Past and Present Differences in Opportunity by Neighborhood

Liana Fox  
US Census Bureau

Catherine Massey  
Welch Consulting and University of Michigan

Jonathan Rothbaum  
US Census Bureau

Disclaimer: This work was made possible by the US 2050 project, supported by the Peter G. Peterson Foundation and the Ford Foundation. The statements made and views expressed are solely the responsibility of the authors. This report is released to inform interested parties of ongoing research and to encourage discussion. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the US Census Bureau.

Abstract

Recent research finds that childhood neighborhoods affect adult economic outcomes, especially for children of low-income parents. However, understanding why one neighborhood results in better outcomes for low-income children than another is extremely challenging using estimates from only one point in time. Because places are shaped by both contemporary and historical factors, it is important to understand geographic differences in opportunity both today and in the past. Using 1940 Census data linked to 1040 tax returns, we examine geographic differences in child outcomes experienced by cohorts born roughly 50 years apart – revealing how intergenerational persistence of status has changed over time both at the national level and at smaller geographic levels. In studying these changes, we hope to shed light on the causes of intergenerational mobility and inequality of opportunity.
Motivation


• Regional differences in opportunity may change over time as neighborhoods are shaped by both contemporary and historical factors

• Place-based opportunity may trend differently for minorities relative to whites

• Difficult to evaluate place-based policies that affect children with only cross section of child outcomes by place

Research Questions

• For older cohorts, what is the association (and causal relationship) between place and their adult outcomes?

• How has place-based mobility changed since the early 20th century?

• Has place-based mobility evolved differently for different groups (blacks vs. whites, for example)?
Data

• 1940 Census
• 2000 Census
• 2001-2015 American Community Survey
  • Location and income information for filers
• Social Security Administration Numident File

Data Linkage – Person Identification Validation System (PVS)

• PVS assigns 9 digit, unique identifiers called Protected Identification Keys (PIKs) via probabilistic matching techniques to surveys and decennial data
• PIKs are used to facilitate removing duplicates and record linkage
• Once ‘PIKed,’ data can be linked to any other data processed through PVS
## Data Linkage Example
### 1922-1940 Cohort

<table>
<thead>
<tr>
<th>PIK</th>
<th>First Name</th>
<th>Last Name</th>
<th>Relation</th>
<th>Occupation</th>
<th>Wages</th>
<th>Age</th>
<th>Birth State</th>
</tr>
</thead>
<tbody>
<tr>
<td>111-11-111</td>
<td>Abraham</td>
<td>Simpson</td>
<td>Head</td>
<td>Security Guard</td>
<td>4,000</td>
<td>26</td>
<td>New York</td>
</tr>
<tr>
<td>222-22-222</td>
<td>Mona</td>
<td>Simpson</td>
<td>Spouse</td>
<td></td>
<td>0</td>
<td>25</td>
<td>New York</td>
</tr>
<tr>
<td>333-33-333</td>
<td>Herb</td>
<td>Simpson</td>
<td>Child</td>
<td></td>
<td>7</td>
<td>1</td>
<td>Springfield</td>
</tr>
<tr>
<td>444-44-444</td>
<td>Homer</td>
<td>Simpson</td>
<td>Child</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 1974 IRS 1040

<table>
<thead>
<tr>
<th>PIK</th>
<th>First Name</th>
<th>Last Name</th>
<th>Relation</th>
<th>Adjusted Gross Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>444-44-444</td>
<td>Homer</td>
<td>Simpson</td>
<td>Head</td>
<td>2,000</td>
</tr>
<tr>
<td>555-55-555</td>
<td>Marge</td>
<td>Simpson</td>
<td>Spouse</td>
<td></td>
</tr>
</tbody>
</table>

### 2000 Longform Census

<table>
<thead>
<tr>
<th>PIK</th>
<th>First Name</th>
<th>Last Name</th>
<th>Relation</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>444-44-444</td>
<td>Homer</td>
<td>Simpson</td>
<td>Head</td>
<td>12 Years</td>
</tr>
<tr>
<td>555-55-555</td>
<td>Marge</td>
<td>Simpson</td>
<td>Spouse</td>
<td>16 Years</td>
</tr>
</tbody>
</table>

### Regression of Child Status on Parent Status

<table>
<thead>
<tr>
<th>Education (Years of Schooling)</th>
<th>1940 LF and ACS Cohort</th>
<th>1940 LF and ACS Cohort</th>
<th>1940 LF and ACS Cohort</th>
<th>1940 LF and ACS Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope</td>
<td>0.3372</td>
<td>0.4017</td>
<td>0.2351</td>
<td>0.3289</td>
</tr>
<tr>
<td>Intercept</td>
<td>9.163</td>
<td>8.736</td>
<td>38.69</td>
<td>33.55</td>
</tr>
<tr>
<td>Observations</td>
<td>2,130,000</td>
<td>218,000</td>
<td>19,110,000</td>
<td>5,180,000</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.16</td>
<td>0.15</td>
<td>0.06</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Method

• Calculate outcomes for children born in each commuting zone following work by Chetty and Hendren (2015)
  • Rank-rank slope and intercept and expected outcome (for below- and above-median children) by location

• Replicate Chetty and Hendren’s (2018) causal estimation of neighborhoods on intergenerational gaps
  • Limit sample to children who moved once during childhood
  • Regress income rank of children who move on origin characteristics, destination characteristics, and parental income rank all interacted with age-at-move fixed effects

Location in Historical Sample

Where did the Simpson children grow up?

1933
Numident
Herb born in New York

1935
Census
Reported living in NY

1940
Census
Observed in Springfield

1939
Numident
Homer born in Springfield

NY: 4 years
Moved from NY
Springfield: 3 years+

Herb’s Exposure by place
Impute move in 1937
Causal Estimate

<table>
<thead>
<tr>
<th>Birth-Move Estimates</th>
<th>Chetty and Hendren Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Causal Effect</strong></td>
<td><strong>Per Year</strong></td>
</tr>
<tr>
<td>1940 Cohort</td>
<td>0.016</td>
</tr>
<tr>
<td>(Per Year)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>2000 Longform/ACS Cohort*</td>
<td>0.040</td>
</tr>
<tr>
<td>Observations</td>
<td>(1-Time movers in sample)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Challenges in 1922-1940 Cohort

- 1940 measures of status (asked of full-count census!)
  - Earnings
    - Wage and salary
    - Self-employment — a dummy variable for >= $50
      - Relatively common — weighted by child: 9% of mothers, 33% of fathers
      - Substantial income missing in data
    - Expect some attenuation of parent-child relationship as parent status is measured with error (life-cycle bias, measurement error in earnings, transitory shocks, etc.)
  - Education
    - Years of schooling
- Child linkage — 70 percent of children are linked
  - Non-random — observables such as race are associated with linkage probability
Solutions

• Inverse Probability Weights
  • Several Weights
    • No weights (full-count census represents all children, weight of 1 to each)
    • IPW – regress dummy for group on family and parent characteristics as well as geographic summary variables at county, CZ, and state level
      • Child PIK IPW (Dummy = is child PIKed)
      • Parent Earnings IPW (Dummy = parent earnings > 0)
      • Both (Dummy = parent earnings > 0 | is child PIKed = 1)
  • Calculate any summary stat/regression coefficient for all samples
  • Evaluate impact of different weights on results, especially for local mobility statistics

Next Steps

• Conduct analysis separately for blacks and whites
• Evaluate CZ changes that are associated with mobility changes
• Predict causal effects by place in 1940 (as in Chetty and Hendren, 2018)