Principles of Empirical Analysis

Lecture 7: Introduction to observational data and quasi-experiments

Miri Stryjan
Spring, 2024

Teacher Presentation

Miri Stryjan

- Assistant Professor in the Economics Department, Aalto BIZ.
- E-mail: miri.stryjan@aalto.fi
- Teaching (approx.) 30%, Research (approx.) 70%
- Research area: Development Economics, Empirical/applied microeconomics
- Languages: English & Svenska!

Outline for Part II

- **In this part, we move on from using experimental data to issues we face when using observational data in estimating causal effects**
  - When and under what type of assumptions can we estimate causal effects from observational data?
- **We will familiarize ourselves with the most common quasi-experimental causal inference methods**
  - Difference-in-differences (DID)
  - Regression Discontinuity Design (RDD)
  - Instrumental Variables (IV)
- **And, designs based on controlling for observable differences**

Outline for Part II

- **Homework 4**
  - You have one more ”practical” homework assignment (homework 4). It will open on Monday 5.2 and has a deadline in the last week of lectures (Feb 14).
  - There will be a related exercise session with Martta Thursday 8.2
- **Reading assignment (homework 5)**
  - For this last homework, we will give you a list of research papers to choose from and you need to answer questions about the paper.
  - This homework will be due one week after the exam.

I will communicate with you through Announcements on mycourses!

Outline for today

- **The limitations of Randomized experiments (RCTs)**
- **What is observational data** and why it is difficult to make causal claims based on observational data?
- **What are quasi-experiments?**
  - This is the theme throughout part II
- **Case study:** studying neighborhood effects with a quasi-experimental study.

Causal inference - some important terms

We want to estimate the effect of some variable, T on some outcome, Y.

T = the independent variable, the variable of interest, the ”treatment”.

Y = the dependent variable, the outcome variable.

Examples:

- **T= education Y= earnings**
- **T= marketing campaign Y= sales**
- **T= carbon tax Y= emissions**
- **T= R&D subsidy Y= innovation**
- **T= fiscal stimulus Y= unemployment**

Prelude: The Limits of RCTs

The Limitations of RCTs

Randomized experiments are often the best way the evaluate the impact of "treatments". They:

- Are simple and transparent - everyone can understand the results
- require fewer assumptions than alternative approaches (in part 2 of the course)

However, RCTs are not always feasible or desirable. They can be

- Too costly (in terms of money or time)
- Unethical
- Unhelpful when studying historical questions – we cannot go back in time and perform an experiment
- Unhelpful for understanding market level (General equilibrium) effects

In such situations, using observational data and clever design can still allow the researcher to study causal questions.

Ethical and practical limitations of RCTs

Experiments should not knowingly harm anyone

• but we still need to understand the effect of potentially harmful things

Meaningful experiments are sometimes very expensive

• on the other hand, policy and business mistakes can also be very costly

→ even large investments in experimentation can be justified

The relevant time horizon may be very long

• sometimes many decades! E.g. when studying the role of a given education on future labor market outcomes throughout adult life.

Hawthorne and John Henry Effects

- the evaluation itself may push people to change their behavior
- This is likely less of a problem with administrative data and long follow-up periods (subjects not reminded about being evaluated)

Examples of ethical limitations

Suppose we want to study the effect of… **(T variables in red)**:

- political propaganda on political views or beliefs
- fertility on women's wages.
- exposure to violence, war or pollution on human capital development

It would not be ethical to attempt to introduce random variation in these treatments, T.

But we still want to learn about the effects of theses T's.

⇒ In such cases: Find some exogenous factor that introduces random variation in the participation in / exposure to war or violence or pollution.

Fundamental limitations of RCTs

Spillovers

- the treatment also affects the control group
- → cannot use the control group to infer what would have happened without the treatment

General equilibrium (GE) effects

- the GE effects may be the main value of some treatments

- **RCTs never capture economy-wide GE effects**
- **Some examples of measuring more limited spillovers with RCTs exist**

Scarcity of potential observations

- some treatments affect entire countries or even the whole world
- we'll never have experimental designs for these treatments

Observational data

Observational data

- **When experimental designs are infeasible, researchers must resort to the use of observational data**
  - Observational data: data that is collected as part of the normal functioning of society's institutions, organizations or firms, e.g. data from administrative records, surveys, sales data, censuses…
  - Also data available online, from apps….
  - Prottoy provided some concrete examples in Lecture 6.
  - Observational study draws inferences from a sample of a population where the independent variable (variable of interest, or "treatment") is not under the control of the researcher
  - This is in contrast with experiments, such as randomized controlled trials (RCTs), where each subject is randomly assigned to a treatment group or a control group (e.g. MTO)

Observational data

- **From part 1 of the course:** we should be suspicious of correlations found in observational data. Correlations are almost certainly not reflecting a causal relationship because the variables were endogenously chosen by people who were making decisions about their lives
  - E.g. education level of the people in the (FLEED) data is not a consequence of randomization on the part of the researcher, but of optimization on part of the individuals in the data
- **This is also what Economic theory teaches us. People's decisions are not "random"! (at least not on average)**
- **Considerable challenges when estimating causal effects with observational data.**

People optimize

Consider the case with T= education Y= earnings

- **In the potential outcomes model:**
  - A treatment must be completely independent of the potential outcomes under consideration in order to measure a causal effect.
  - BUT if the person is choosing level of education based on what she thinks is best, then it necessarily violates this independence condition.
  - Economic theory predicts choices will be endogenous, and thus naive correlations are misleading!
- **Keep in mind selection bias**

Observational alternatives to experiments*

1. **Selection on observables: treatment and control groups differ from each other only w.r.t. observable characteristics**
   - multivariate regression, subclassification, matching
2. **Selection on unobservables: treatment and control groups differ from each other in unobservable characteristics**
   - Something unexpected happened that affected some people and not others in an “almost random” manner – (case study of today)
   - Treatment and controls are observed before and after treatment – difference-in-differences (DID)
   - Selection mechanism is known – regression discontinuity designs (RDD)
   - Exogenous variable induces variation in treatment – instrumental variables (IV)

*In addition, ***structural approaches*** offer an alternative to experiments. We will not cover them in this course.

Natural or quasi-experiments

- **Sometimes the researcher is "lucky" and a government policy or nature affects households (or firms etc.) in a way that resembles an experiment**
- **Such situations are referred to as "natural experiments" or "quasi-experiments"**
- **[quasi = from Latin *quasi* 'as if, almost']**

Natural or quasi-experiments

- **Broad term that refers to many different situations and require different research methods (IV, RDD, DID)**
- Sometimes the researcher is “lucky” and a government policy or nature affects households (or firms etc.) in a way that resembles an experiment
- **These instances are referred to as “natural experiments” or “quasi-experiments”**
  - Historical episodes that provide observable, “almost” random variation in treatment status
  - These might be law changes that affect some people, but not others => control and treatment groups

The rest of this lecture

- **We will discuss a situation where a policy created experiment-like condition that can be analyzed just as an experiment!**
- **We will continue to use neighborhood effects as the running example by illustrating the challenges in studying neighborhood effects without an experiment**
- **Quasi-experimental evidence on neighborhood effects**
  - Chyn, Eric. 2018. "Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children." *American Economic Review*, 108 (10): 3028-56. https://www.aeaweb.org/articles?id=10.1257/aer.20161352

Quasi-experiment and neighborhood effects

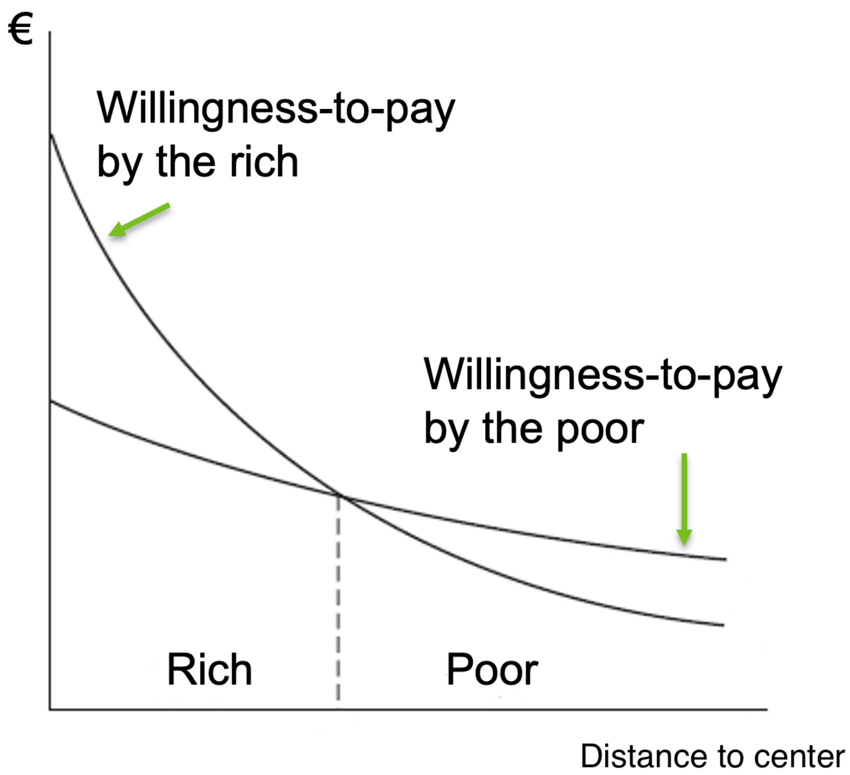
Segregation in a model city

- **Let's assume that there are two residential areas in a city with a fixed supply of housing**

Area 1: Central city	Area 2: Suburb
Historic city center with beautiful architecture, historical monuments and easy access to jobs and variety of services	Far away from city center with lower quality amenities and less services

- **Let's also assume there are only two types of households**
  - Type 1: Rich; Type 2: Poor
  - Both types work in the city center and dislike commuting

Where will the rich end up living?



With these assumptions and a free market, the city will be segregated by income:

The center is attractive for everyone -> prices will be higher in the center.

Rich live in the city center; poor cannot afford to live there, poor live in the suburbs.

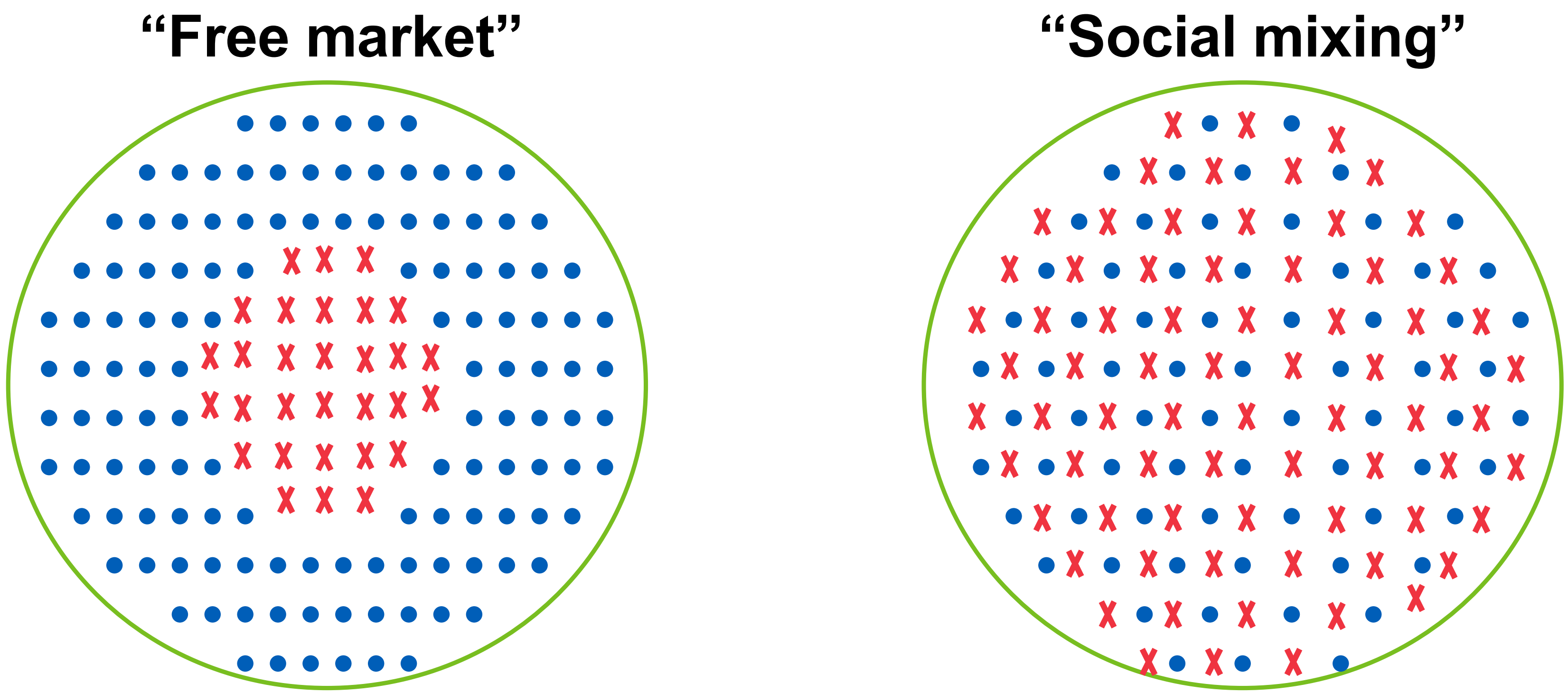
Segregation is the consequence of income inequality, quality differences of neighborhoods and optimizing behavior

Whether this is good or bad depends on neighborhood effects

Neighborhood effects

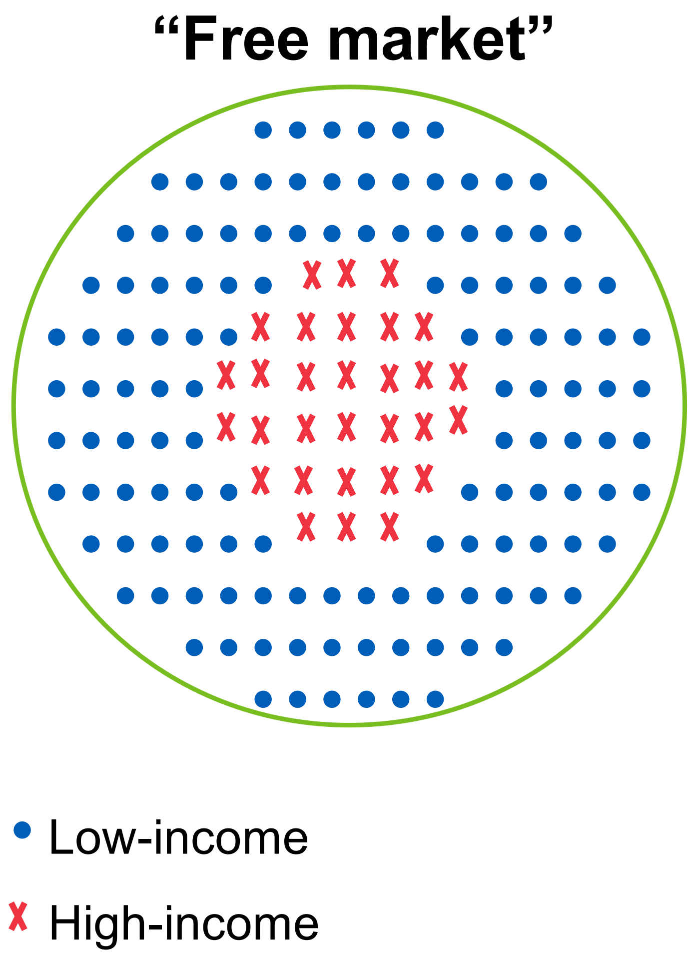
- **Idea that the neighborhood you live in can have direct and indirect effect on several socio-economic outcomes.**
  - Especially for children, the neighborhood they grow up in can matter.
- **Why do you think?**
- The neighborhood an individual grows up in can determine her peer group, education, role models and aspirations.
- Poor neighborhoods provide fewer opportunities than more affluent (richer) neighborhoods.

Which of these cities would be better for the citizens?



- Low-income
- High-income

The observed situation resembles this

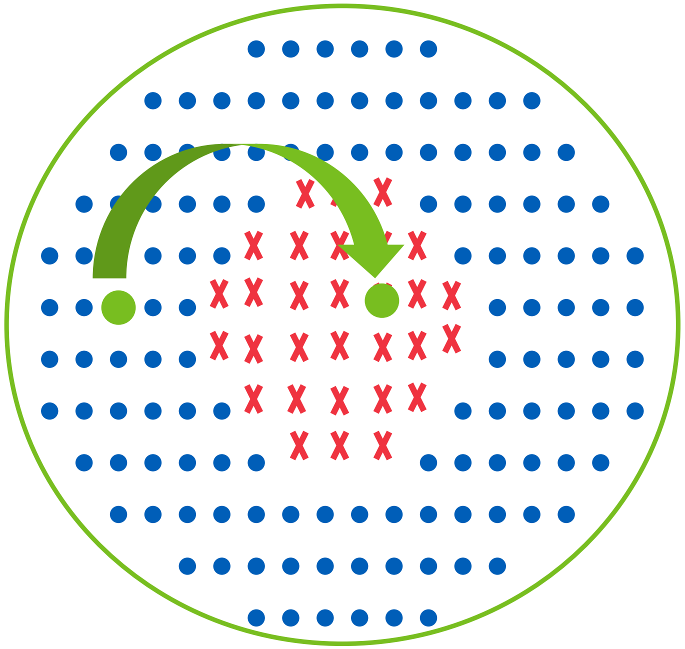


Suppose we wanted to estimate the effect of living next to high income families.

The city is segregated as in the picture.

Why is it difficult to estimate the effect using observational data on the city's residents? Who would we compare? Are there any selection problems?

One low-income family



- What if we narrow down our question and focus on a poor family that *moves* from low income to high income areas?
  - *Neighborhood quality would increase*
  - *The children would have different role models and peers*

- Question: Would the children in the family benefit if the family moved next to high-income families?

Housing market mechanism and selection bias

Discuss potential alternative factors that could affect both T and Y

Parental resources

T: Living in richer neighborhood: better nbd quality and peer group

?

Y: Child's outcomes: Education, future wages, employment

- Children who grow up in rich neighborhoods (nbd) do better later in life
- But is this just a correlation due to optimization behavior by parents or a causal effect?

Housing market mechanism and selection bias

- There is selection into neighborhood based on parental resources.
- The same resources that affect the neighborhood also affect the child outcome directly.

Parental resources

T: Living in richer neighborhood: better nbd quality and peer group

?

Y: Child's outcomes: Education, future wages, employment

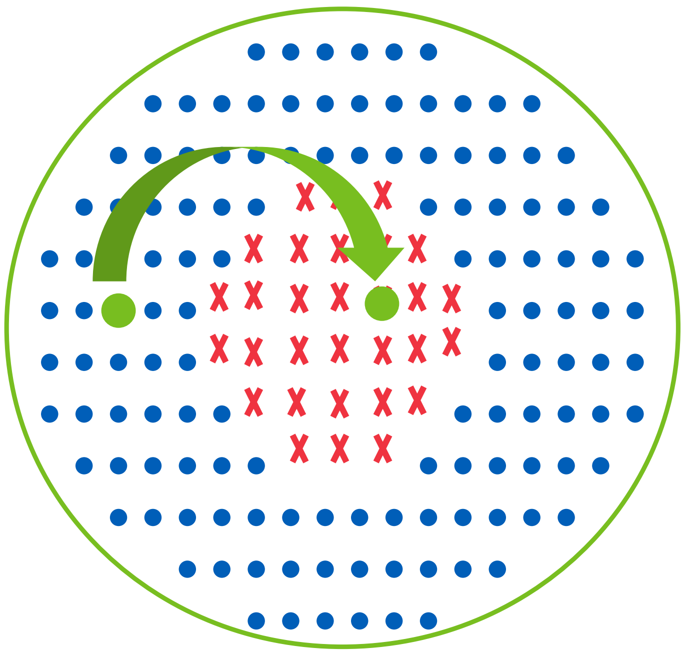
- How can we isolate the link between our "treatment" living in a rich neighborhood and our outcomes?

Controlling for observable differences?

- **One way would be to control for observable differences**
  - This would mean we can *compare people who are similar on observable and measurable characteristics:* have the same initial income, level of education etc., but live in different quality neighborhoods.
- **However, if families are assumed to be similar, why did the families make different residential location choices?**
  - low-income parents who make the effort to move to a higher quality nbd probably also use more of other resources in parenting than observably similar parents who remain in the poor nbd.
  - There are likely *Unobservable differences*

Public housing demolition as a quasi-experiment

One low-income family



- What if we *provided* one low-income family the resources to move to the other residential area?
  - *Neighborhood quality would increase*
  - *The children would have different role models and peers*

- Question: Would the children in the family benefit if the family moved next to high-income families?

Chyn (2018, AER)

American Economic Review 2018, 108(10): 3028–3056
https://doi.org/10.1257/aer.20161352

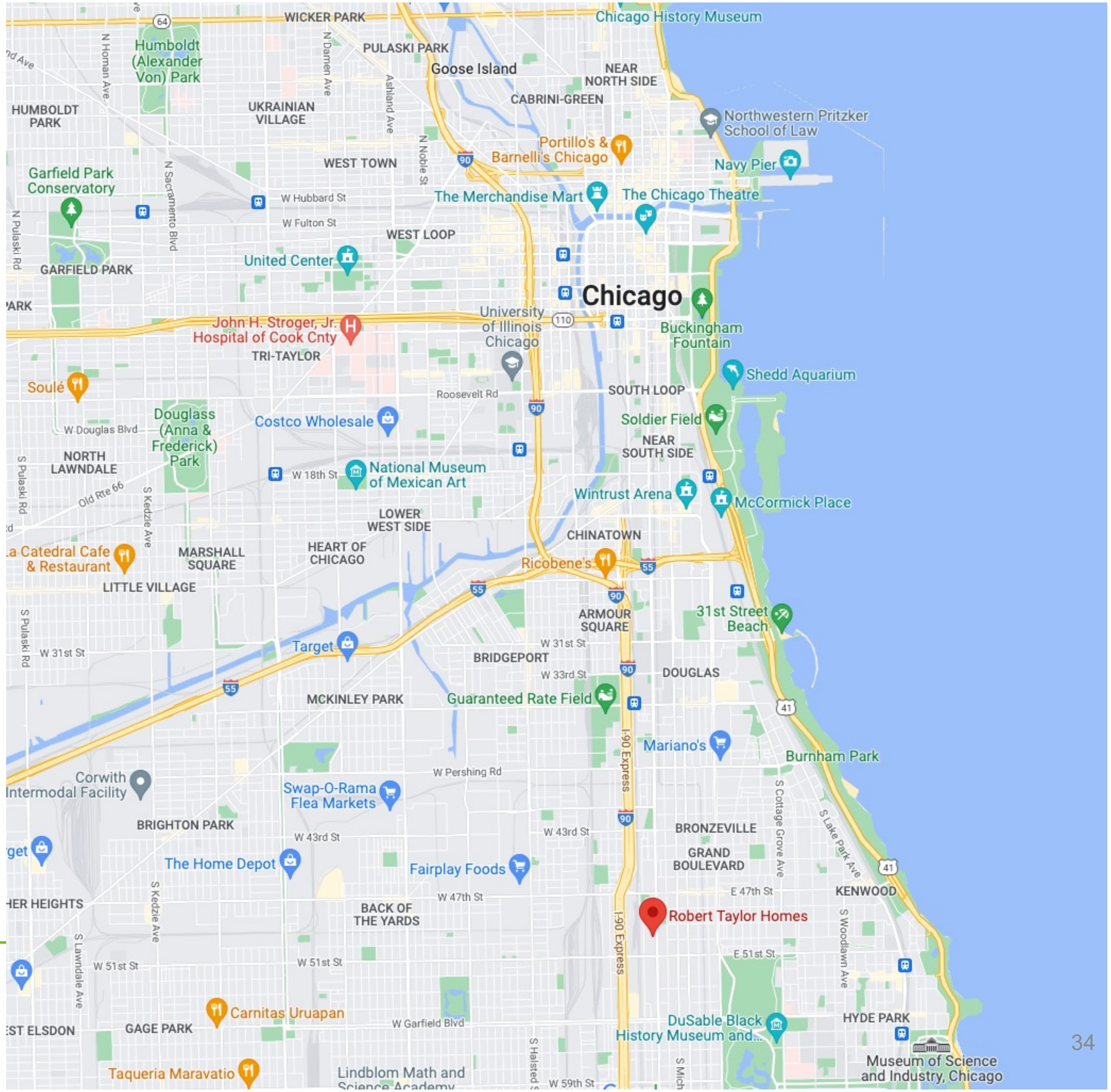
Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children†

By Eric Chyn*

This paper provides new evidence on the effects of moving out of disadvantaged neighborhoods on the long-run outcomes of children. I study public housing demolitions in Chicago, which forced low-income households to relocate to less disadvantaged neighborhoods using housing vouchers. Specifically, I compare young adult outcomes of displaced children to their peers who lived in nearby public housing that was not demolished. Displaced children are more likely to be employed and earn more in young adulthood. I also find that displaced children have fewer violent crime arrests. Children displaced at young ages have lower high school dropout rates. (*JEL* H75, I38, J13, R23, R38)

Example: Robert Taylor Homes project

Example:
Robert Taylor Homes project



Aalto University
School of Business

Chyn (2018, AER)

- **Studies the case of Chicago where the housing authority began reducing its stock of public housing during the 1990s**
  - The authority targeted some buildings with poor maintenance for demolition while leaving nearby buildings untouched
  - Residents of buildings selected for demolition received Section 8 housing vouchers and were forced to relocate
- **This policy created a treatment and a control group "naturally" or by accident**
  - The housing authority was not planning to divide the residents into control and treatment groups for research purposes
  - The researcher was not involved in creating these groups
  - Quasi-experiment!

Chyn (2018, AER)

- **Note that this is a particular type of quasi-experiment that you can analyze exactly as if it was a randomized experiment**
  - This is usually not the case!

The paper

1. **Discuss what is the treatment exactly? And how can this study give us insights different from related research?**
    - Everyone complies
    - Here the treatment is a combination of many things
    - See how much the neighborhood poverty rate changes
2. **Present the kind of observational data that is used.**
3. **Check that groups really look like they are randomized**
    - Pre-treatment covariates must be balanced across groups (balance tests)
4. **Main results**
    - & Heterogeneity w.r.t sex and age etc.

Research design

- **The research design compares the young adult outcomes of displaced and non-displaced children from the same public housing project**
  - I.e. compares the treatment and control groups
  - Displaced children are "treated"
  - Non-displayed children are "control"
  - The Treatment, or variable of interest is "being displaced"

  - **Key point:** If these two groups of children and their households were similar before the demolition, differences in later-life outcomes can be attributed to relocation to another neighborhood

Key assumption I

- **The decision about <u>which buildings to demolish</u> were unrelated to the characteristics of the tenants**
  - Why do you think this is important?
- **This assumption should be valid if the tenant selection mechanism did not allow households to self-select into buildings**
  - Within a given housing project, the households were (as-good-as) randomly assigned to buildings
  - Waiting lists: there are more applicants than housing units
  - With severe need for affordable housing and few outside options, people would choose the unit they are offered

Key assumption I

- **In this type of research design, one needs to carefully show that the households and children were similar in the control and treatment group prior to treatment (demolition)**
  - If they are similar in terms of characteristics that the researcher can observe, it is plausible that they are similar also in terms of the characteristics the researcher does not observe
  - Need to perform Balance tests!

Key assumption II

- **Demolition has no effects on the children whose building was not demolished (control group)**

- **Prior research on the same demolitions shows that crime fell in the projects after demolitions, so this condition is violated.**
  - Both the treatment and the control group might benefit from the demolition.
  - If crime in a neighborhood has adverse effects on children, Chyn's results might be biased toward zero (underestimates).
    - *This is less of a problem than if the effect would have been biased upwards!*

Data

The data that I use to test whether demolition has an impact on long-run outcomes of children is drawn from multiple administrative sources. Specifically, I combine building records from the CHA and social assistance (i.e., TANF/AFDC, Food Stamps, and Medicaid) case files (1994–1997) from the Illinois Department of Human Services (IDHS) to create a sample of children who lived in public housing and were affected by demolition during the 1990s.[19] I obtain information on baseline (prior to displacement due to demolition) characteristics and long-run outcomes by merging the sample of children with unemployment insurance wage records (1995–2009) from the Illinois Department of Employment Security (IDES), comprehensive arrest records (up to 2009) from the Illinois State Police (ISP), and IDHS assistance files (1989–2009). In complementary research, Chyn, Jacob, and Ludwig (2017) also link the sample of displaced and non-displaced children to records from the Chicago Public Schools (CPS) and the National Student Clearinghouse (NSC) to analyze schooling outcomes.[20]

Balance test

The next table presents comparisons between treatment and control groups for “baseline” variables, i.e. variables measured before the demolition.

Table 1—Comparison of Displaced and Non-Displaced Children and Adults at Baseline (*Prior to Demolition*)

	All children		Male children		Female children		Adults	
	Control mean (1)	Difference: treated–control, within estimate (2)	Control mean (3)	Difference: treated–control, within estimate (4)	Control mean (5)	Difference: treated–control, within estimate (6)	Control mean (7)	Difference: treated–control, within estimate (8)
Demographics								
Age	11.714	0.035 (0.159)	11.548	0.145 (0.196)	11.873	−0.070 (0.186)	28.851	0.810 (0.312)
Male (= 1)	0.489	−0.008 (0.017)					0.128	−0.001 (0.011)
Teen mom (= 1)†							0.371	−0.018 (0.024)
Past arrests (#)								
Violent	0.015	0.005 (0.007)	0.028	0.011 (0.014)	0.004	−0.003 (0.009)	0.185	−0.017 (0.032)
Property	0.011	0.010 (0.009)	0.018	0.015 (0.014)	0.004	0.004 (0.010)	0.156	0.016 (0.020)
Drugs	0.025	0.000 (0.013)	0.054	0.017 (0.023)	0.000	−0.018 (0.012)	0.166	0.031 (0.022)
School outcomes								
Enrolled (= 1)	0.948	0.003 (0.015)	0.946	−0.009 (0.017)	0.949	0.014 (0.016)		
Reading score (SD)	−0.443	0.024 (0.074)	−0.477	−0.045 (0.087)	−0.410	0.074 (0.074)		
Math score (SD)	−0.449	0.048 (0.061)	−0.509	0.007 (0.077)	−0.393	0.073 (0.065)		
Economic activity								
Employed (= 1)							0.173	0.006 (0.016)
Earnings‡							$1,493.75	−$45.91 (193.358)
Observations (individuals)		5,250		2,547		2,703		4,331

Point estimate

Standard error

95% confidence interval = 0.81 ±1.96*0.312

Main effects

The next tables and graphs present main effects of the treatment on the outcome variables.

Table 2—Impact of Demolition on Household Neighborhood Characteristics

	3 years after demolition		8 years after demolition	
	Control mean (1)	Difference: treated–control, within estimate (2)	Control mean (3)	Difference: treated–control, within estimate (4)
HH has address (= 1)	0.777	0.014	0.656	0.011
		(0.021)		(0.020)
Only HHs with address				
Tract characteristics:				
Black (percent)	94.897	–2.801	90.042	–1.055
		(1.125)		(1.257)
Below poverty (percent)	64.208	–14.264	40.858	–2.771
		(2.729)		(2.353)
Violent crime rate	68.855	–29.522	30.801	–2.371
		(5.807)		(4.714)
Observations (HHs)		2,767		2,767
Observations (HHs with address)		2,162		1,824

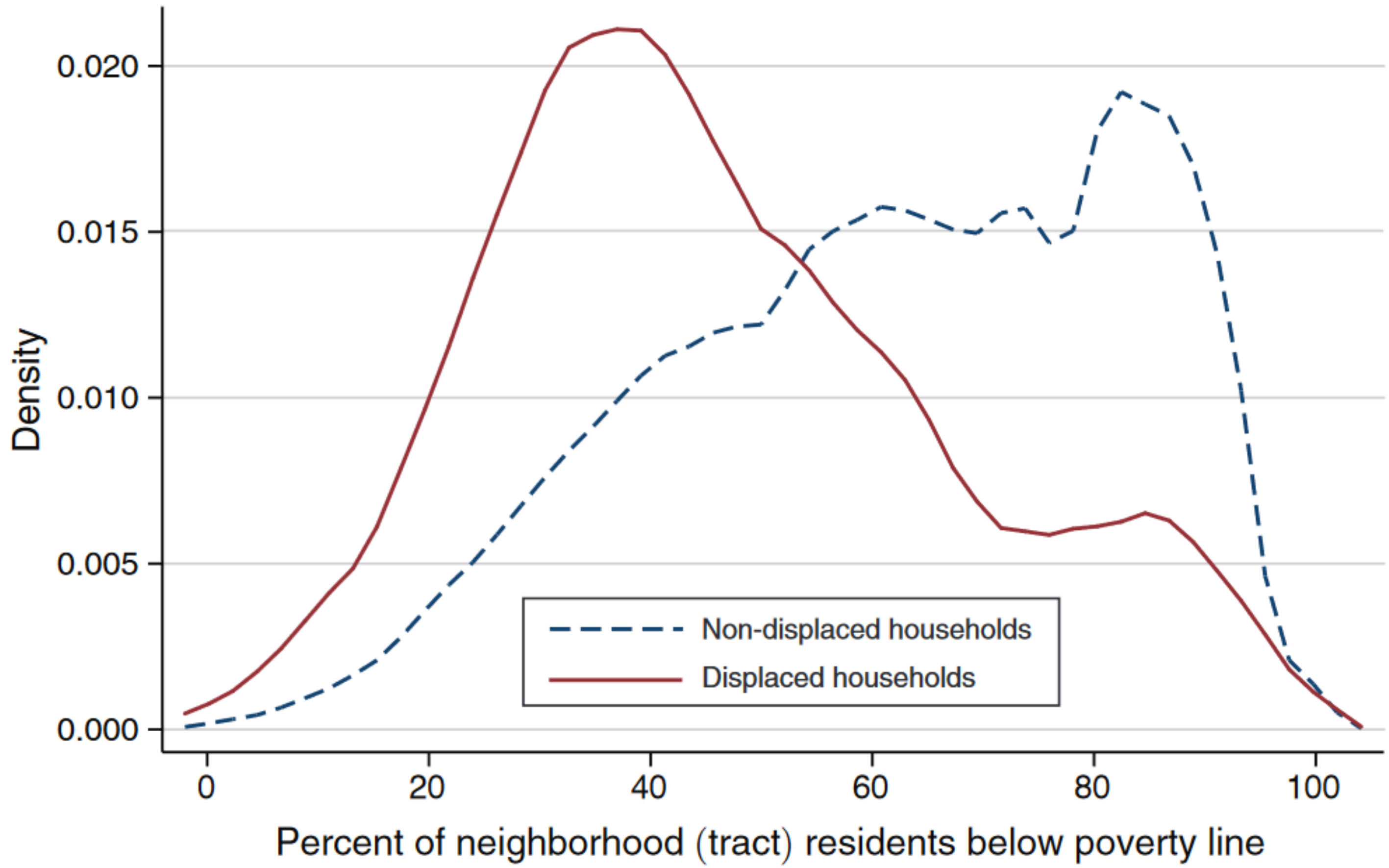


FIGURE 1. DENSITY OF NEIGHBORHOOD POVERTY AFTER DEMOLITION

Notes: The figure shows statistics for the duration-weighted average poverty rate for each household in the sample ($N = 2{,}767$). I compute the average over all locations for the household regardless of whether a child is still present.

TABLE 3—IMPACT OF DEMOLITION ON ADULT LABOR MARKET OUTCOMES OF CHILDREN

	Control mean (1)	Difference: treated–control, within estimate (2)
Employed (= 1)	0.419	0.040
		(0.014)
Employed full-time (= 1)	0.099	0.013
		(0.006)
Earnings	\$3,713.00	\$602.27
		(153.915)
Earnings (> 0)	\$8,856.91	\$587.56
		(222.595)
Observations		35,382
Individuals		5,246

Main effects: interpretation

Main effects:

Families who were replaced ended up in “better” (less poor) neighborhoods. But the effect did not persist after 8 years.

There were positive effects on employment and wages (measuread after age 19) on the children who were displaced, compared to the control group who stayed in the housing projects.

Heterogenous effects

The following tables and figures examine if there are heterogenous effects, i.e. if the effect on the outcome variables differ by subgroup.

Relevant subgroups?

- **By sex**
- **By age at the time of demolition**
- **Others?**

TABLE 4—IMPACT OF DEMOLITION ON ADULT LABOR OUTCOMES OF CHILDREN BY SEX

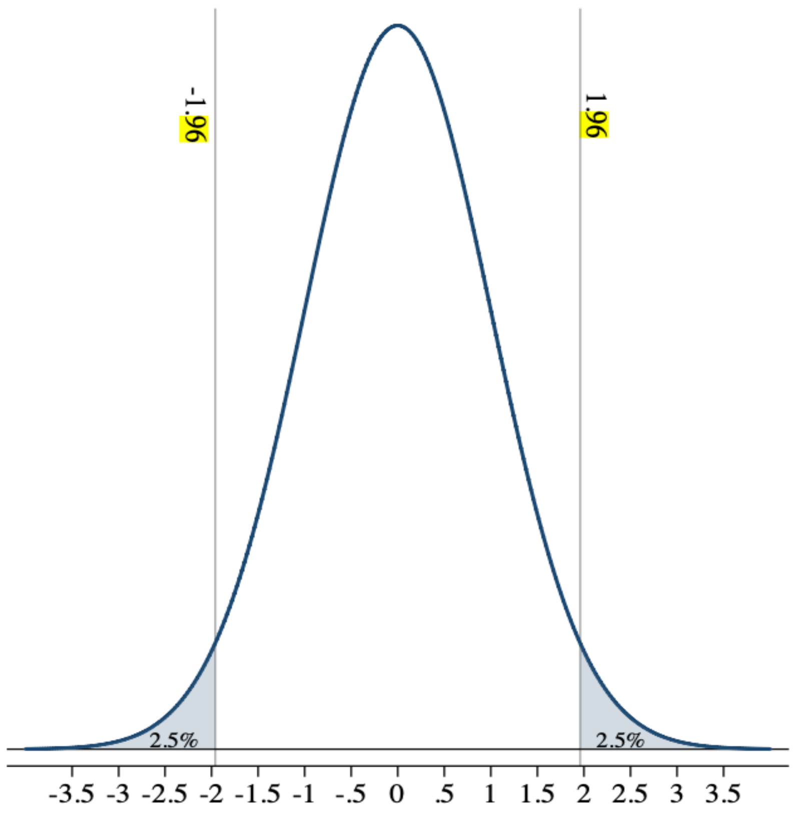
	Males		Females	
	Control mean (1)	Difference: treated–control, within estimate (2)	Control mean (3)	Difference: treated–control, within estimate (4)
Employed (= 1)	0.325	0.017 (0.019)	0.505	0.066 (0.014)
Employed FT (= 1)	0.080	0.013 (0.008)	0.117	0.015 (0.008)
Earnings	\$2,946.51	\$417.46 (236.705)	\$4,416.94	\$806.22 (188.520)
Earnings (> 0)	\$9,055.43	\$552.21 (439.299)	\$8,739.53	\$609.26 (274.111)
Observations		16,876		18,506
Individuals		2,546		2,700

Statistical significance – recall from lecture 5

Critical values and a rule-of-thumb

- Critical value is a point in the test distribution corresponding to a specific p-value
  - in large samples, a t-statistic of **1.96** corresponds to a p-value of 0.05 in a 2-sided test

→ A common rule-of-thumb is to call a result "statistically significant" if the point estimate is at least twice as large as its standard error



Statistical significance

- **We typically call estimates ”statistically significant” if $p < .05$**
- **i.e. if there was no effect, differences as extreme as the one we observed between treatment and control would occur less than 1 out of 20 times**

Let’s discuss significance and the 95% CI in the context of Table 4.

The point estimate for the treat-control difference is 0.066.

The standard error of this point estimate is 0.014.

In other words, the lower and upper boundaries of the 95% confidence interval for this estimate are

0.066 - (1.96*0.014)= 0.066-0.0274 = **0.03856**

0.066 + (1.96*0.014)= 0.066+0.0274 = **0.09344**

Since the entire CI lies above 0, we can conclude that there is a “significant” effect different from 0.

Discussion of Table 4

This is the female columns from prev page

Among females in the control group (kids who did not get displaced) the average employment is just over 50%.

The treatment effect (discussed more on the right side of the table) is 6.6%. If we compare this effect to the control group mean, we see that due to the treatment, the likelihood to be employed increased by 0.066/0.505 = approximately 13%

Females	
Control mean (3)	Difference: treated–control, within estimate (4)
0.505	0.066
	(0.014)
0.117	0.015
	(0.008)
\$4,416.94	\$806.22
	(188.520)
\$8,739.53	\$609.26
	(274.111)
	18,506
	2,700

The point estimate for the treat-control difference is 0.066, i.e. in the treated (displaced) group, the employment rate is 6.6 ppt. higher.

The standard error of this point estimate is 0.014.

In other words, the difference between treatment and control females is significant at the 5% level since

0.066 - (1.96*0.014)= 0.066-0.0274 >0

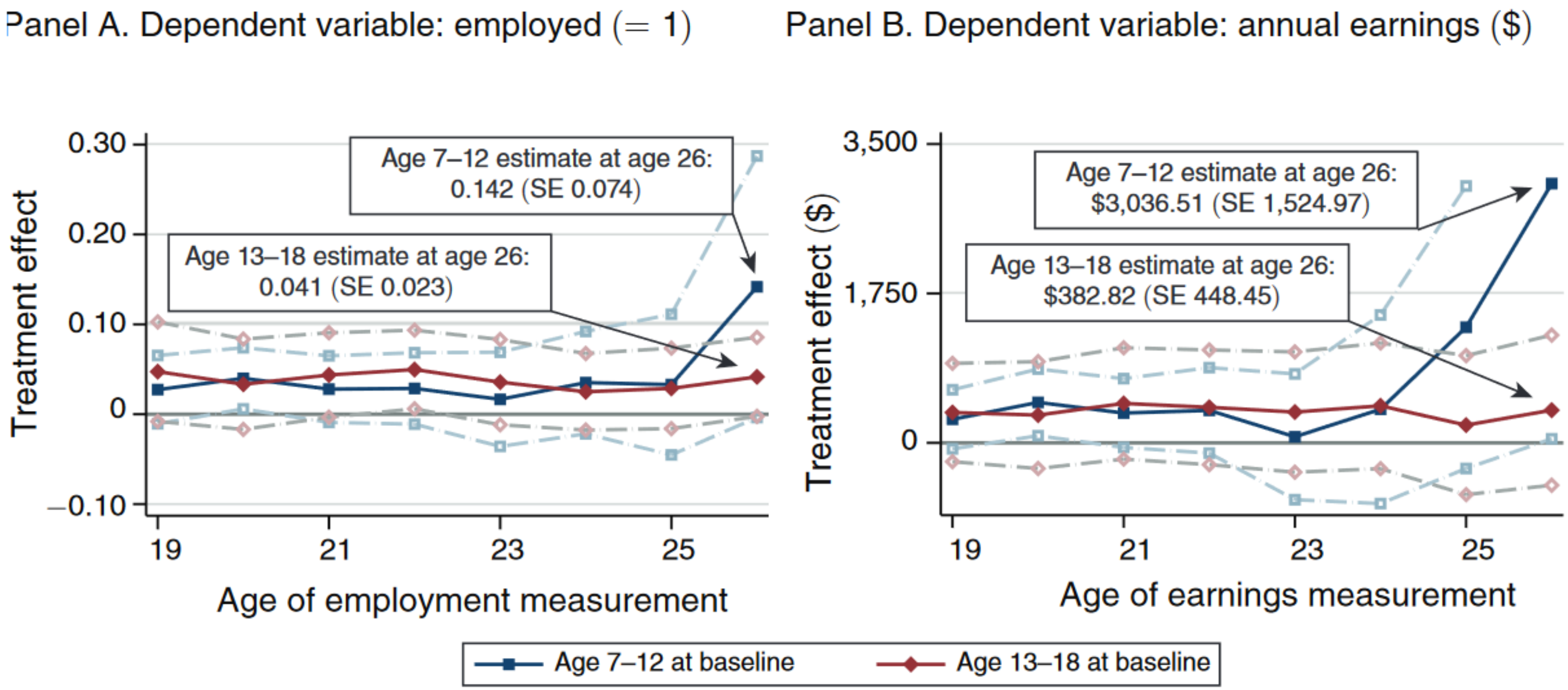


Figure 2. Impact on Employment and Earnings by Age of Measurement

Notes: Each point on the figure is an estimate from the following model:

$$Y_{it} = \sum_{j=19}^{26} \alpha_j D_{i,b} \mathbf{1}(age_{i,t} = j) + X_i'\theta + \psi_p + \delta_t + \epsilon_{it},$$

where i, t, b, and p index individuals, years, buildings, and projects, respectively. See Section IV for further details.

For comparison: Chetty et al., 2016 (MTO) found:

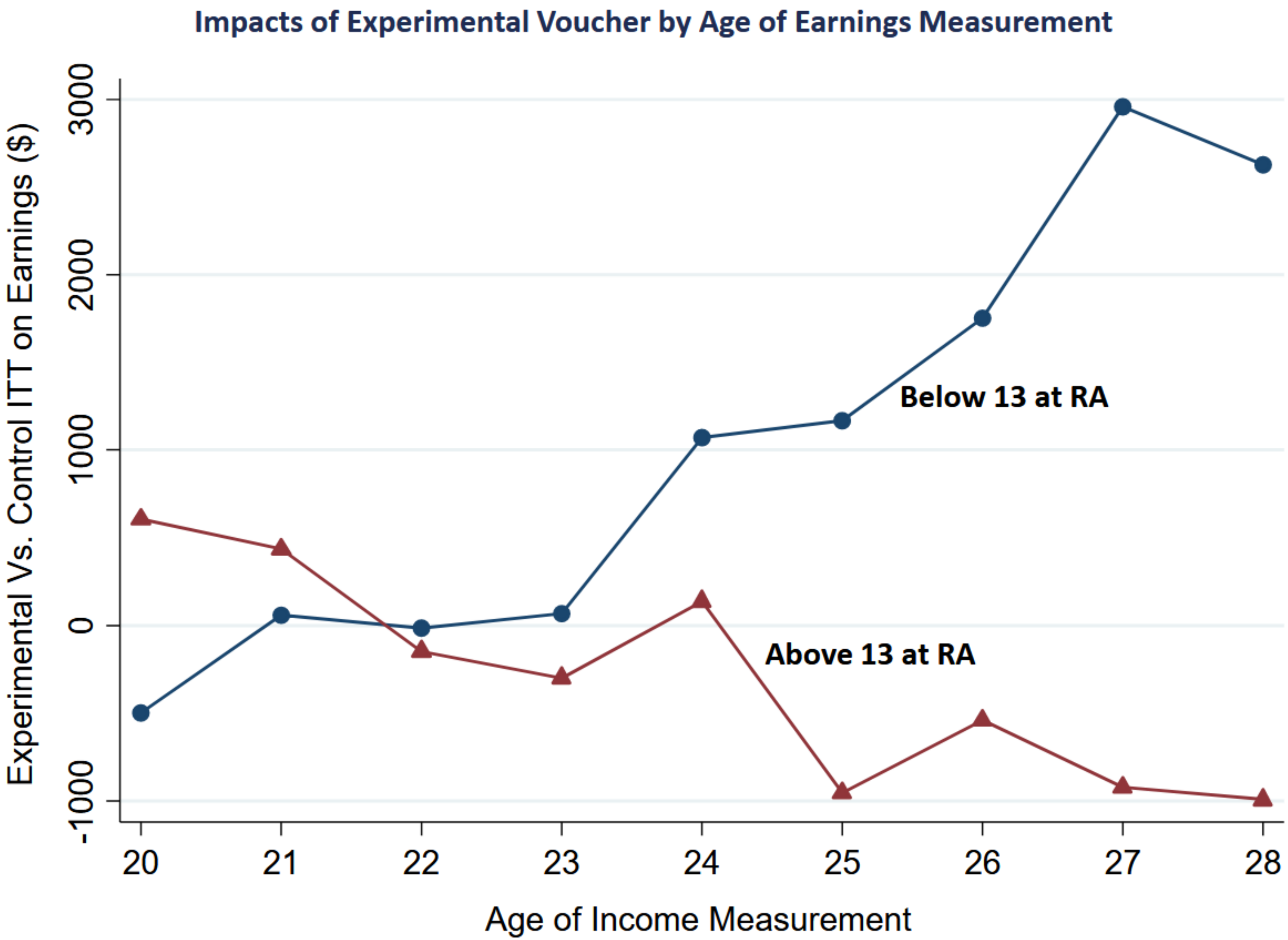


TABLE 5—IMPACT OF DEMOLITION ON CRIME OF CHILDREN

	All		Males		Females	
	Control mean (1)	Difference: treated–control, within estimate (2)	Control mean (3)	Difference: treated–control, within estimate (4)	Control mean (5)	Difference: treated–control, within estimate (6)
Number of arrests						
Violent	0.072	–0.010	0.106	–0.017	0.039	–0.004
		(0.004)		(0.006)		(0.005)
Property	0.034	0.006	0.041	0.009	0.028	0.003
		(0.003)		(0.006)		(0.003)
Drug	0.103	–0.005	0.193	–0.016	0.018	0.005
		(0.011)		(0.018)		(0.008)
Other	0.154	–0.25	0.268	–0.037	0.046	–0.014
		(0.011)		(0.015)		(0.008)
Observations		56,629		27,246		29,383
Individuals		5,250		2,547		2,703

Notes: The control mean statistic in column 1 refers to averages for non-displaced individuals. The mean difference between displaced and non-displaced children in columns 2, 4, and 6 are computed from the regression specified in equation (1). Robust standard errors are clustered at the public housing building level.

Discussion I

The paper aims to estimate the effect on poor families of moving from a poor to a richer neighborhood.

- Both Chetty et al. (MTO paper) and Chyn find that younger kids benefit more
- Chetty et al. even find negative estimates for older kids (although not statistically significant)
- **Why do you think this is?**

Discussion II

- **Internal validity** **(do we learn the true effect for the treated population?)**
- Are the statistical inferences about causal effects valid for the population being studied?
  - That is, are we free of selection bias for example?

- **External validity** **(can we extrapolate to other populations?)**
  - Can the statistical inferences be generalized from the population and setting studied to other populations and settings, where the “setting” refers to the legal, policy, and physical environment and related salient features?
  - For example, can we learn something concerning Helsinki or other cities from the Chicago experience (or the MTO)?

Chyn 2018:

From introduction:

“The findings in this paper suggest these positive benefits of neighborhood change are not limited to the type of households that volunteered for the MTO experiment. “

Recap

- **We discussed what is meant by “observational data”**
- **The challenges of estimating causal effects with observational data can be considerable**
  - Economic theory predicts that choices will be endogenous, and thus naive correlations are misleading
- **Sometimes we can make use of “natural” of “quasi-experiments”**
  - Historical episodes that provide observable, quasi- or “as good as” random variation in treatment
- **In most cases, internal validity of quasi-experiments is not as strong as internal validity with experimental designs**