0	0.417	(0.054)	0
Kept same job (uncond.)	0.024	(0.009)	0	0.045	(0.025)	0
Has new job (uncond.)	0.251	(0.027)	0	0.336	(0.056)	0
PBF Beneficiary	0.424	(0.036)	0	0.34	(0.08)	0

Average Potential Outcomes by Relatives' Types (Kitagawa 2015)

Back

- Second-stage analogue of the previous test: with two location choices (d, r), assumptions AY1–AY2 identify some of the mean potential outcomes $Y(d, r)$ separately for each complier **relative type**: $\mathbb{E}[Y(1,0)|B]$ for those who stay behind (B) and $\mathbb{E}[Y(1,1)|F]$ for followers (F)
- Validity now requires implied probabilities in $[0, 1]$ *type by type*; violations reported as on previous slide; estimated among relatives living nearby.

	$E[Y(1,0)\|B]$			$E[Y(1,1)\|F]$		
	Est.	(s.e.)	$\#\hat{q} < .1$	Est.	(s.e.)	$\#\hat{q} < .1$
Outcome	(1)	(2)	(3)	(1)	(2)	(3)
Any employment	0.677	(0.038)	0	0.596	(0.055)	0
Any informal employment	0.475	(0.041)	0	0.363	(0.065)	0
Any formal employment	0.318	(0.038)	0	0.323	(0.061)	0
Kept same formal job (uncond.)	0.028	(0.012)	0	0.013	(0.012)	0
Has new formal job (uncond.)	0.247	(0.034)	0	0.259	(0.058)	0
Receives PBF	0.484	(0.046)	0	0.603	(0.067)	0

Appendix: Partial identification of subLATEs

Decomposing LATE into sub-LATEs

Back

- 2 endogenous variables (D and R) and only 1 instrument (Z)

Decomposing LATE into sub-LATEs

Back

- 2 endogenous variables (D and R) and only 1 instrument (Z)
- The TSLS of Y on D identifies (Kirkeboen et al. 2016, Mountjoy 2022):

$$\beta_{IA} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]} = \mathbb{E}[Y(1,R) - Y(0,R)|T_{J(i)} = C]$$

$$= \pi_1 LATE[Y(1,0) - Y(0,0)|R(1) = R(0) = 0, T_{J(i)} = C]$$

$$+ \pi_2 LATE[Y(1,1) - Y(0,0)|R(1) > R(0), T_{J(i)} = C]$$

$$+ \pi_3 LATE[Y(1,1) - Y(0,1)|R(1) = R(0) = 1, T_{J(i)} = C]$$

where:

- $T_{J(i)} = C := D(1) > D(0)$
- π_i are complier shares

Decomposing LATE into sub-LATEs

Back

- 2 endogenous variables (D and R) and only 1 instrument (Z)
- The TSLS of Y on D identifies (Kirkeboen et al. 2016, Mountjoy 2022):

$$\beta_{IA} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]} = \mathbb{E}[Y(1,R) - Y(0,R)|T_{J(i)} = C]$$

$$= \pi_1 LATE[Y(1,0) - Y(0,0)|R(1) = R(0) = 0, T_{J(i)} = C]$$

$$+ \pi_2 LATE[Y(1,1) - Y(0,0)|R(1) > R(0), T_{J(i)} = C]$$

$$+ \pi_3 LATE[Y(1,1) - Y(0,1)|R(1) = R(0) = 1, T_{J(i)} = C]$$

where:

- $T_{J(i)} = C := D(1) > D(0)$
- π_i are complier shares

Relatives who **stay behind**,

Decomposing LATE into sub-LATEs

Back

- 2 endogenous variables (D and R) and only 1 instrument (Z)
- The TSLS of Y on D identifies (Kirkeboen et al. 2016, Mountjoy 2022):

$$\beta_{IA} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]} = \mathbb{E}[Y(1,R) - Y(0,R)|T_{J(i)} = c]$$

$$= \pi_1 LATE[Y(1,0) - Y(0,0)|R(1) = R(0) = 0, T_{J(i)} = c]$$
$$+ \pi_2 LATE[Y(1,1) - Y(0,0)|R(1) > R(0), T_{J(i)} = c]$$
$$+ \pi_3 LATE[Y(1,1) - Y(0,1)|R(1) = R(0) = 1, T_{J(i)} = c]$$

where:

- $T_{J(i)} = c := D(1) > D(0)$
- π_i are complier shares

Relatives who **stay behind**, Relatives **followers**

Decomposing LATE into sub-LATEs

Back

- 2 endogenous variables (D and R) and only 1 instrument (Z)
- The TSLS of Y on D identifies (Kirkeboen et al. 2016, Mountjoy 2022):

$$\beta_{IA} = \frac{\mathbb{E}[Y|Z=1]-\mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1]-\mathbb{E}[D|Z=0]} = \mathbb{E}[Y(1,R)-Y(0,R)|T_{J(i)}=c]$$

$$= \pi_1 LATE[Y(1,0)-Y(0,0)|R(1)=R(0)=0, T_{J(i)}=c]$$

$$+ \pi_2 LATE[Y(1,1)-Y(0,0)|R(1)>R(0), T_{J(i)}=c]$$

$$+ \pi_3 LATE[Y(1,1)-Y(0,1)|R(1)=R(0)=1, T_{J(i)}=c]$$

where:

- $T_{J(i)} = c := D(1) > D(0)$
- π_i are complier shares

Relatives who **stay behind**, Relatives **followers**, Relatives **always-movers**

Decomposing LATE into sub-LATEs

Back

- 2 endogenous variables (D and R) and only 1 instrument (Z)
- The TSLS of Y on D identifies (Kirkeboen et al. 2016, Mountjoy 2022):

$$\beta_{IA} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]} = \mathbb{E}[Y(1,R) - Y(0,R)|T_{J(i)} = c]$$

$$= \pi_1 LATE[Y(1,0) - Y(0,0)|R(1) = R(0) = 0, T_{J(i)} = c]$$

$$+ \pi_2 LATE[Y(1,1) - Y(0,0)|R(1) > R(0), T_{J(i)} = c]$$

$$\cancel{+ \pi_3 LATE[Y(1,1) - Y(0,1)|R(1) = R(0) = 1, T_{J(i)} = c]}$$

where:

- $T_{J(i)} = c := D(1) > D(0)$
- π_i are complier shares

Relatives who **stay behind**, Relatives **followers**, ~~Relatives **always-movers**~~

Decomposing LATE into sub-LATEs

Back

- 2 endogenous variables (D and R) and only 1 instrument (Z)
- The TSLS of Y on D identifies (Kirkeboen et al. 2016, Mountjoy 2022):

$$
\begin{aligned}
\beta_{IA} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]} &= \mathbb{E}[Y(1,R) - Y(0,R)|T_{J(i)} = C] \\
&= \pi_0 LATE[Y(1,R) - Y(0,R)|R(1) = 0, T_{J(i)} = C] \\
&\quad + \pi_1 LATE[Y(1,R) - Y(0,R)|R(1) = 1, T_{J(i)} = C]
\end{aligned}
$$

where:

- $T_{J(i)} = C := D(1) > D(0)$
- π_i are complier shares

Relatives who **stay behind**, Relatives **followers**

Decomposing LATE into sub-LATEs

Back

- 2 endogenous variables (D and R) and only 1 instrument (Z)
- The TSLS of Y on D identifies (Kirkeboen et al. 2016, Mountjoy 2022):

$$\beta_{IA} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]} = \mathbb{E}[Y(1,R) - Y(0,R)|T_{J(i)} = C]$$
$$= \pi_0 LATE_0 + \pi_1 LATE_1$$

where:

- ▷ $T_{J(i)} = C := D(1) > D(0)$
- ▷ π_i are complier shares

Relatives who **stay behind**, Relatives **followers**

Estimation: No point identification. Estimate sharp bounds (Horowitz and Manski 1995, Dutz et al. 2024). More details are provided on the next slides ("Bounding Sub-LATEs").

Bounding Sub-LATEs

Back

- Consider $LATE_0 = \mathbb{E}[Y(1,0) - Y(0,0) | R(1) = R(0) = 0, T_{J(i)} = C]$

Bounding Sub-LATEs

Back

- Consider $LATE_0 = \mathbb{E}[Y(1,0) - Y(0,0) | R(1) = R(0) = 0, T_{J(i)} = C]$
- Under assumptions:
  - $\mathbb{E}[Y(1,0)|\cdot]$ identified via carefully chosen Wald estimand
  - $\mathbb{E}[Y(0,0)|\cdot]$ **not point identified** but can be bounded

Bounding Sub-LATEs

Back

- Consider $LATE_0 = \mathbb{E}[Y(1,0) - Y(0,0)|R(1) = R(0) = 0, T_{J(i)} = C]$
- Under assumptions:
  - $\mathbb{E}[Y(1,0)|\cdot]$ identified via carefully chosen Wald estimand
  - $\mathbb{E}[Y(0,0)|\cdot]$ **not point identified** but can be bounded

Identified:

$$\mathbb{E}[Y(1,0)|R(1) = R(0) = 0, T_{J(i)} = C] = \frac{\mathbb{E}[YD(1-R)|Z=1] - \mathbb{E}[YD(1-R)|Z=0]}{\mathbb{E}[D(1-R)|Z=1] - \mathbb{E}[D(1-R)|Z=0]}$$

Bounding Sub-LATEs

Back

- Consider $LATE_0 = \mathbb{E}[Y(1,0) - Y(0,0) | R(1) = R(0) = 0, T_{J(i)} = C]$
- Under assumptions:
  - $\mathbb{E}[Y(1,0)|\cdot]$ identified via carefully chosen Wald estimand
  - $\mathbb{E}[Y(0,0)|\cdot]$ **not point identified** but can be bounded

Identified:

$$\mathbb{E}[Y(0,0)|R(0) = 0, T_{J(i)} = C] = \frac{\mathbb{E}[Y(1-D)(1-R)|Z=1] - \mathbb{E}[Y(1-D)(1-R)|Z=0]}{\mathbb{E}[(1-D)(1-R)|Z=1] - \mathbb{E}[(1-D)(1-R)|Z=0]}$$

Notice groups $\big(R(1) = R(0) = 0\big) \subset \big(R(0) = 0\big)$

Bounding Sub-LATEs

Back

- Consider $LATE_0 = \mathbb{E}[Y(1,0) - Y(0,0) | R(1) = R(0) = 0, T_{J(i)} = C]$
- Under assumptions:
  - $\mathbb{E}[Y(1,0)|\cdot]$ identified via carefully chosen Wald estimand
  - $\mathbb{E}[Y(0,0)|\cdot]$ **not point identified** but can be bounded
- Bounding $E[Y(0,0)|\cdot]$:

$$\underbrace{\mathbb{E}[Y(0,0)|R(0) = 0, T_{J(i)} = C]}_{\text{known}} = \underbrace{\omega_{R(0)=0}}_{\text{known}} \underbrace{\mathbb{E}[Y(0,0)|R(1) = R(0) = 0, T_{J(i)} = C]}_{\text{target}}$$
$$+ \underbrace{(1 - \omega_{R(0)=0})}_{\text{known}} \underbrace{\mathbb{E}[Y(0,0)|R(1) > R(0), T_{J(i)} = C]}_{\text{unknown but bounded}}$$

With $\omega_{R(0)=0}$ point identified

Bounding Sub-LATEs

Back

- Consider $LATE_0 = \mathbb{E}[Y(1,0) - Y(0,0) | R(1) = R(0) = 0, T_{J(i)} = C]$
- Under assumptions:
  - $\mathbb{E}[Y(1,0)|\cdot]$ identified via carefully chosen Wald estimand
  - $\mathbb{E}[Y(0,0)|\cdot]$ **not point identified** but can be bounded
- Bounding $E[Y(0,0)|\cdot]$:

$$\underbrace{\mathbb{E}[Y(0,0)|R(0) = 0, T_{J(i)} = C]}_{\text{known}} = \underbrace{\omega_{R(0)=0}}_{\text{known}} \underbrace{\mathbb{E}[Y(0,0)|R(1) = R(0) = 0, T_{J(i)} = C]}_{\text{target}}$$
$$+ \underbrace{(1 - \omega_{R(0)=0})}_{\text{known}} \underbrace{\mathbb{E}[Y(0,0)|R(1) > R(0), T_{J(i)} = C]}_{\text{unknown but bounded}}$$

- Dutz et al. (2024) bound target parameter using trimming arguments from Horowitz and Manski (1995)
- *Intuition:* The target mean is *smallest* if the target group occupies the **bottom** ω-fraction of the mixture distribution, and *largest* if it occupies the **top** ω-fraction

Bounding Sub-LATEs

Back

- Consider $LATE_0 = \mathbb{E}[Y(1,0) - Y(0,0) | R(1) = R(0) = 0, T_{J(i)} = c]$
- Under assumptions:
  - $\mathbb{E}[Y(1,0)|\cdot]$ identified via carefully chosen Wald estimand
  - $\mathbb{E}[Y(0,0)|\cdot]$ **not point identified** but can be bounded
- Bounding $E[Y(0,0)|\cdot]$:

$$\underbrace{\mathbb{E}[Y(0,0)|R(0) = 0, T_{J(i)} = c]}_{\text{known}} = \underbrace{\omega_{R(0)=0}}_{\text{known}} \underbrace{\mathbb{E}[Y(0,0)|R(1) = R(0) = 0, T_{J(i)} = c]}_{\text{target}}$$

$$+ \underbrace{(1 - \omega_{R(0)=0})}_{\text{known}} \underbrace{\mathbb{E}[Y(0,0)|R(1) > R(0), T_{J(i)} = c]}_{\text{unknown but bounded}}$$

- Dutz et al. (2024) bound target parameter using trimming arguments from Horowitz and Manski (1995)
- *Formally:* For binary Y, with identified mixture mean μ:

$$\max\left\{0, \frac{\mu-(1-\omega)}{\omega}\right\} \leq \mathbb{E}[Y(0,0)|R(1) = R(0) = 0, T_{J(i)} = c] \leq \min\left\{1, \frac{\mu}{\omega}\right\}$$

Bounding Sub-LATEs

Back

- Consider $LATE_0 = \mathbb{E}[Y(1,0) - Y(0,0) | R(1) = R(0) = 0, T_{J(i)} = c]$
- Under assumptions:
  - ▷ $\mathbb{E}[Y(1,0)|\cdot]$ identified via carefully chosen Wald estimand
  - ▷ $\mathbb{E}[Y(0,0)|\cdot]$ **not point identified** but can be bounded
- Bounding $E[Y(0,0)|\cdot]$:

$$\underbrace{\mathbb{E}[Y(0,0)|R(0) = 0, T_{J(i)} = c]}_{\text{known}} = \underbrace{\omega_{R(0)=0}}_{\text{known}} \underbrace{\mathbb{E}[Y(0,0)|R(1) = R(0) = 0, T_{J(i)} = c]}_{\text{target}}$$

$$+ \underbrace{(1 - \omega_{R(0)=0})}_{\text{known}} \underbrace{\mathbb{E}[Y(0,0)|R(1) > R(0), T_{J(i)} = c]}_{\text{unknown but bounded}}$$

- Dutz et al. (2024) bound target parameter using trimming arguments from Horowitz and Manski (1995)
- *Formally:* For binary Y, with identified mixture mean μ:

$$\max\left\{0, \frac{\mu-(1-\omega)}{\omega}\right\} \leq \mathbb{E}[Y(0,0)|R(1) = R(0) = 0, T_{J(i)} = c] \leq \min\left\{1, \frac{\mu}{\omega}\right\}$$

- Estimate endpoints via functions of TSLS estimates of the relevant Wald estimands

Appendix: Decomposing Substitution Effects by Pre-lottery Employment Status

Decomposing the Formal / Informal Employment Effects

Back

Winning lowers relatives' formal employment (-0.102) and raises informal ($+0.142$). *Where?* flows out of each baseline state are:

Baseline sector ($t = -2$)	Share	$\to$ Formal at $t = 4$ LATE (s.e.)	$\to$ Informal at $t = 4$ LATE (s.e.)
Formal	0.26	-0.057 (0.041)	$+0.043$ (0.036)
Informal	0.29	-0.010 (0.021)	$+0.095$ (0.060)
Both	0.03	-0.002 (0.025)	$+0.033$ (0.026)
Unemployed	0.42	-0.033 (0.034)	-0.030 (0.033)
Total (level)	1.00	-0.102 (0.053)	$+0.142$ (0.077)

Note: Pooled (4 cities); nearby relatives ($\leq$ 3 km). Each cell $=$ LATE on $\mathbf{1}[\text{baseline sector}] \times \mathbf{1}[\text{sector at } t = 4]$. "Both" $=$ formal *and* informal at baseline.

Evidence for switching: previously employed individuals are **replacing** formal with informal work

Stayers vs Followers: a Separation-Channel Effect

Back

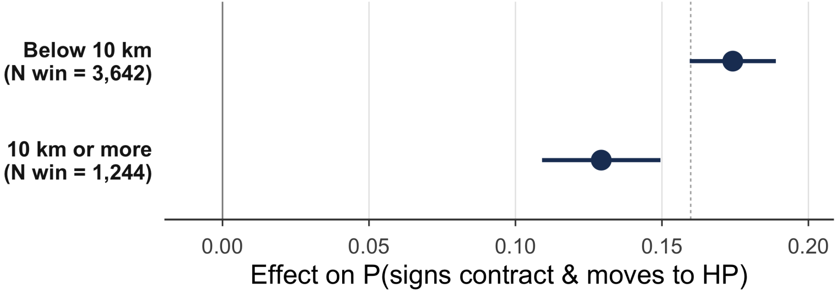
Splitting each employment flow into relatives **left behind** ($R=0$) vs **followers** ($R=1$)

	$\to$ Formal at $t=4$		$\to$ Informal at $t=4$	
Baseline sector ($t=-2$)	$LATE_0$ (stay)	$LATE_1$ (follow)	$LATE_0$ (stay)	$LATE_1$ (follow)
Formal-only	[−0.247, −0.022]	[−0.595, 0.132]	[0.002, 0.084]	[−0.082, 0.181]
Informal-only	[−0.003, 0.047]	[−0.171, −0.010]	[0.263, 0.300]	[−0.107, 0.012]
Both	[−0.021, 0.021]	[−0.059, 0.076]	[0.028, 0.037]	[0.055, 0.085]
Unemployed	[−0.033, 0.038]	[−0.205, 0.023]	[−0.045, 0.055]	[−0.195, 0.127]
Total (level)	[−0.304, −0.079]	[−0.505, 0.221]	[0.248, 0.473]	[−0.322, 0.405]

Appendix: Relocation Choices by Baseline Distance to HP

Candidates’ Relocation Choices, by Distance from the HP

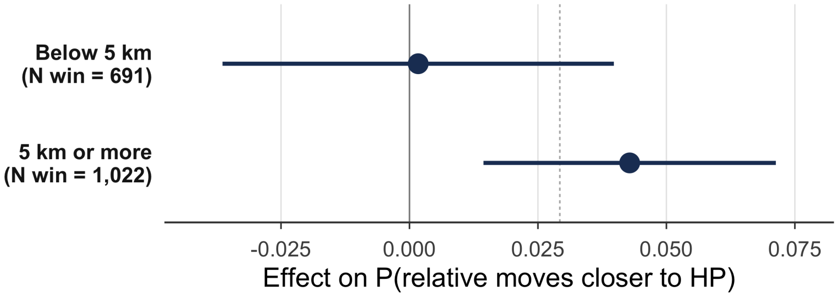
back



Note: Effect of the lottery win on P(signs contract & moves to HP), splitting candidates below vs. at-or-above 10 km from the HP. Dashed line = full-sample estimate.

Location Choices of Relatives, by Distance from the HP

back



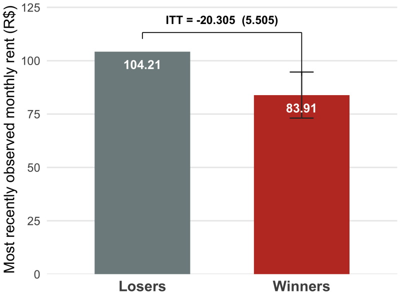
Note: Effect of the lottery win on P(relative moves closer to HP), splitting relatives below vs. at-or-above 15 km from the HP. Dashed line = full-sample estimate. (ITT)

Appendix: Effect of Winning on Rent and Neighborhood Quality

Effect on Rent (Rio): Candidates

Back

Outcome: **most recently observed monthly rent** up to $t=4$ (last non-missing value carried forward across gap years; an observed rent of 0 (a non-renter) is kept).



Note: “Winners” bar $=$ loser mean $+$ ITT; whiskers $=$ 95% CI. FE $=$ lottery $\times$ strata; s.e. clustered by candidate. N_0/N_1: candidates 517,127/1,091. Z $=$ winner_reserve.

Winning cuts candidates’ rent by $\sim$R$20

Winning moves candidates to poorer neighborhoods

Back

Neighborhood of residence four years after the lottery (CEP 5-digit).

Appendix: Placebo – Effect on Neighbors' Location Choices

Effects on **neighbors'** location choices

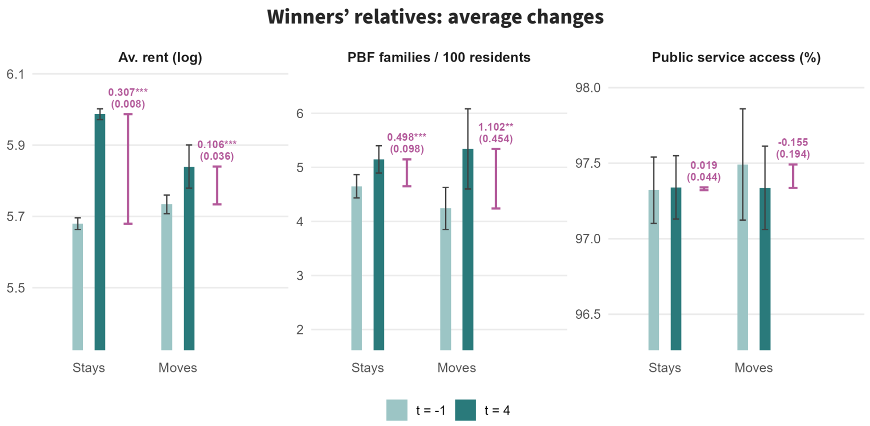
Back

	Loser mean	Estimate	(s.e.)
Variable	(1)	(2)	(3)
Neighbor has post-address	0.915	0.003	(0.005)
Pair has post-address	0.898	0.011*	(0.006)
Moves from baseline	0.232	0.000	(0.012)
Moves closer to HP	0.065	0.006	(0.013)
Distance Moved (uncond)	1.985	0.167	(0.194)
Distance to Candidate	4.247	2.654*	(1.362)
N pairs	2,106,307 / 15,860		
N candidates	72,503 / 741		
N neighbors	179,003 / 9,626		

Appendix: Relatives' neighborhood quality

Followers relocate in poorer neighborhoods

Back



Relatives of winners living nearby at baseline, split by observed response (moves = moved toward the candidate's project $R = 1$); 90% CIs, SEs clustered by candidate.

Appendix: Relatives' Labor Market Responses (Broader Set of Outcomes)

Labor Market Response of Relatives Nearby (Full Set)

				Complier			
	Non-taker	RF	OLS	Untreated	LATE	*LATE*$_0$	*LATE*$_1$
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any employment	0.607	0.003	-0.042*	0.604	-0.016	[-0.063, 0.16]	[-0.238, 0.482]
	(0.001)	(0.012)	(0.023)	(0.065)	(0.069)		
Any informal employment	0.354	0.030**	0.005	0.221	0.141*	[0.214, 0.437]	[-0.325, 0.396]
	(0.001)	(0.013)	(0.026)	(0.071)	(0.077)		
Any formal employment	0.299	-0.017*	-0.057***	0.417	-0.102*	[-0.215, -0.008]	[-0.476, 0.244]
	(0.001)	(0.009)	(0.019)	(0.054)	(0.053)		
Total formal earnings (uncond.)	1453.815	-93.676	-307.356**	2033.139	-568.099		
	(4.658)	(69.034)	(154.073)	(386.873)	(402.431)		
Number of jobs (uncond.)	1.301	-0.071	-0.265**	1.749	-0.473	[-2.263, -0.783]	[-0.088, 4.689]
	(0.003)	(0.050)	(0.106)	(0.293)	(0.291)		
Kept same formal job (uncond.)	0.020	-0.004	-0.013	0.045	-0.021	[-0.001, 0.028]	[-0.095, -0.002]
	(0.000)	(0.004)	(0.009)	(0.025)	(0.026)		
Has new formal job (uncond.)	0.233	-0.015	-0.039*	0.336	-0.085	[-0.198, 0.025]	[-0.486, 0.234]
	(0.001)	(0.010)	(0.020)	(0.056)	(0.057)		
Receives PBF	0.427	0.019	0.012	0.340	0.084	[0.069, 0.292]	[-0.17, 0.55]
	(0.001)	(0.015)	(0.029)	(0.080)	(0.087)		

Back to overall LATEs Back to sub-LATEs

Appendix: More on Selection Model

Selection Model

Location choices:

$$D_{J(i)} = \mathbb{1}\{\mu_c(X_{J(i)}, Z_{J(i)}) + \nu_{cJ(i)} > 0\}$$
$$R_i = \mathbb{1}\{\mu_n(X_i, D_{J(i)}) + \nu_{ni} > 0\}$$

Selection Model

Location choices:

$$D_{J(i)} = \mathbb{1}\{\mu_c(X_{J(i)}, Z_{J(i)}) + \nu_{cJ(i)} > 0\}$$
$$R_i = \mathbb{1}\{\mu_n(X_i, D_{J(i)}) + \nu_{ni} > 0\}$$

- Monotonicity: $\mu_c(X_{J(i)}, 1) > \mu_c(X_{J(i)}, 0)$ and $\mu_n(X_i, 1) > \mu_n(X_i, 0)$
- Exclusion: $\mu_n(X_{J(i)}, D_i)$ not a function of Z_i

Back

Selection Model

Location choices:

$$D_{J(i)} = \mathbb{1}\{\mu_c(X_{J(i)}, Z_{J(i)}) + \nu_{cJ(i)} > 0\}$$
$$R_i = \mathbb{1}\{\mu_n(X_i, D_{J(i)}) + \nu_{ni} > 0\}$$

- $(\nu_{cJ(i)}, \nu_{ni})$ are unobserved tastes, moving costs, and other constraints to mobility

$$(\nu_{cJ(i)}, \nu_{ni}) | X_i, Z_{J(i)} \sim N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

Back

Selection Model

Location choices:

$$D_{J(i)} = \mathbb{1}\{\mu_c(X_{J(i)}, Z_{J(i)}) + \nu_{cJ(i)} > 0\}$$
$$R_i = \mathbb{1}\{\mu_n(X_i, D_{J(i)}) + \nu_{ni} > 0\}$$

- $(\nu_{cJ(i)}, \nu_{ni})$ are unobserved tastes, moving costs, and other constraints to mobility

$$(\nu_{cJ(i)}, \nu_{ni}) | X_i, Z_{J(i)} \sim N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

Potential labor market outcomes:

$$E[Y(d, r) | X, Z, \nu_c, \nu_n] = \psi_{(d,r)}(X) + \beta_{(d,r)c}\nu_c + \beta_{(d,r)n}\nu_n$$

- (Roy) Those who are more likely to move benefit from moving:

$$\beta_{(1,0)n} < \beta_{(1,1)n}, \quad \beta_{(0,0)n} < \beta_{(0,1)n}$$

Back

Identification of Selection Model

Back

1. **Difference out ψ_{dr} using the instrument.** Differencing across Z within the cell:

$$\Delta_Z\, \mathbb{E}[Y \mid X, d, r] = \beta_{(d,r)c}\, \Delta_Z \lambda_c(X, d, r) + \beta_{(d,r)n}\, \Delta_Z \lambda_n(X, d, r).$$

   The lottery shifts only the cell's selection terms (relatives of compliers switching locations).

2. **Identify the selection slopes from variation in X.** For every two covariate values x_1, x_2 we have a 2×2 system

$$\begin{bmatrix} \Delta_Z\, \mathbb{E}[Y \mid x_1, d, r] \\ \Delta_Z\, \mathbb{E}[Y \mid x_2, d, r] \end{bmatrix} = \mathcal{M}_{dr}(x_1, x_2) \begin{bmatrix} \beta_{(d,r)c} \\ \beta_{(d,r)n} \end{bmatrix},$$

   so $(\beta_{(d,r)c}, \beta_{(d,r)n})$ is point-identified whenever $\det \mathcal{M}_{dr}(x_1, x_2) \neq 0$.

Key restriction (a la (Kline and Walters 2016)): selection on unobservables acts "the same way" for every X; the $X \times Z$ interaction provides the identifying variation.

Appendix: Toy model

Framework: The Two Values of Family Proximity

A relative chooses consumption c, leisure ℓ, and market hours h, taking proximity to the candidate, $P \in \{0, 1\}$, as given (à la Becker, 1965)

- Preferences: $u(c, \ell) + \theta \mathbf{P}$;
  ◇ $\theta \geq 0$: direct utility from proximity (companionship — **amenity** channel)
- Time constraint: $h + \ell \leq T + \tau \mathbf{P}$;
  ◇ $\tau \geq 0$: time transferred by the nearby candidate (childcare, help at home — **production** channel)
- Budget constraint: $c = wh + y \quad \Rightarrow c + w\ell = wT + y + w\tau \mathbf{P}$

Framework: The Two Values of Family Proximity

A relative chooses consumption c, leisure ℓ, and market hours h, taking proximity to the candidate, $P \in \{0, 1\}$, as given (à la Becker, 1965)

- Preferences: $u(c, \ell) + \theta \mathbf{P}$;
  ◇ $\theta \geq 0$: direct utility from proximity (companionship — **amenity** channel)
- Time constraint: $h + \ell \leq T + \tau \mathbf{P}$;
  ◇ $\tau \geq 0$: time transferred by the nearby candidate (childcare, help at home — **production** channel)
- Budget constraint: $c = wh + y \quad \Rightarrow c + w\ell = wT + y + w\tau \mathbf{P}$

Value of proximity (envelope theorem; $\lambda \equiv u_c$ is the marginal utility of income):

$$\frac{1}{\lambda} \frac{\partial V}{\partial \mathbf{P}} = \underbrace{\theta / \lambda}_{\text{amenity value}} + \underbrace{w\tau}_{\text{production value}}$$

Framework: The Two Values of Family Proximity

A relative chooses consumption c, leisure ℓ, and market hours h, taking proximity to the candidate, $P \in \{0, 1\}$, as given (à la Becker, 1965)

- Preferences: $u(c, \ell) + \theta \mathbf{P}$;
  ◇ $\theta \geq 0$: direct utility from proximity (companionship — **amenity** channel)
- Time constraint: $h + \ell \leq T + \tau \mathbf{P}$;
  ◇ $\tau \geq 0$: time transferred by the nearby candidate (childcare, help at home — **production** channel)
- Budget constraint: $c = wh + y \quad \Rightarrow c + w\ell = wT + y + w\tau \mathbf{P}$

Value of proximity (envelope theorem; $\lambda \equiv u_c$ is the marginal utility of income):

$$\frac{1}{\lambda}\frac{\partial V}{\partial \mathbf{P}} = \underbrace{\theta/\lambda}_{\text{amenity value}} + \underbrace{w\tau}_{\text{production value}}$$

When does a relative follow? After candidate relocates, restoring $\mathbf{P} = 1$ requires moving at cost κ_i:

$$\text{Follow} \iff \underbrace{\theta_i/\lambda_i + w_i\tau_i}_{\text{value of proximity}} \geq \kappa_i$$

Back: What we do

Back: Sorting test

Framework: Sorting on Gains Identifies the Production Channel

Relative differ in θ_i/λ_i, $w_i\tau_i$, and κ_i. Which source of heterogeneity *selects* the followers?

The values relate **differently** to the earnings gain from proximity, $\Delta \equiv Y(1,1) - Y(1,0)$:

Framework: Sorting on Gains Identifies the Production Channel

Relative differ in θ_i/λ_i, $w_i\tau_i$, and κ_i. Which source of heterogeneity *selects* the followers?

The values relate **differently** to the earnings gain from proximity, $\Delta \equiv Y(1,1) - Y(1,0)$:

- **Production** value *is* an earnings gain: a relative with high τ values proximity more *and* earns more when near ($\Delta = w\tau$ absent income effects)

  $\Rightarrow$ production-driven choices select followers with high Δ: **positive sorting**, $E[\Delta|F] > E[\Delta|B]$ (Roy 1951; Borjas 1987)

Framework: Sorting on Gains Identifies the Production Channel

Relative differ in θ_i/λ_i, $w_i\tau_i$, and κ_i. Which source of heterogeneity *selects* the followers?

The values relate **differently** to the earnings gain from proximity, $\Delta \equiv Y(1,1) - Y(1,0)$:

- **Production** value *is* an earnings gain: a relative with high τ values proximity more *and* earns more when near ($\Delta = w\tau$ absent income effects)

  $\Rightarrow$ production-driven choices select followers with high Δ: **positive sorting**, $E[\Delta|F] > E[\Delta|B]$ (Roy 1951; Borjas 1987)

- **Amenity** value is not: a relative with high θ values proximity more, but earns *no more* when near

  $\Rightarrow$ amenity-driven choices select followers unrelated to Δ: **zero sorting** (negative, if family time crowds out work time)

Framework: Sorting on Gains Identifies the Production Channel

Relative differ in θ_i/λ_i, $w_i\tau_i$, and κ_i. Which source of heterogeneity *selects* the followers?

The values relate **differently** to the earnings gain from proximity, $\Delta \equiv Y(1,1) - Y(1,0)$:

- **Production** value *is* an earnings gain: a relative with high τ values proximity more *and* earns more when near ($\Delta = w\tau$ absent income effects)

  $\Rightarrow$ production-driven choices select followers with high Δ: **positive sorting**, $E[\Delta|F] > E[\Delta|B]$ (Roy 1951; Borjas 1987)

- **Amenity** value is not: a relative with high θ values proximity more, but earns *no more* when near

  $\Rightarrow$ amenity-driven choices select followers unrelated to Δ: **zero sorting** (negative, if family time crowds out work time)

- $\Rightarrow$ **positive sorting on Δ can only come from the production channel**
  - Empirically: $\Delta_i = Y_i(1,1) - Y_i(1,0)$; the choice index maps to the latent ν_n in the selection model, so the test is the Roy restriction $\beta_{(1,0)n} < \beta_{(1,1)n}$

Back to What We Do | Back to Sorting test

Appendix: Estimates for Betas (Selection Model)

Loadings: informal employment & earnings

Back

Control-function loadings (β_c, β_n) by (D, R) cell — distance-pair spec.

Informal employment

	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
β_c	-0.039	-0.116	0.134	-0.106
β_n	0.403	0.142	0.155	-0.343

$\Delta\beta_n = -0.498 < 0$: substitution toward informal (formal$+$/informal$-$).

Earnings (R$)

	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
β_c	-212	718	461	-219
β_n	-3967	-1510	-2999	-50

$\Delta\beta_n = +2949 > 0$: positive, but the DiD is not significant.

Loading standard errors are not in the current data export (Spec-C draws store only aggregated sub-LATEs); DiD/LATE s.e. on the main slide are the reported ones.

Identification and Model Fit

Back

- Identification based on interaction of X (lottery strata and distance from HP) with Z
  - Semiparametric restrictions on unselected potential outcome distributions (Kline and Walters 2016)
  - Allows treatment effects to vary by X parametrically
  - Main restriction: selection on unobservables works "the same way" for every value of the covariates

Identification and Model Fit

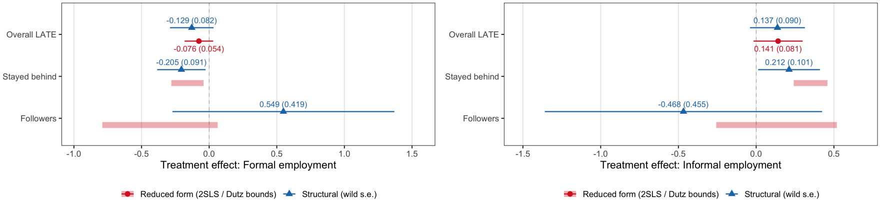
Back

- Identification based on interaction of X (lottery strata and distance from HP) with Z
- The model replicates the data for those who stayed behind (followers are noisy)

Identification and Model Fit

Back

- Identification based on interaction of X (lottery strata and distance from HP) with Z
- The model replicates the data for those who stayed behind (followers are noisy)



Related Literature

A Selection of Related Literature

- **Effects of subsidized housing programs on recipients.** MCMV: da Mata and Mation (2018), Machado and Rachter (2020a;b), Belchior et al. (2023), da Mata and Mation (2025), Leite Mariante and Ribeiro (2025). Other contexts: e.g., Jacob and Ludwig (2012), Chetty et al. (2016), Barnhardt et al. (2017), Chyn (2018), Franklin (2020), Bergman et al. (2024).
- **Migration and social networks.** Altonji and Card (1991), Munshi (2003), Beaman (2012), Munshi (2020), Stuart and Taylor (2021), Sahai and Bailey (2022), Diamond and Gaubert (2022), Blumenstock et al. (2025), Koenen and Johnston (2025)
- **Family proximity, childcare, and labor supply.** Posadas and Vidal-Fernández (2013), Compton and Pollak (2014), Clark et al. (2019), Zamarro (2020)
- **Switching between formal and informal employment in developing countries.** Maloney (1999), Bosch and Maloney (2010), Günther and Launov (2012), Meghir et al. (2015), Ulyssea (2018; 2020), Berniell et al. (2021), Gerard and Gonzaga (2021), Engbom et al. (2022), Martínez Leyva (2025), Derenoncourt et al. (2025)
- **Do movers select on labor-market gains?** Roy (1951), Borjas (1987), Chiquiar and Hanson (2005), McKenzie and Rapoport (2010), Kennan and Walker (2011), Bryan et al. (2014), Bryan and Morten (2019), Lagakos et al. (2023)
- **Methods.** Kirkeboen et al. (2016), Kline and Walters (2016), Mountjoy (2022)