Sorting and Geographic Variation in Intergenerational Mobility

Jonathan Rothbaum
Social, Economic and Housing Statistics Division
US Census Bureau

2017 ASSA Annual Meeting
January 6, 2017

This presentation is to inform interested parties of ongoing research and to encourage discussion. The views expressed in this research are those of the author and not necessarily those of the U.S. Census Bureau.

Research Question

- Which places cause better and worse outcomes for children?
  - Policymakers – inform decisions about where to focus resources and attention
  - Parents – a way to invest in children that they currently lack information about

Background

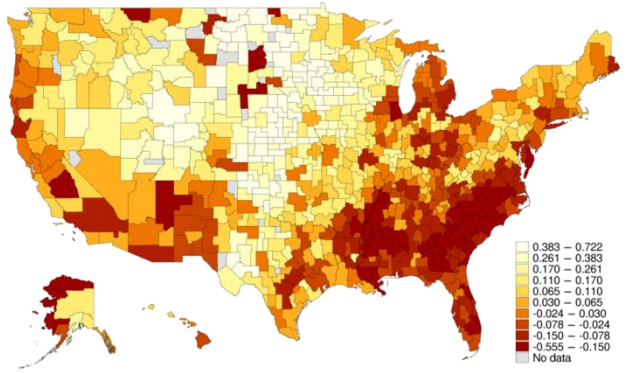
- Place matters for child outcomes
  - Randomization Experiments – Gould et al., 2011; Damm and Dustman, 2014; Damm 2014, MTO work (Katz et al., 2001; Kling et al., 2007; etc.); Oreopoulos, 2003; among others
  - Chetty, Hendren, Klein, and Saez (CHKS, 2014) – location is strongly associated with intergenerational mobility
    - Best places have more mobility than most mobile countries
    - Worst places have less mobility than any country with data
  - Chetty and Hendren (2016a) – study movers to separate causal effects of place from geographic variation in parent characteristics (sorting)
    - Estimate that at least 50% of the variation in CHKS is causal
    - Each year in a location, child's adult income converges 4% to location average

Background

- Chetty and Hendren (2016b) – estimate which specific places cause good/bad outcomes (as opposed to general result about effect of good vs. bad places from 2016a)
  - Use movers (at different ages) to control for family fixed effects
  - Unfortunately, even with a sample of > 1 million, the estimates are very imprecise
  - Forecast estimate – accept bias for precision

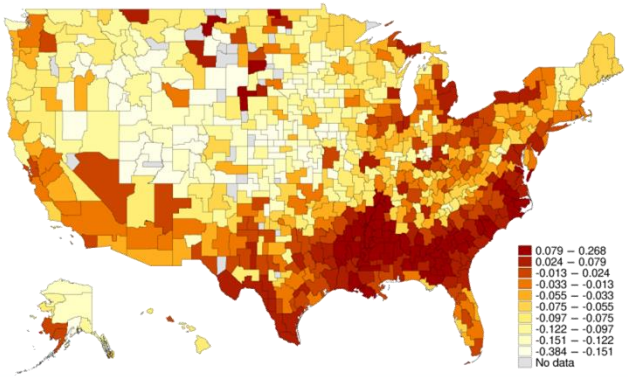
Where We're Headed: Reducing Sorting Bias

Unadjusted Forecast



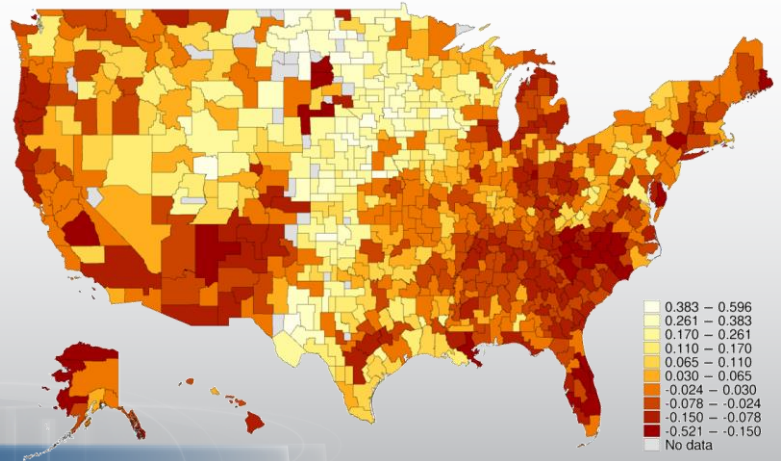
+

Sorting Adjustment



Sorting-Adjusted Estimate

=



Model

Suppose outcome for child i depend on parent and location characteristics:

$$y_i = \alpha + \underbrace{\beta X_i}_{\text{Family Effects (Sorting)}} + \underbrace{\gamma Z_c}_{\text{Location Effects}} + \underbrace{\sum_{j=1}^{J}\sum_{k=1}^{K} \delta_{jk} x_{ji} z_{kc}}_{\text{Heterogeneous Location Effects}} + e_i$$

y_i = Child outcome

X_i = Parent/Family Characteristics, $X_i = (x_1, \ldots, x_J)$

Z_c = Location Characteristics, $Z_c = (z_{1c}, \ldots, z_{Kc})$

Location Estimates

- Observed outcomes in location c for permanent residents (stayers, s)

$$\bar{y}_{sc} = \alpha + \beta \bar{X}_{sc} + \gamma Z_c + \sum_{j=1}^{J} \sum_{k=1}^{K} \delta_{jk} \bar{x}_{jsc} z_{kc}$$

Sorting | Location Effects | Heterogeneous Location Effects

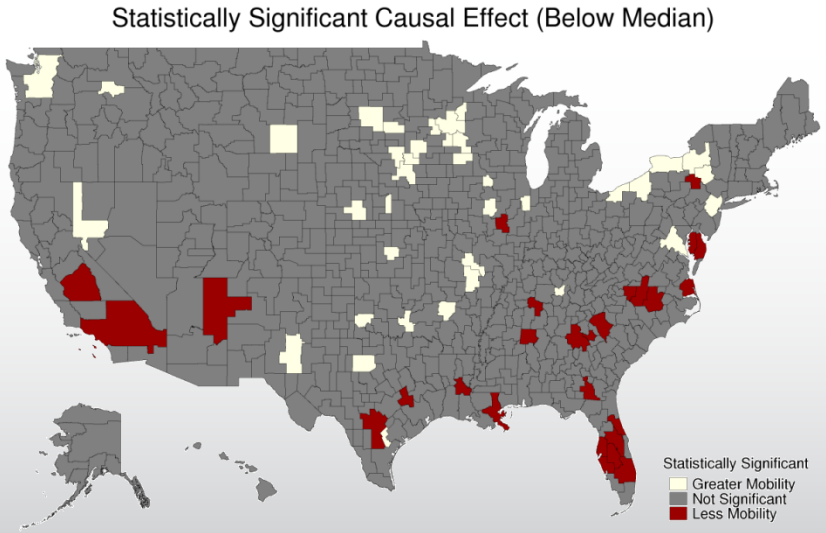
$\bar{y}_{sc}$ = Expected child outcome in location c

$\bar{X}_{sc}$ = Average parent characteristics in location c,

$$\bar{X}_{sc} = \left(\bar{x}_{1sc}, \ldots, \bar{x}_{Jsc}\right)$$

Chetty and Hendren Causal Estimates

- Mover sample: ~1.9 million observations across 741 CZs (~2,500 per CZ)
- Yields very imprecise estimates

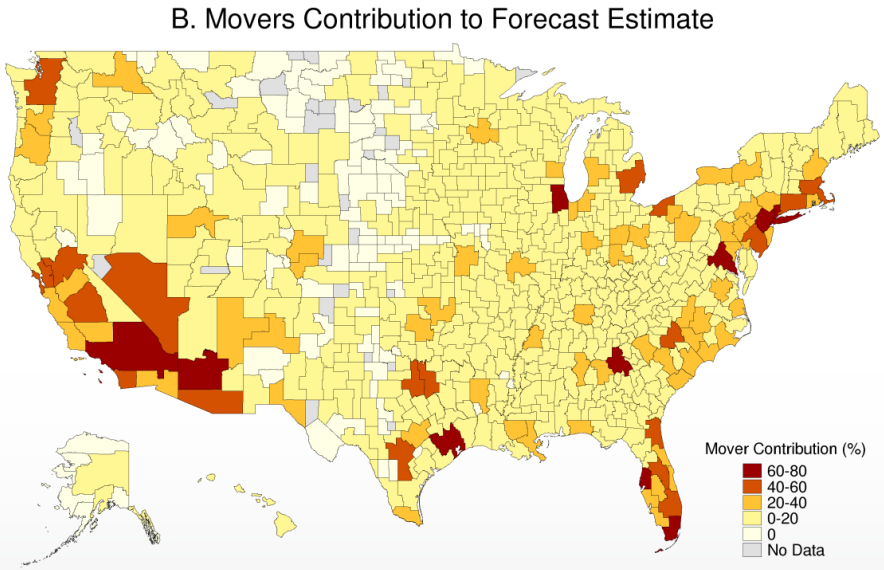


Source: Author's calculation from Chetty and Hendren's raw causal estimates (data available at http://www.equality-of-opportunity.org/).

Forecast

- Two sources of information
  - Permanent residents ($\bar{y}_{sc}$) – very precise, but biased
  - Movers (μ_{mc}) – very imprecise, but unbiased
- Combine as in a forecast
  - Weight on permanent residents is a function of uncertainty in mover estimate

Forecast Bias



- Bias is decreasing in weight on mover term

Controlling for Sorting

- Assuming we cannot get precise causal estimates from movers, we could theoretically adjust for sorting (a) as:

$$\underbrace{\bar{y}_{ac}}_{\text{Sorting-adjusted estimates in location } c} = \underbrace{\bar{y}_{sc}}_{\text{Biased estimate from permanent residents}} - \underbrace{\beta \bar{X}_{sc}}_{\text{Sorting Adjustment}}$$

To estimate adjustment

1) $\bar{X}_{sc}$ = parent chars in location c

2) β = causal effect of X on child outcome

Partial Sorting Adjustment

- Suppose some family characteristics are observed
- If we regress child outcomes on observed characteristics:
  - Correlation with unobserved family characteristics– may not be a concern
  - Correlation with location effects– definitely a problem

Partial Sorting Adjustment

- Consider two regressions:
  1. Child outcome on observed characteristics only ($\tilde{\beta}_O$)
  2. Child outcome on observed characteristics and location characteristics ($\tilde{\beta}_{O|Z}$)
     - If $\tilde{\beta}_O \approx \tilde{\beta}_{O|Z}$, then that is strong evidence that omitted variable bias would not confound the partial sorting adjustment

Estimating $\tilde{\beta}_o$: CPS-SIPP DER (CSD)

- Survey Data
  - Current Population Survey Annual Social and Economic Supplement (CPS ASEC)
    - 91,94,96-09
  - Survey of Income and Program Participation (SIPP) Gold Standard File
    - 90-93, 96, 01, 04, and 08
  - Information on parent-child links, parent characteristics (education, age, single/married, etc.), location
- Administrative Data - SSA and IRS Detailed Earnings Records
  - Extract from Master Earnings File linked to surveys
  - W-2 wage and self-employment earnings for all individuals from 1978-2012
- Baseline Sample – 49,559 children

Estimating $\overline{X}_{sc}$:1990 Longform Decennial Census

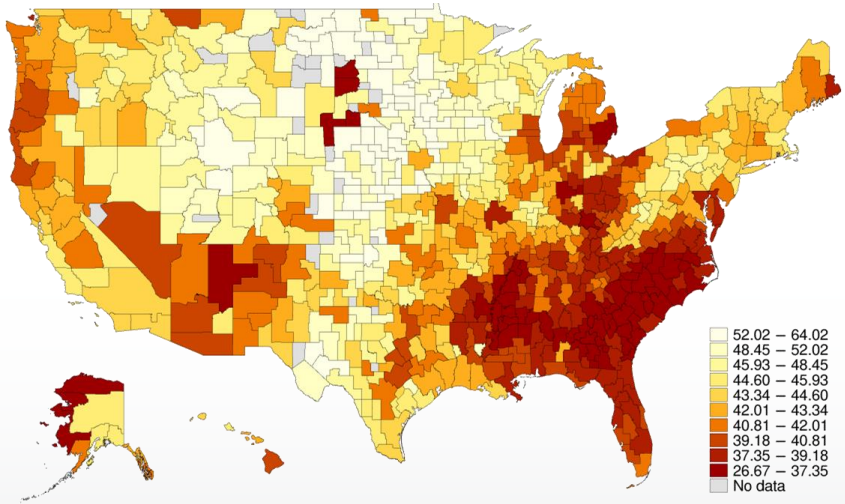
- Used to adjust for sorting in estimating the causal effects of place
- Sample
  - includes 5.1 million families with 10 million children 18 or under (from an estimated population of 32 million child families with 64 million total children)
  - Contains all the parent characteristic and location information present in the CSD file
  - Calculate parent rank from survey response
  - Assign expected child rank based on parent rank and Chetty et al. (2014) absolute and relative mobility parameters

Sorting Adjustment and Bias

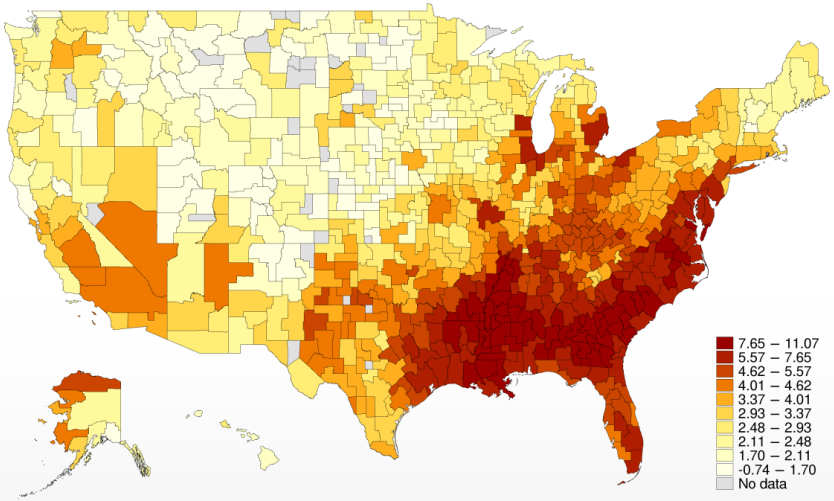
		Proxy for Location Effect					
		Permanent Residents		Unadjusted Forecast		Raw Causal	
Dependent Variable = Child Rank	Baseline (1)	(2)	(3)	(4)	(5)	(6)	(7)
Parent Rank	0.215***	0.218***	0.374***	0.208***	0.287***	0.216***	0.260***
	(0.014)	(0.013)	(0.082)	(0.013)	(0.086)	(0.014)	(0.085)
Black	-16.19***	-15.27***	-16.26***	-16.05***	-17.16***	-16.04***	-17.29***
	(1.16)	(1.40)	(3.69)	(1.44)	(2.41)	(1.20)	(2.14)
Black*Female	15.48***	15.32***	15.62***	15.62***	13.77***	15.33***	14.81***
	(1.12)	(1.46)	(3.25)	(1.60)	(2.06)	(1.23)	(2.46)
Most Educated Parent							
<High School	-6.19***	-6.20***	-6.67***	-5.90***	-8.18***	-6.28***	-4.77**
	(1.04)	(1.04)	(2.30)	(1.13)	(1.73)	(1.03)	(2.00)
College+	8.42***	8.72***	9.81***	8.50***	9.33***	8.75***	8.42***
	(1.09)	(1.09)	(2.90)	(1.05)	(2.78)	(1.11)	(2.14)
Interacted with Parent Rank							
Black	0.061***	0.054**	0.120**	0.063***	0.106**	0.062***	0.095***
	(0.021)	(0.024)	(0.054)	(0.023)	(0.049)	(0.021)	(0.034)
<High School	0.031	0.036	0.101*	0.037	0.154***	0.032	0.022
	(0.030)	(0.030)	(0.057)	(0.029)	(0.053)	(0.029)	(0.052)
College+	-0.039**	-0.042**	-0.075*	-0.041**	-0.041	-0.045**	-0.029
	(0.017)	(0.018)	(0.045)	(0.017)	(0.039)	(0.018)	(0.031)
Location Effect		4.80		13.41**		-1.48	
		(4.32)		(5.94)		(2.16)	
Constant	44.50***	44.17***	44.13***	45.06***	41.65***	44.52***	52.43***
	(0.87)	(0.86)	(4.25)	(0.80)	(4.27)	(0.88)	(4.46)
Causal Interacted with parent rank and baseline model dummies		X		X		X	
Causal quintile dummies (interacted with all baseline model variables, 3rd quintile excluded)			X		X		X
R-Squared	0.11	0.11	0.12	0.12	0.12	0.11	0.12
Observations	49,559	49,102	49,102	49,102	49,102	48,277	48,277

Sorting Adjustment

CHKS (Non-Causal)



Sorting Adjustment

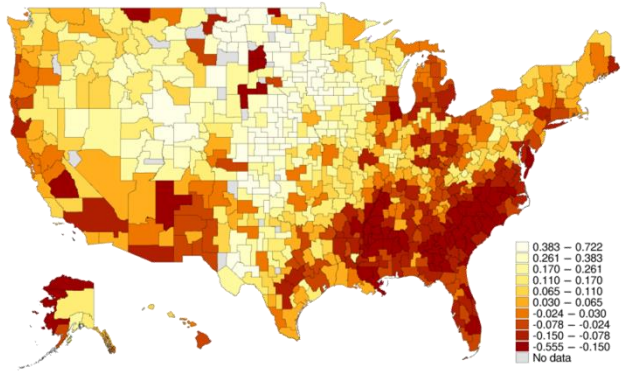


Source: Chetty et al., 2014

- Correlation of sorting adjustment with observed mobility = -0.69
- Sorting adjustment reduces the variance by 46 percent, compared to Chetty and Hendren estimate of 30-50 percent of variation due to sorting

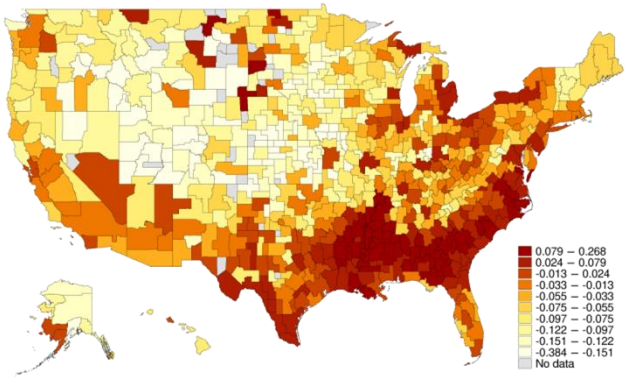
Sorting Adjustment – Causal

Chetty and Hendren



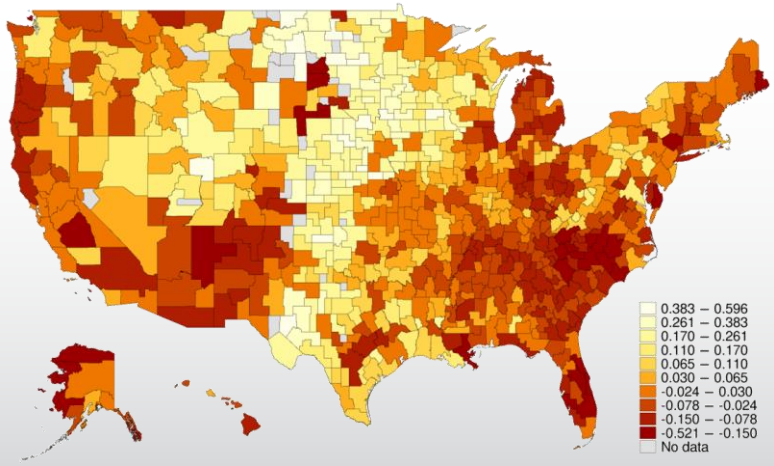
+

Causal Sorting Adjustment



Sorting-Adjusted Causal Estimate

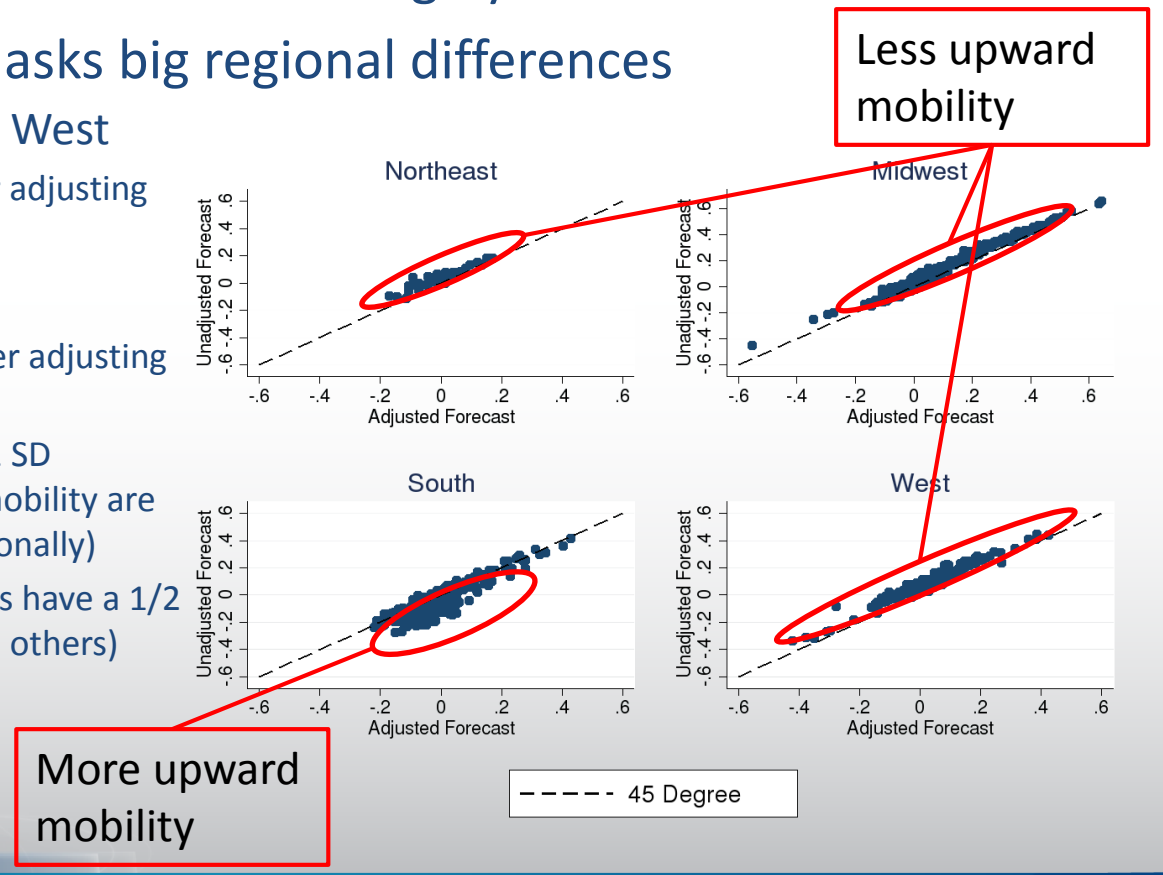
=



- Correlation of sorting adjustment with forecast = -0.46

Sorting-Adjusted Causal Estimates By Region

- CH and sorting-adjusted estimates are highly correlated
- However, correlation masks big regional differences
  - Northeast, Midwest, and West
    - Less upward mobility after adjusting for sorting (above 45°)
  - South
    - More upward mobility after adjusting for sorting
    - 90% of CZs with over a 1/2 SD improvement in upward mobility are in the South (85 of 94 nationally)
    - About 30% of Southern CZs have a 1/2 SD improvement (2% of all others)



Variation Reduced By

- Region – 74 percent
- Division – 50 percent
- State – 39 percent

Most Affected CZs (Largest 50)

CZ	Sorting Adjustment (ΔRank/Year)	Percent Difference in Income at 26 (20 Years Exposure)	Change in Rank (Among Largest 50)
Baltimore, MD	0.042	2.7	6
New Orleans, LA	0.032	2.0	1
St. Louis, MO	0.028	1.8	3
Jacksonville, FL	0.026	1.6	6
Milwaukee, WI	0.025	1.6	10
San Antonio, TX	-0.037	-2.3	-5
Denver, CO	-0.038	-2.4	-7
Portland, OR	-0.042	-2.6	-10
Salt Lake City, UT	-0.055	-3.4	-2
Manchester, NH	-0.062	-3.9	-13

Most Affected Counties (Largest 100)

County	Sorting Adjustment (ΔRank/Year)	Percent Difference in Income at 26 (20 Years Exposure)	Change in Rank (Among Largest 100)
Washington, DC	0.129	8.1	36
Baltimore City, MD	0.086	5.4	17
Shelby, TN	0.065	4.1	11
Jefferson, AL	0.062	3.9	21
Prince George's, MD	0.061	3.8	20
Pima, AZ	-0.040	-2.5	-10
Essex, MA	-0.041	-2.6	-10
Norfolk, MA	-0.047	-2.9	-1
Salt Lake, UT	-0.048	-3.0	-15
Bernalillo, NM	-0.061	-3.8	-16

Best and Worst CZs (Largest 50)

Rank	CZ	Causal Effect (ΔRank/Year)	Percent Difference in Income at 26 (20 Years Exposure)
1	Washington, DC	0.142	8.9
2	Seattle, WA	0.118	7.4
3	Fort Worth, TX	0.069	4.3
4	Minneapolis, MN	0.060	3.8
5	Salt Lake City, UT	0.047	2.9
46	Raleigh, NC	-0.187	-11.7
47	Port St. Lucie, FL	-0.188	-11.8
48	Charlotte, NC	-0.192	-12.0
49	New Orleans, LA	-0.194	-12.2
50	Fresno, CA	-0.220	-13.8

Best and Worst Counties (Largest 100)

Rank	County	Causal Effect (ΔRank/Year)	Percent Difference in Income at 26 (20 Years Exposure)
1	Bergen, NJ	0.272	17.0
2	DuPage, IL	0.247	15.5
3	Fairfax, VA	0.238	14.9
4	Bucks, PA	0.198	12.4
5	Snohomish, PA	0.175	11.0
96	Mecklenburg, NC	-0.193	-12.1
97	Orange, FL	-0.202	-12.6
98	Palm Beach, FL	-0.210	-13.2
99	Fresno, CA	-0.242	-15.2
100	Hillsborough, FL	-0.254	-15.9