Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings

Raj Chetty, Harvard and NBER
John N. Friedman, Harvard and NBER
Emmanuel Saez, UC Berkeley and NBER

April 2012

Identifying Policy Impacts

- Two central challenges in identifying the impacts of tax policies:

  1. Difficult to find comparison groups to estimate causal impacts of policies [Meyer 1995, Gruber 2008]

  2. Difficult to identify long run impacts from short-run responses to tax changes

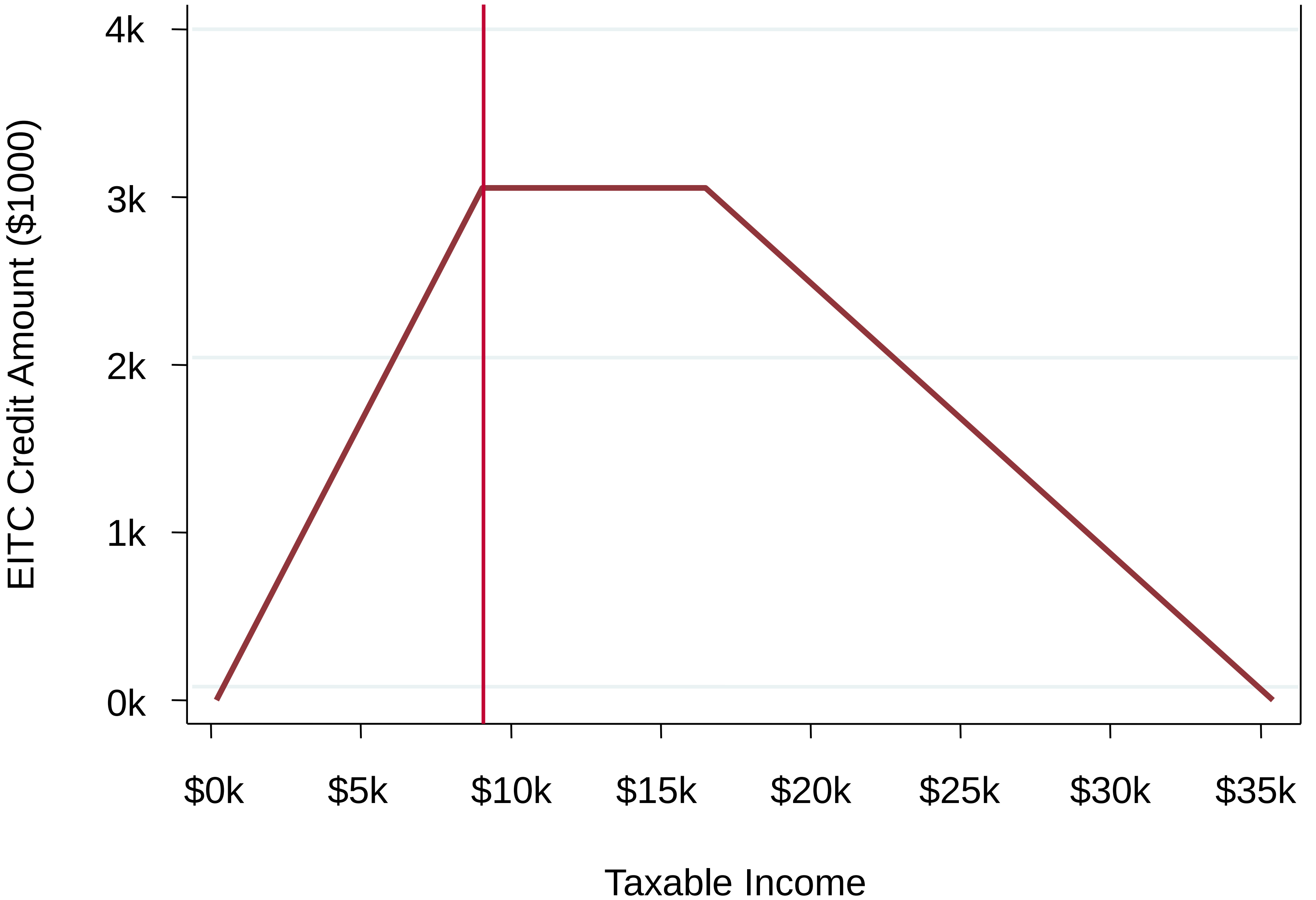
     - Many people are uninformed about tax and transfer policies [Brown 1968, Bises 1990, Chetty and Saez 2009]

     - Workers face switching costs for labor supply [Cogan 1981, Altonji and Paxson 1992, Chetty et al. 2011]

Overview

- We address these challenges by exploiting differences across neighborhoods in knowledge about tax policies
  - Idea: use cities with low levels of information about tax policies as "control groups" for behavior in the absence of tax policy
- Apply this approach to characterize the impacts of the Earned Income Tax Credit (EITC) on the earnings distribution in the U.S.
  - EITC provides refunds of up to $5,000 to approximately 20 million households in the U.S.

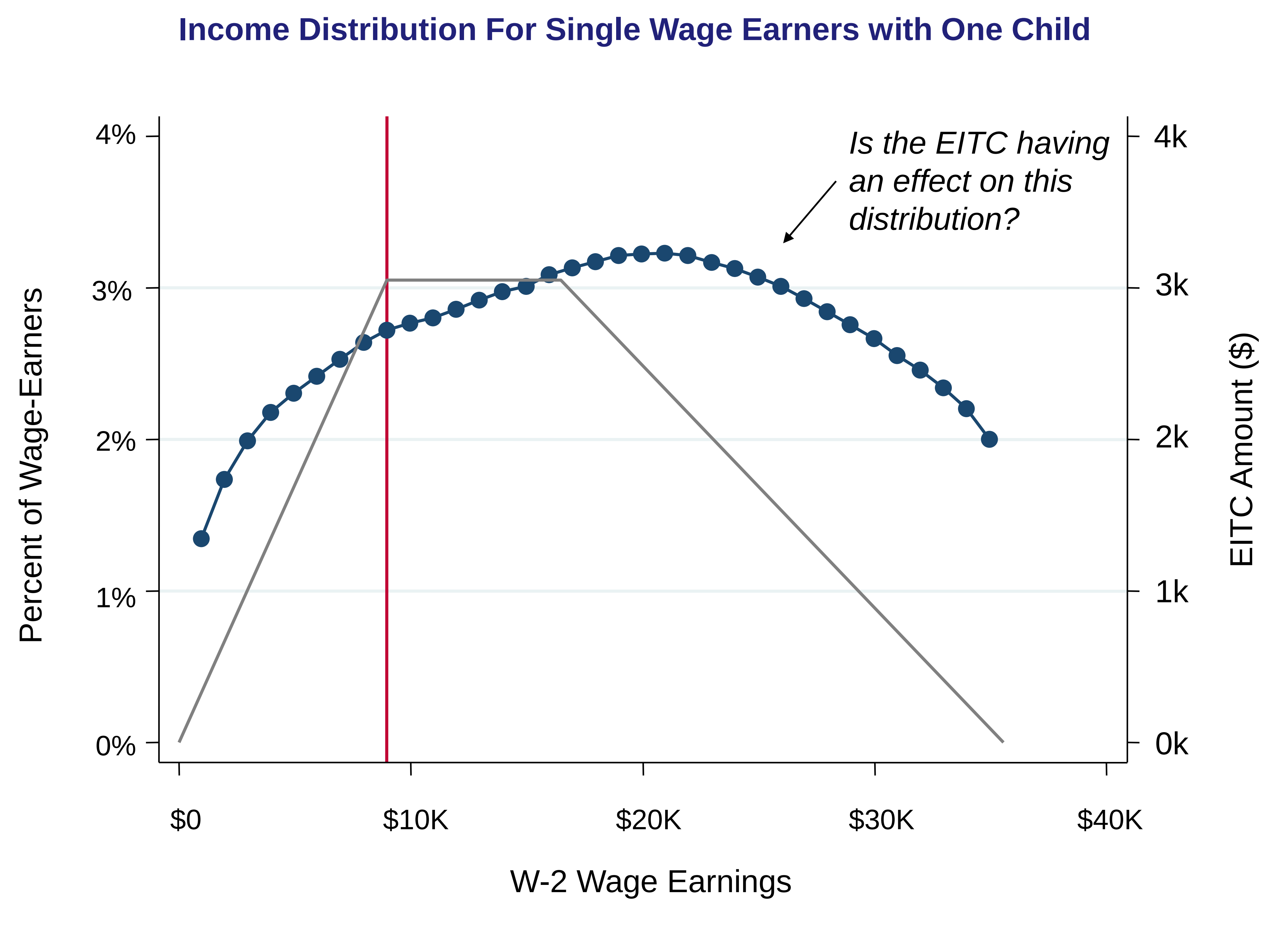
Earned Income Tax Credit Schedule for Single Earners with One Child



Relationship to Prior Work

- Large literature has studied the impacts of EITC on labor supply
  [Eissa and Liebman 1996, Meyer and Rosenbaum 2001, Meyer 2002, Grogger 2003, Hoynes 2004, Gelber and Mitchell 2011]
  - Clear evidence of impacts on *participation* (extensive margin)
  - But no clear, non-parametric evidence on impacts of EITC on *earnings distribution* (intensive margin)
  - Same pattern in studies of labor supply elasticities more generally
- Observed extensive responses may be larger because more people know about existence of EITC refund than shape of schedule
  - Gains from re-optimization are 2^{nd}-order on intensive but 1^{st} order on extensive margin → frictions attenuate intensive responses [Chetty 2011]

Income Distribution For Single Wage Earners with One Child



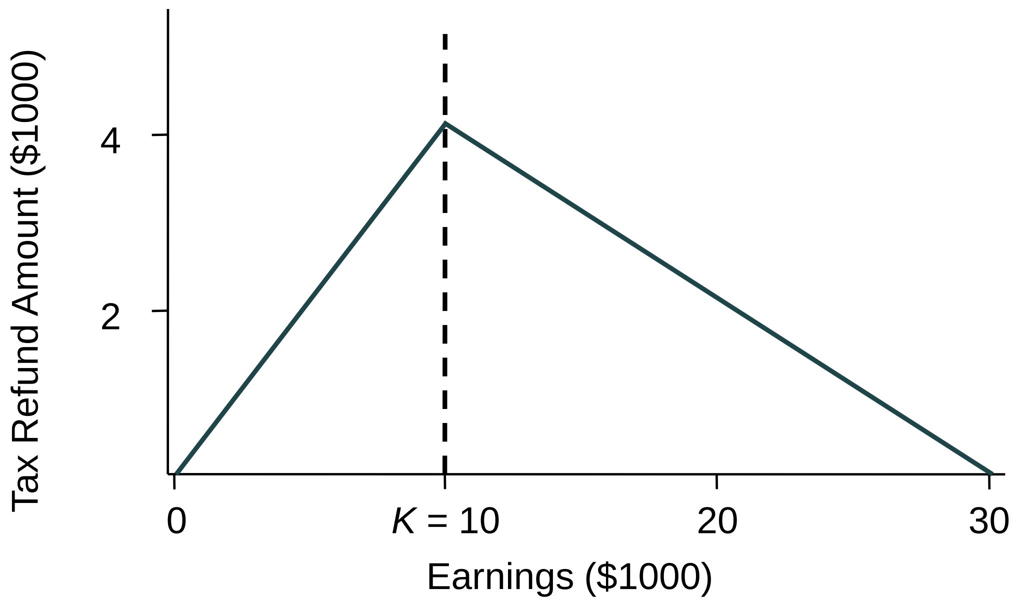
Income Distribution For Single Wage Earners with One Child
Is the EITC having an effect on this distribution?
4%
3%
2%
1%
0%
4k
3k
2k
1k
0k
$0
$10K
$20K
$30K
$40K
Percent of Wage-Earners
EITC Amount ($)
W-2 Wage Earnings

Outline

1. Conceptual Framework
2. Data and Institutional Background
3. A Proxy for Knowledge: Sharp Bunching via Income Manipulation
4. Using Neighborhood Effects to Uncover Wage Earnings Responses
5. Implications for Tax Policy

Stylized Model: Tax System

- Workers face a two-bracket income tax system $\tau = (\tau_1, \tau_2)$ and choose earnings z=wl to maximize quasi-linear utility *u(c,l)*
  - Tax rate of $\tau_1 < 0$ when reported income is below K
  - Marginal tax rate of $\tau_2 > 0$ for reported income above K
  - Tax refund maximized when income is K → bunching around K



Neighborhoods

- Cities indexed by $c = 1, \ldots, N$
- Cities differ only in one attribute: knowledge of tax code
- In city c, fraction λ_c of workers know about tax subsidy for work
  - Others optimize as if tax rates are 0 (i.e. subsidy is lump-sum)
  - With quasi-linear utility, workers with no knowledge behave as they would with no taxes
  - More generally, our technique recovers compensated elasticities
- Firms pay workers fixed wage rate in all cities

Identifying Tax Policy Impacts

- Goal: identify how taxes affect earnings distribution $F(z \mid \tau)$ with average level of knowledge in economy:

$$\Delta F(z \mid \tau) = F(z \mid \tau > 0, \bar{\lambda}_c) - F(z \mid \tau = 0, \bar{\lambda}_c)$$

- Challenge: potential outcome without taxes $F(z \mid \tau = 0, \bar{\lambda}_c)$ unobserved

- Our solution: earnings behavior with no *knowledge* about taxes is equivalent to earnings behavior with no taxes

$$F(z \mid \tau = 0, \bar{\lambda}_c) = F(z \mid \tau > 0, \lambda_c = 0)$$

$$\Rightarrow \Delta F(z \mid \tau) = F(z \mid \tau > 0, \bar{\lambda}_c) - F(z \mid \tau > 0, \lambda_c = 0)$$

Identifying Tax Policy Impacts

- Let μ_c represent a measure of bunching in earnings around kink K
  - Ex: size of EITC refund, fraction of individuals in plateau

- We identify $\mu_c(\lambda_c = 0)$ using an estimating equation of the form

$$\mu_c = \alpha + \beta\lambda_c + \eta_c$$

  - Key orthogonality condition to estimate β: $\lambda_c \perp \eta_c$

  - Identification requires that cities with different levels of knowledge do not have other attributes that affect the earnings distribution

  → Quasi-experimental research design to account for omitted variables

Data and Sample Definition

- Selected data from population of U.S. income tax returns, 1996-2009
  - Includes 1040's and all information forms (e.g. W-2's)
  - For non-filers, we impute income and ZIP from W-2's
  - For joint filers, code income as total household income or W-2's

- Sample restriction: individuals who at least once between 1996-2009: (1) file a tax return, (2) have income < \$50,000, (3) claim a dependent

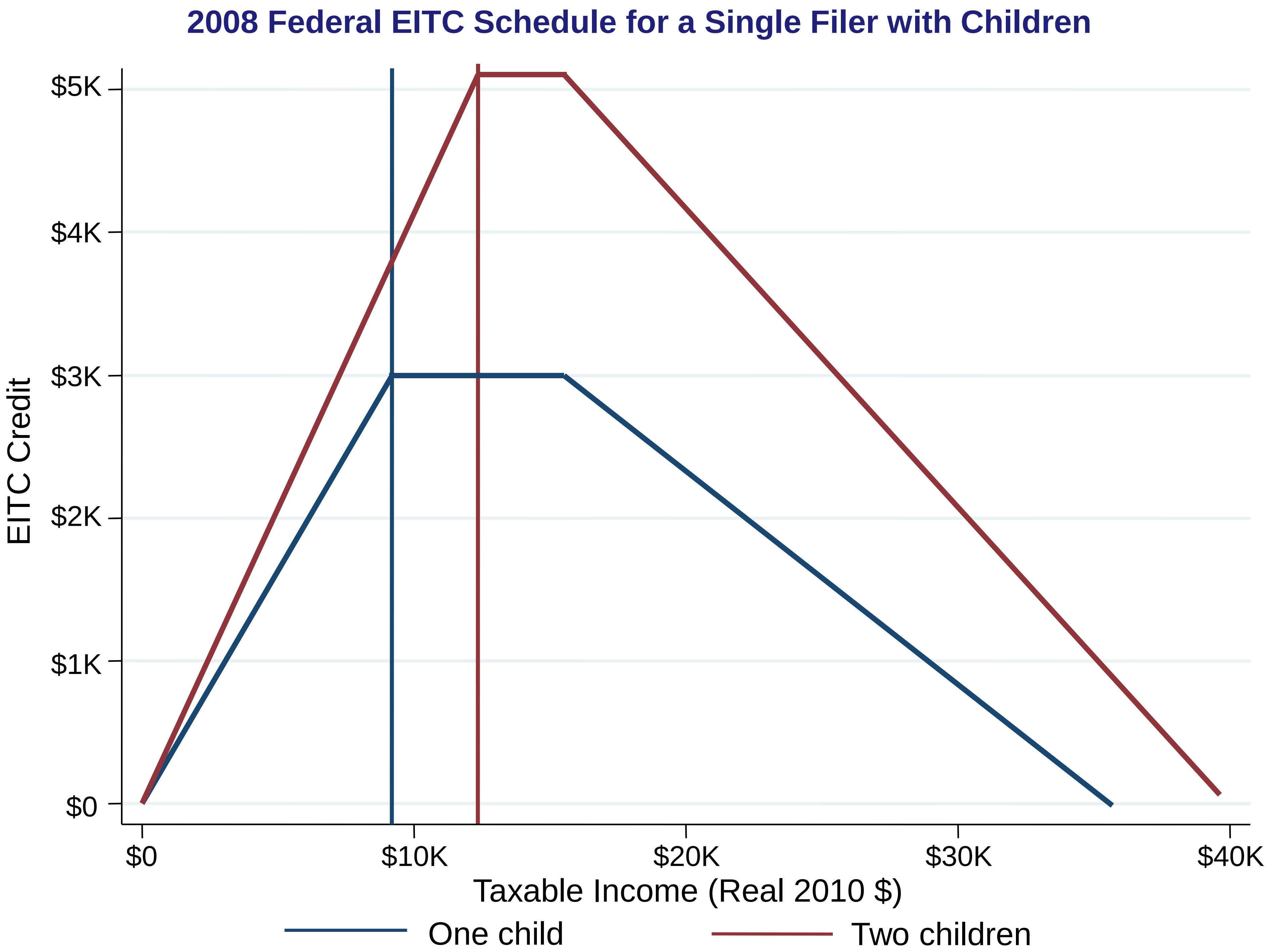
- Sample size after restrictions:
  - 77.6 million unique taxpayers
  - 1.09 billion taxpayer-year observations on income

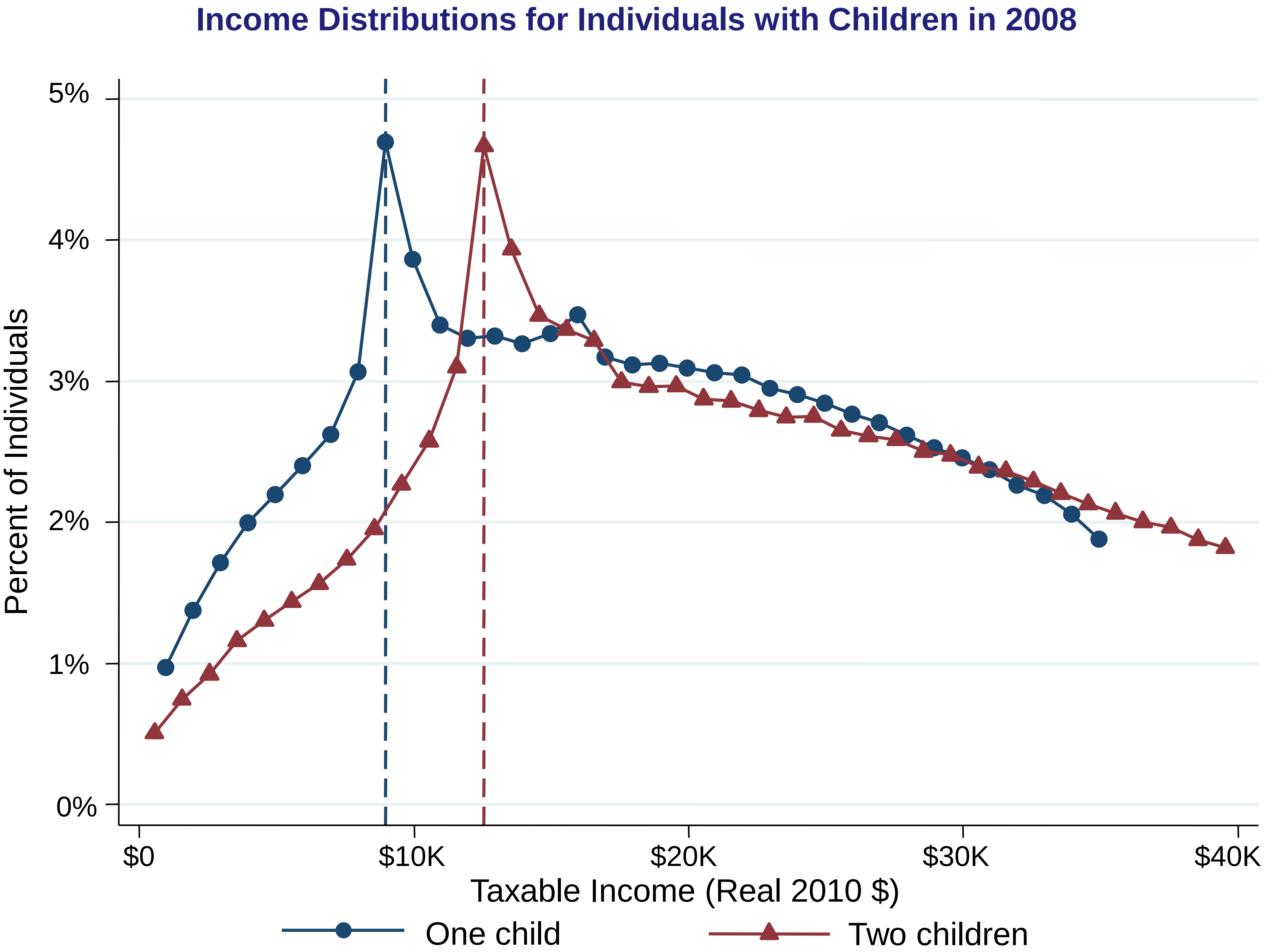
Summary Statistics

Variable	Mean
Income	$23,641
Self Employed	17.1%
Married	29%
Number of Children	1.11
Female (among single filers)	61%

Self Employment Income vs. Wage Earnings

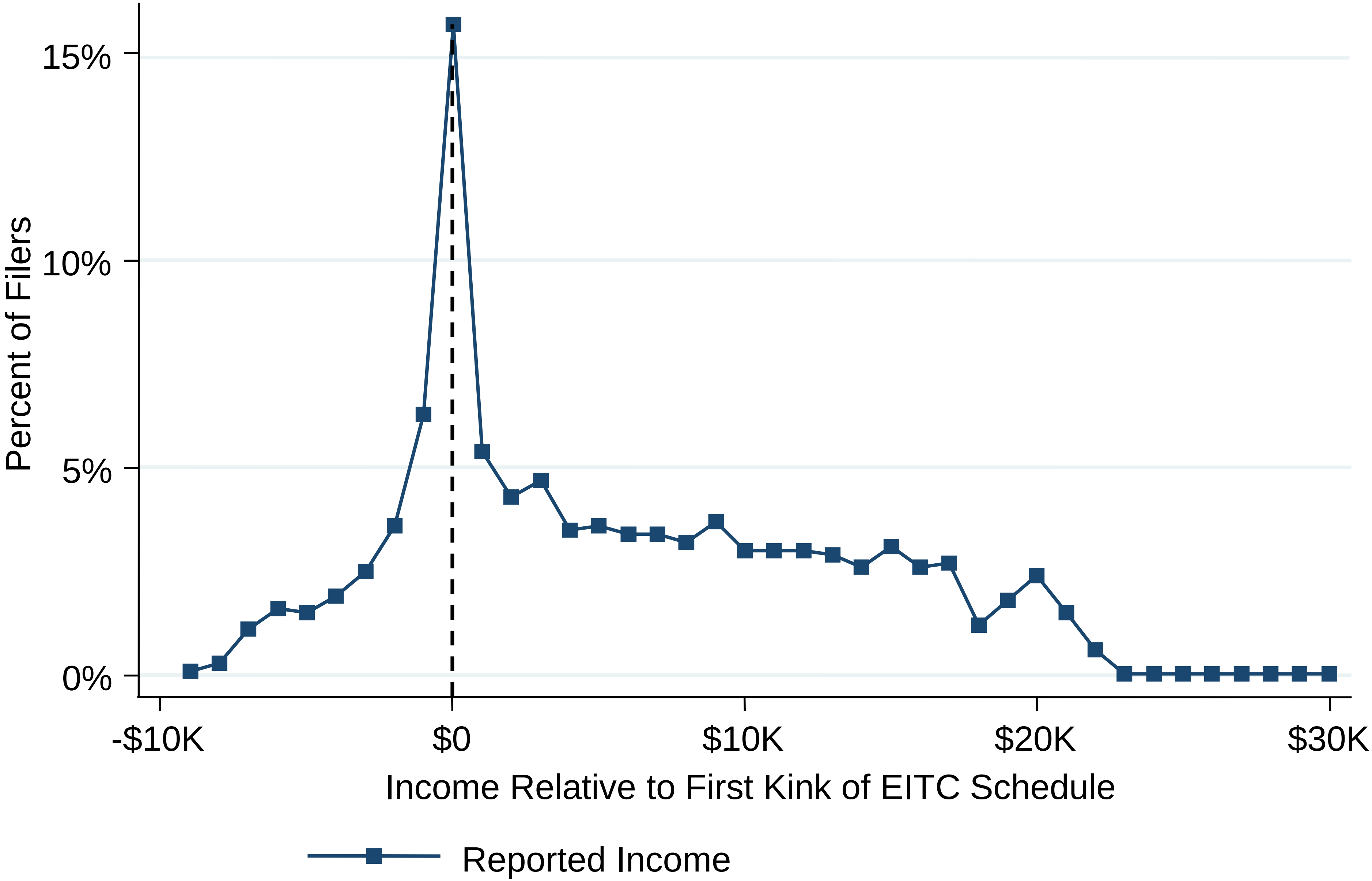
- Critical distinction: wage earnings vs. self-employment income
  - Self employed = filers with any Schedule C income
  - Wage earners = filers with no Schedule C income
- Self-employment income is self-reported → easy to manipulate
- Wage earnings are directly reported to IRS by employers
  - Therefore more likely to reflect “real” earnings behavior
- Analyze misreporting due to EITC using National Research Program Tax Audit data (joint with Peter Ganong, Kara Leibel, Alan Plumley)


2008 Federal EITC Schedule for a Single Filer with Children
EITC Credit
$5K
$4K
$3K
$2K
$1K
$0
$0
$10K
$20K
$30K
$40K
Taxable Income (Real 2010 $)
One child
Two children



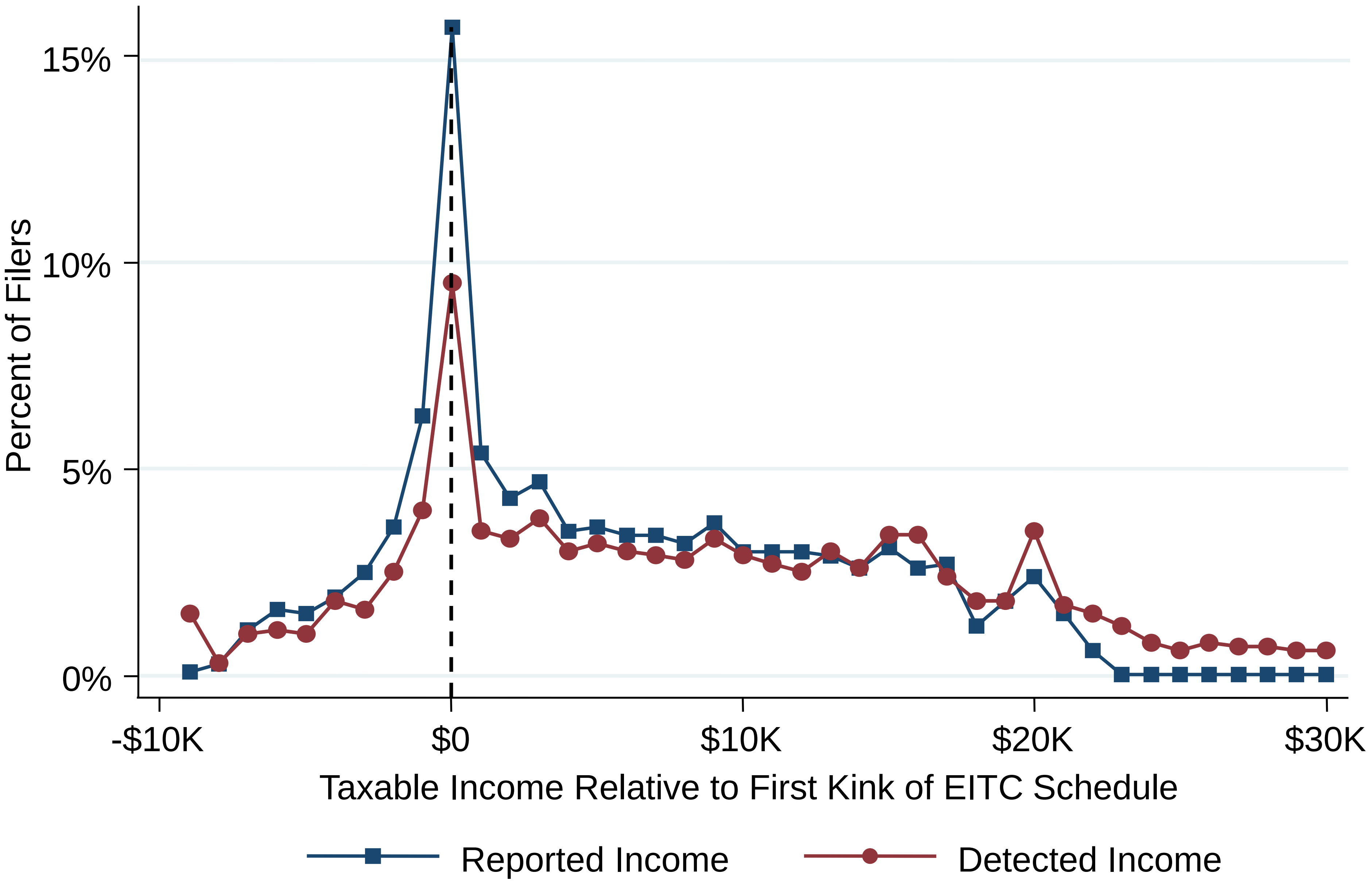
Income Distributions for Individuals with Children in 2008
5%
4%
3%
2%
1%
0%
Percent of Individuals
$0
$10K
$20K
$30K
$40K
Taxable Income (Real 2010 $)
One child
Two children

Reported vs. Audited Income Distributions for SE EITC Filers in 2001
National Research Program Tax Audit Data



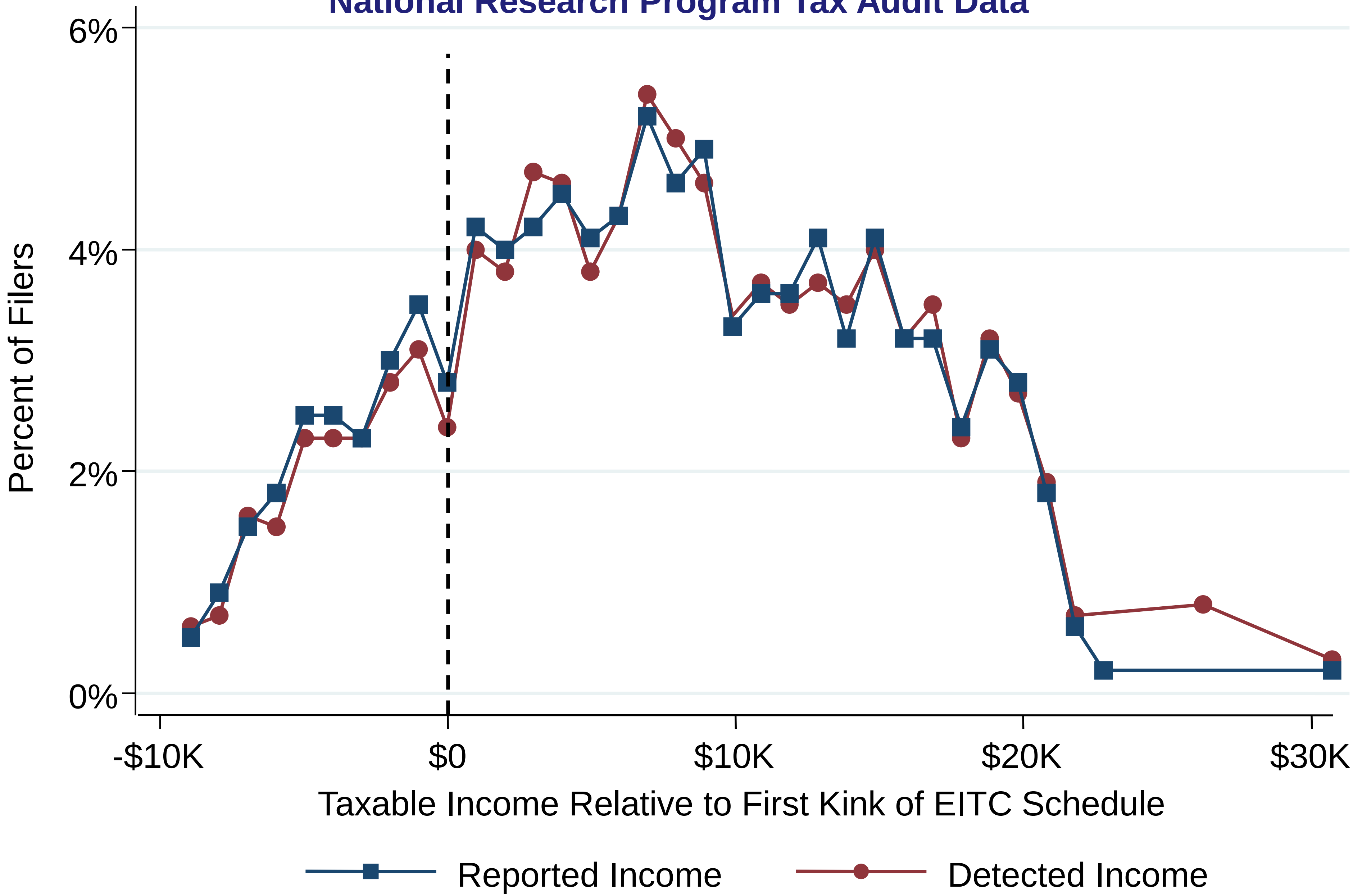
Source: IRS TY01 NRP reporting compliance study of individual income tax returns for those reporting dependent children; amounts reflect only what was detected by the auditors, weighted to population levels.

Reported vs. Audited Income Distributions for SE EITC Filers in 2001
National Research Program Tax Audit Data



Source: IRS TY01 NRP reporting compliance study of individual income tax returns for those reporting dependent children; amounts reflect only what was detected by the auditors, weighted to population levels.

Reported vs. Audited Income Distributions for EITC Wage Earners with Children
National Research Program Tax Audit Data



Source: IRS TY01 NRP reporting compliance study of individual income tax returns for those reporting dependent children; amounts reflect only what was detected by the auditors, weighted to population levels.

Empirical Implementation: Proxy for Knowledge

- We proxy for knowledge λ_c using sharp bunching at refund-maximizing kink among the self-employed
  - Intuition: use amount of misreporting to measure local tax knowledge
- Workers make two choices: earnings (z_i) and reported income ($\hat{z}_i$)
  - Fraction θ_c of workers face 0 cost of non-compliance → report $\hat{z}_i = K$
  - Remaining workers face infinite cost of non-compliance → set $\hat{z}_i = z_i$
- Fraction who report $\hat{z}_i = K$ is proportional to local knowledge:

$$f_c = \theta_c \lambda_c$$

Empirical Implementation: Lower Bound

- Recall ideal estimating equation from the model

$$\mu_c = \alpha + \beta\lambda_c + \eta_c$$

- We instead estimate the feasible regression

$$\mu_c = \alpha + \hat{\beta}f_c + \eta_c$$

- Our proxy f_c is a noisy measure of true knowledge λ_c
  - Differences across cities in f_c may be due to other determinants of tax compliance θ_c rather than knowledge λ_c
  - This measurement error attenuates estimate of β

  → *Lower bound* on estimated impact of EITC

Empirical Implementation: Point Estimate

- Stronger assumption: No sharp bunching → no knowledge about EITC schedule

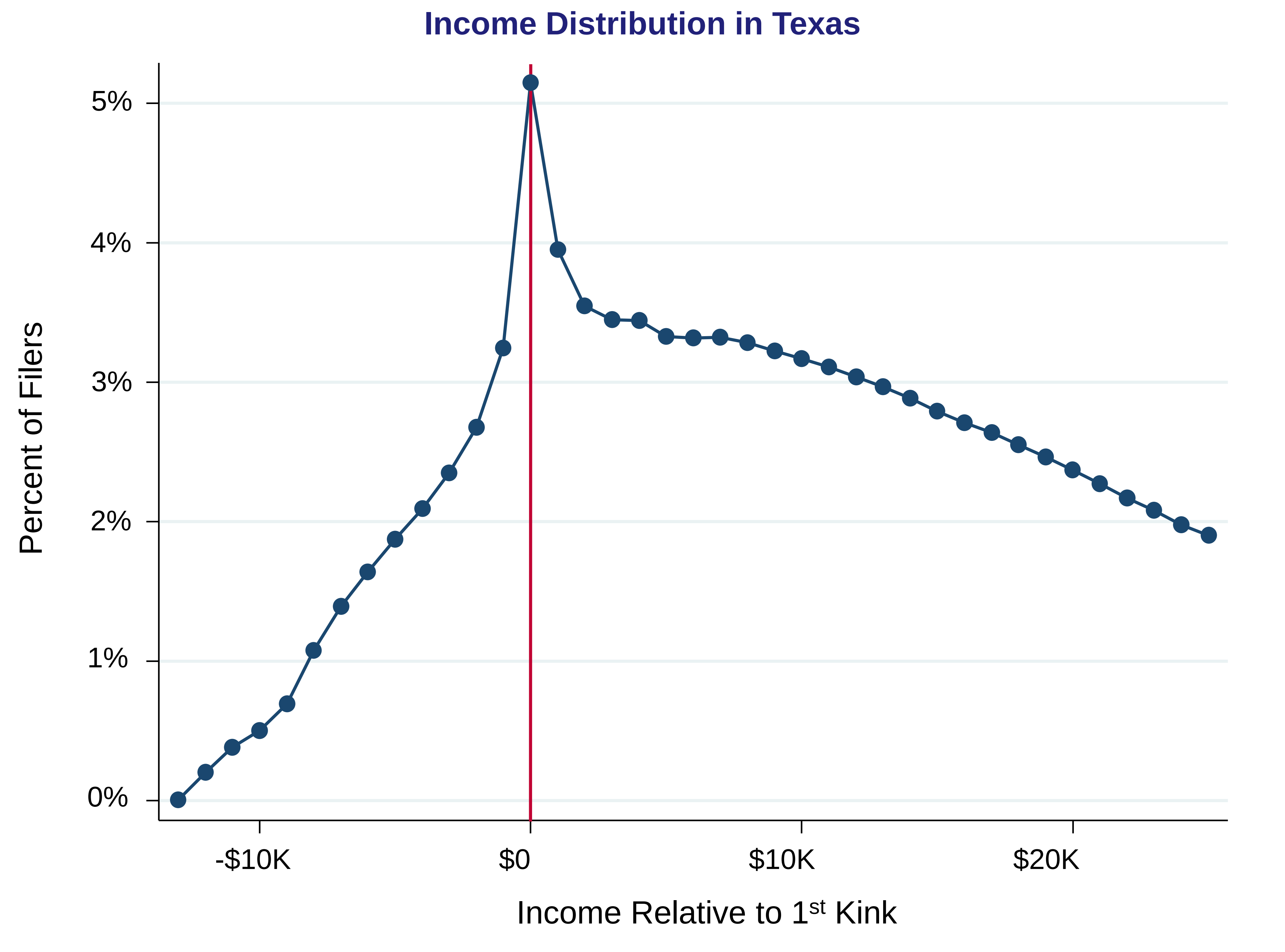
$$f_c = 0 \rightarrow \lambda_c = 0$$

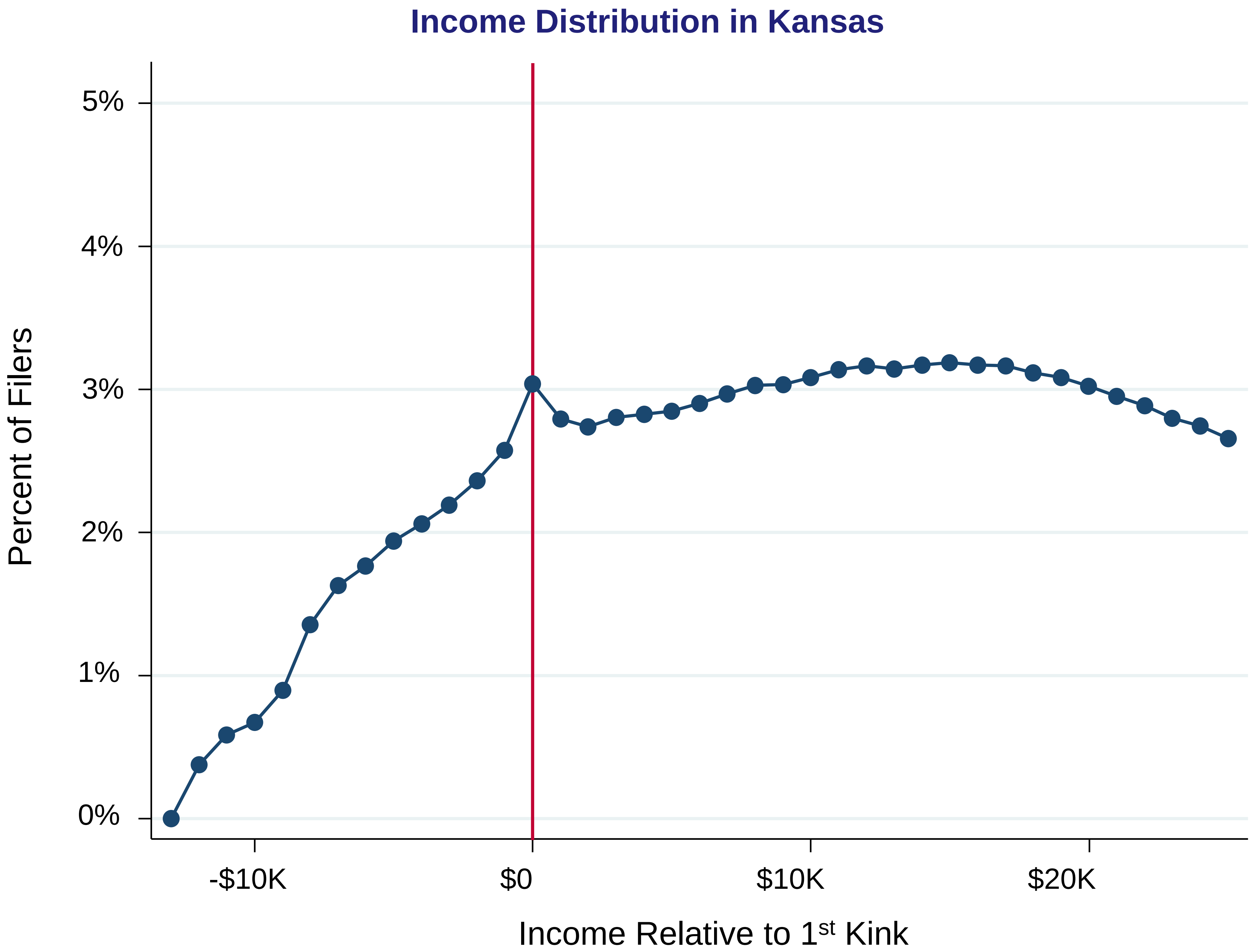
- Under this assumption, we obtain a point estimate of impact of EITC on earnings distribution with average knowledge level in economy
  - Compare aggregate distribution in economy to distribution of wage earnings in neighborhoods with $f_c = 0$

- After showing main results, we present evidence suggesting that individuals in low bunching areas completely ignore EITC

Outline of Empirical Analysis

- Step 1: Develop a proxy for knowledge about the EITC in each neighborhood using sharp bunching among self-employed

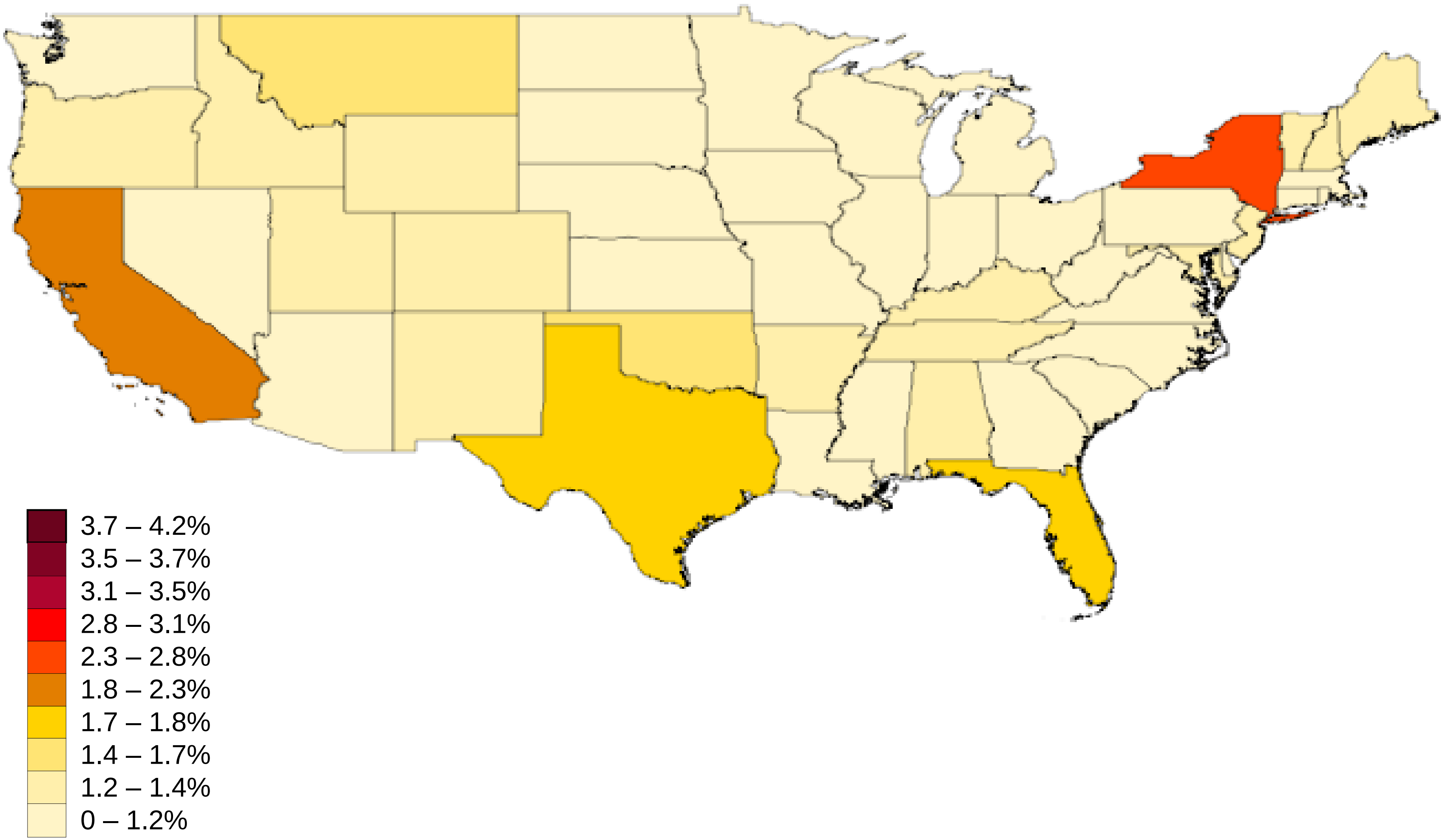

Income Distribution in Texas
Percent of Filers
5%
4%
3%
2%
1%
0%
-$10K
$0
$10K
$20K
Income Relative to 1st Kink


Income Distribution in Kansas
Percent of Filers
5%
4%
3%
2%
1%
0%
-$10K
$0
$10K
$20K
Income Relative to 1st Kink

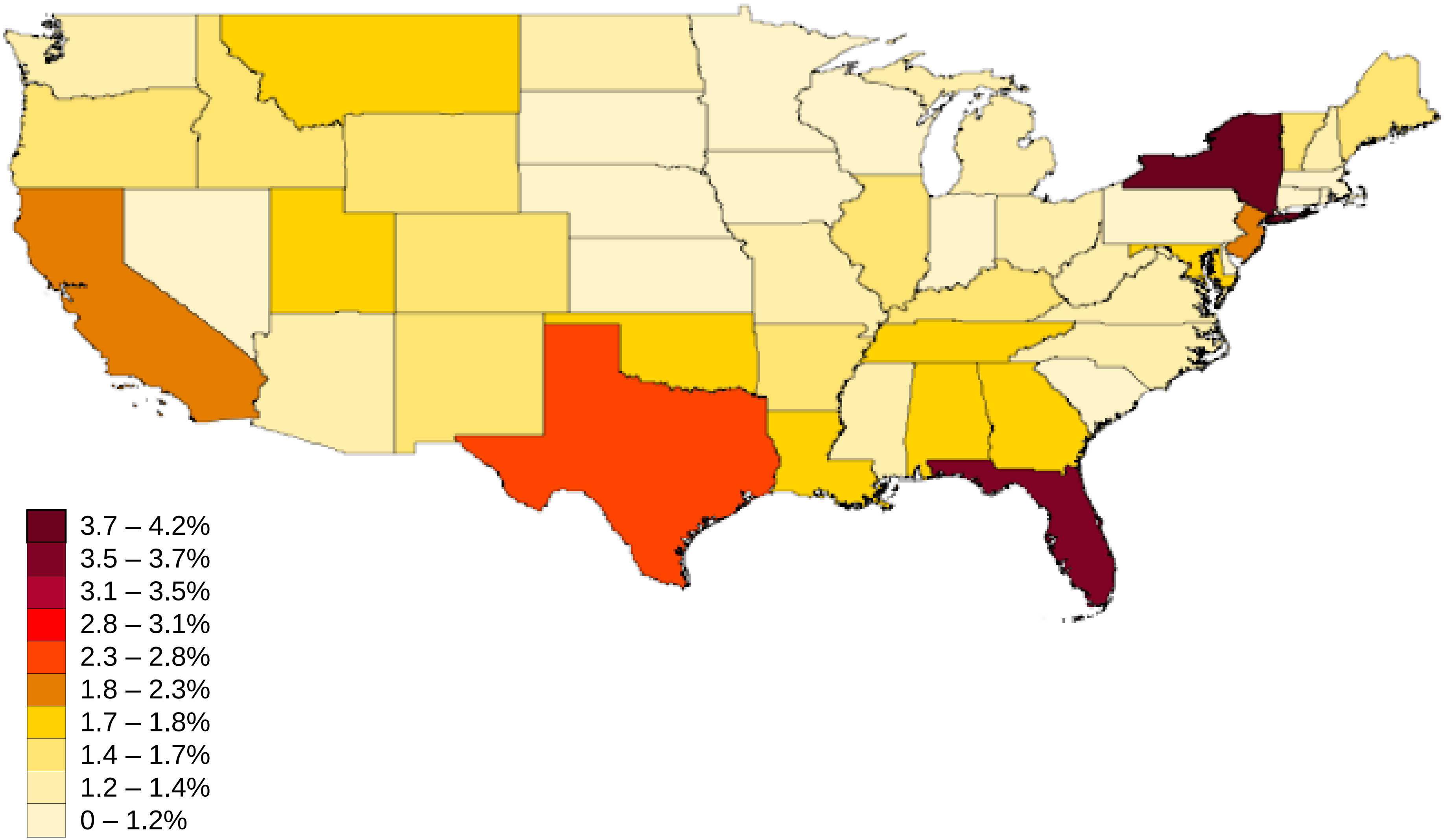
Neighborhood-Level Measure of Bunching

- Self-employed sharp bunching
  - Fraction of EITC-eligible tax filers who report income at first kink and have self-employment income
  - Essentially measures fraction of individuals who manipulate reported income to maximize EITC refund in each neighborhood
- Begin by examining spatial evolution of sharp self-employed bunching across the United States

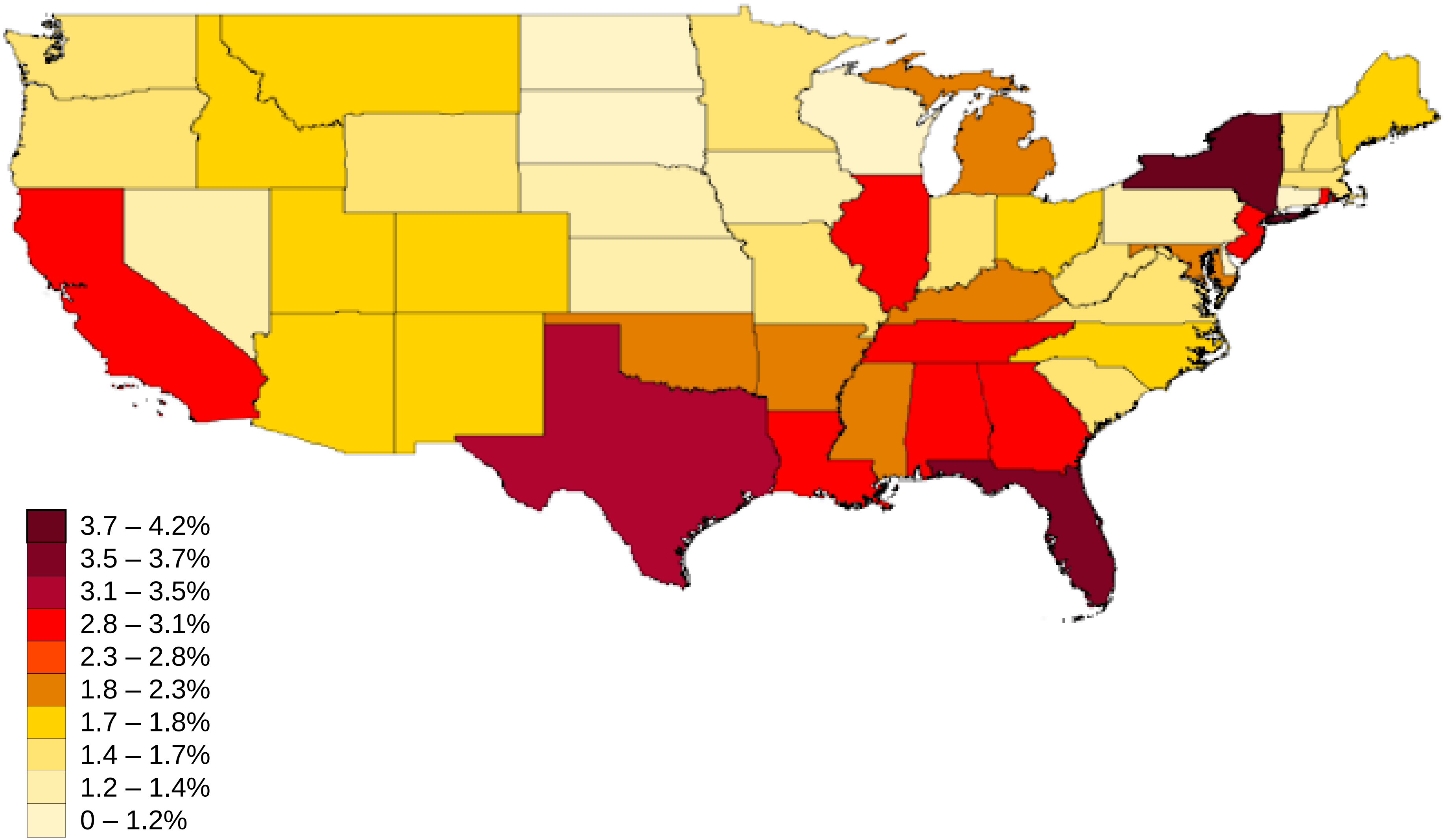
Self-Employed Sharp Bunching in 1996



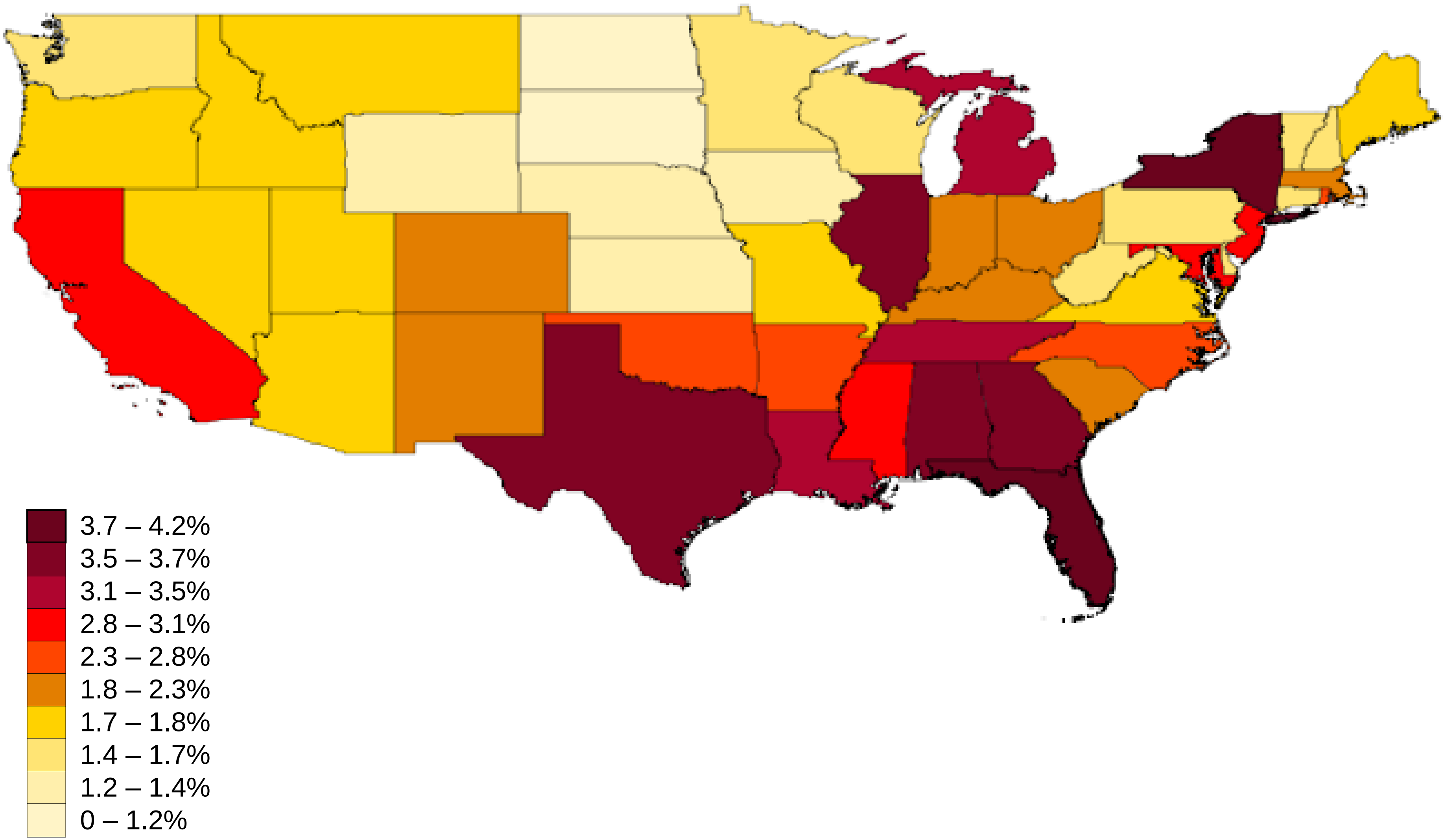
Self-Employed Sharp Bunching in 1999



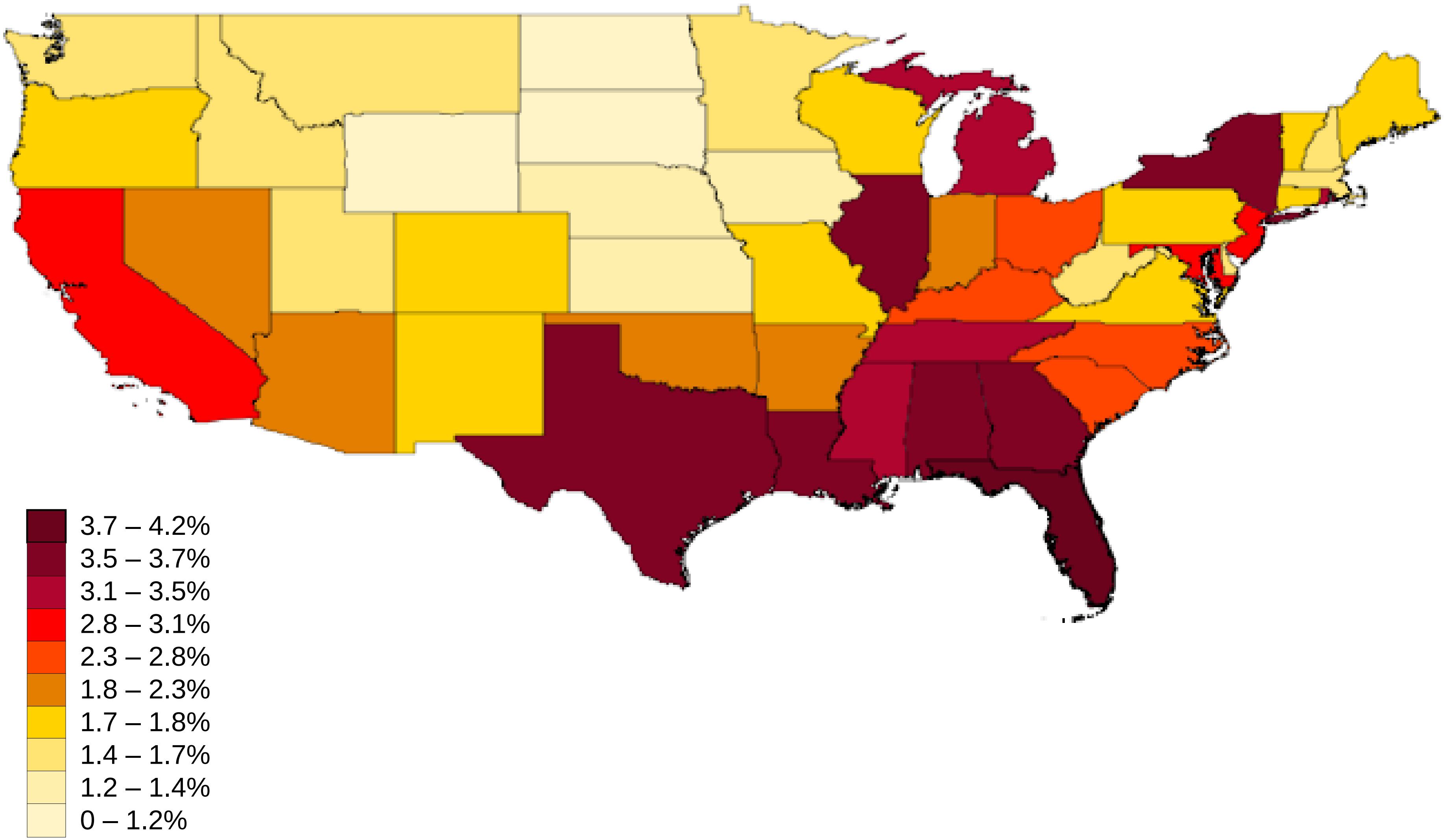
Self-Employed Sharp Bunching in 2002



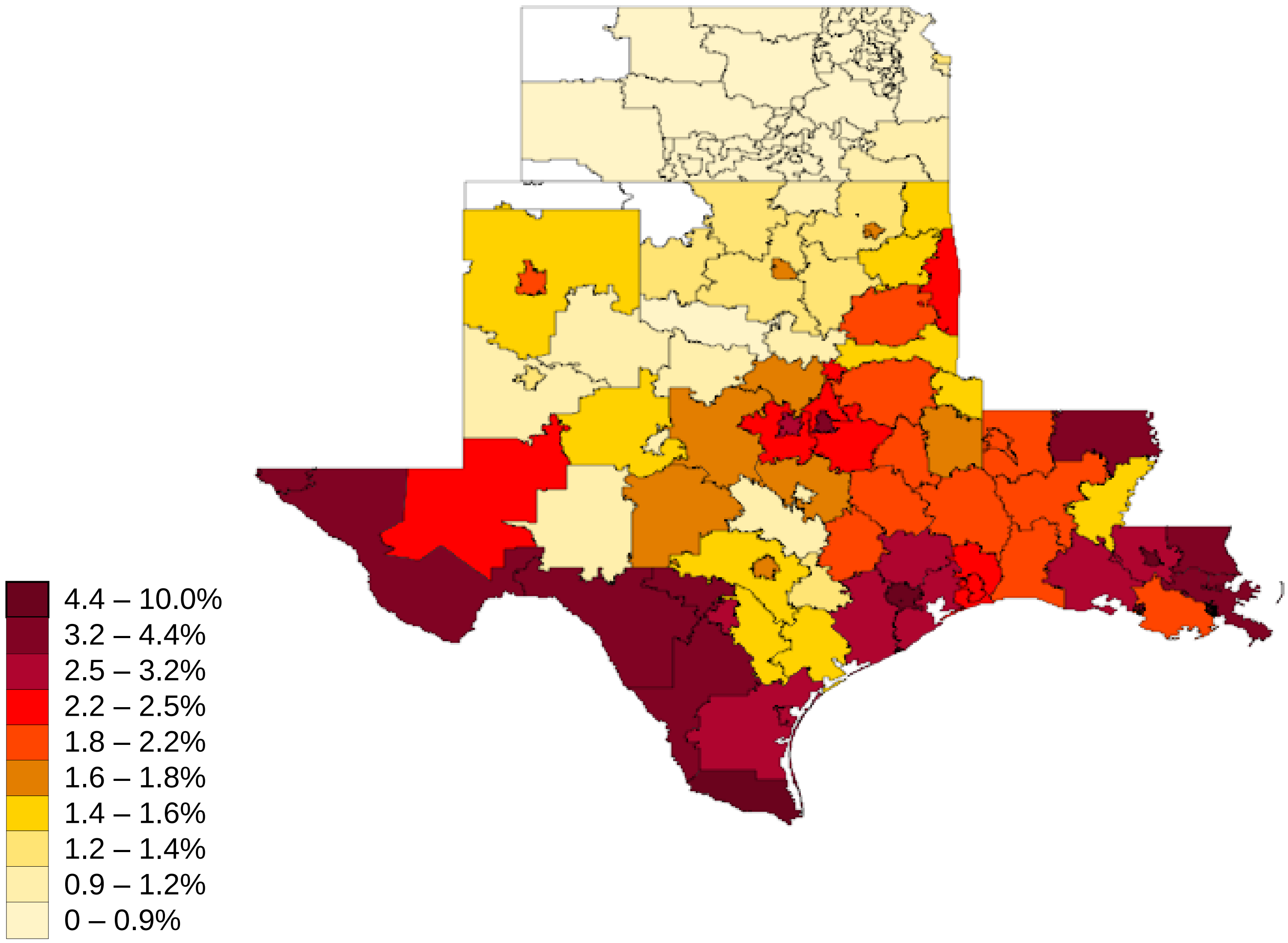
Self-Employed Sharp Bunching in 2005



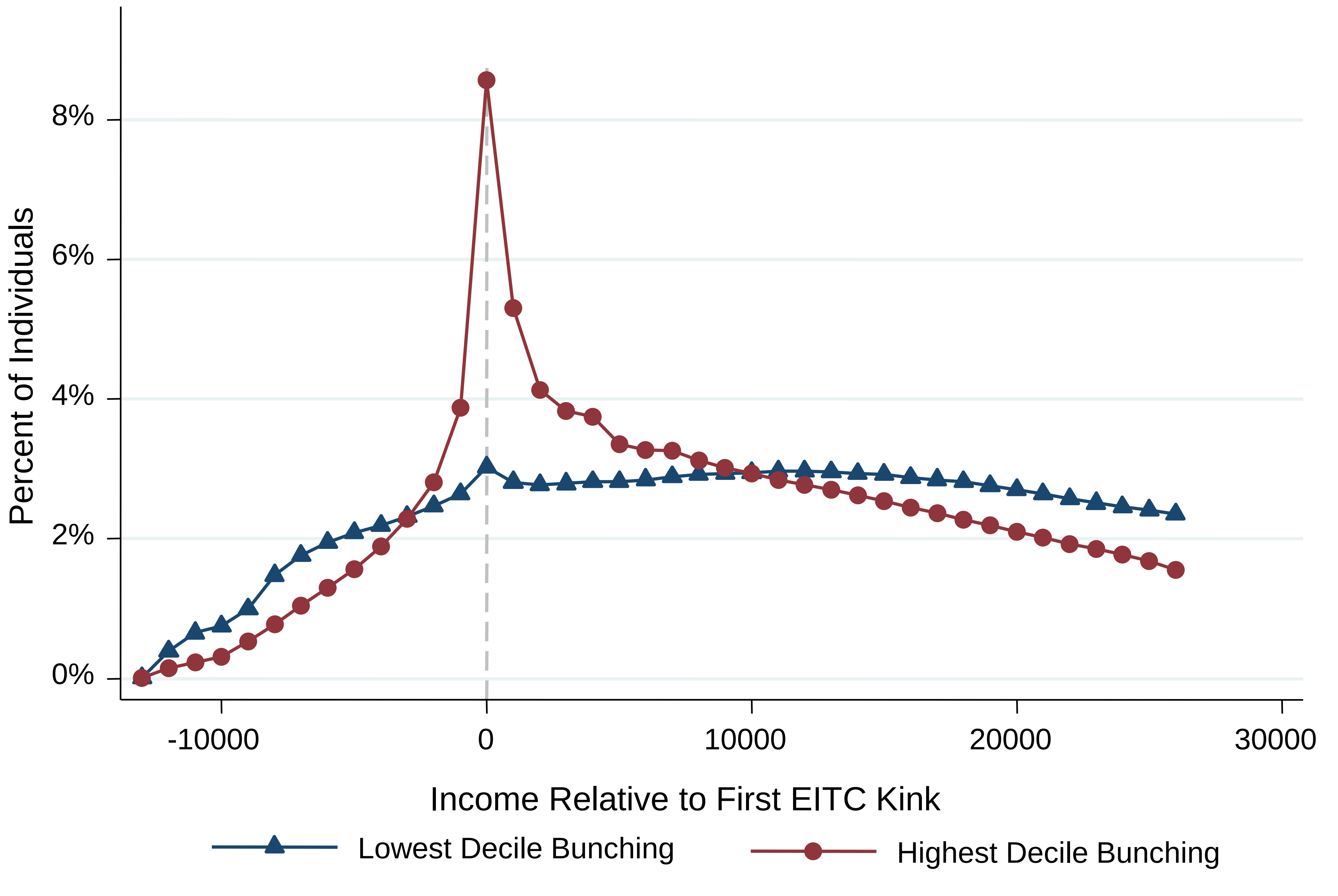
Self-Employed Sharp Bunching in 2008



Self-Employed Sharp Bunching in 2008
by 3-Digit Zip Code in Kansas, Louisiana, Oklahoma, and Texas



Income Distributions in Lowest vs. Highest Deciles of Sharp Bunching



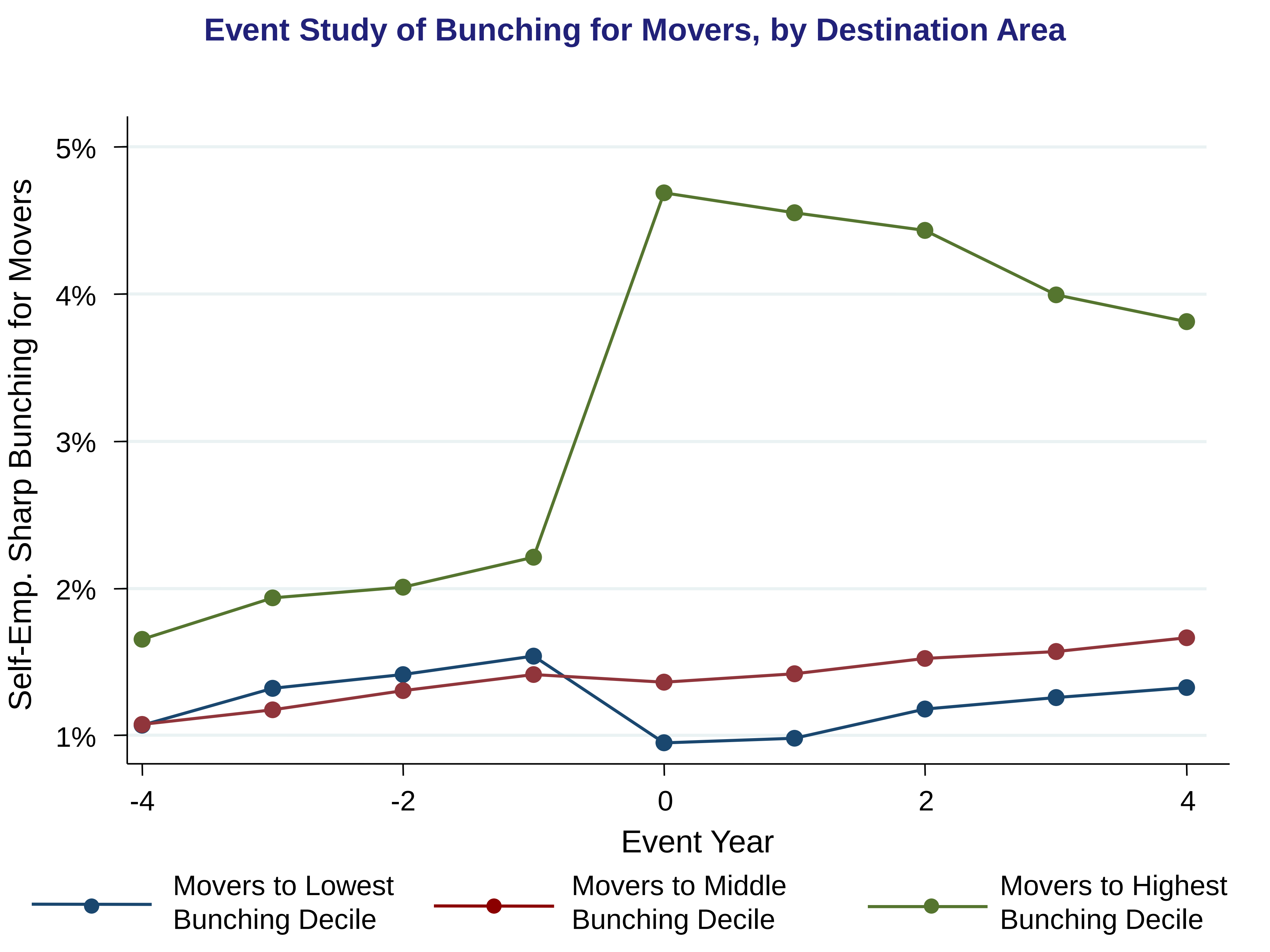
Outline of Empirical Analysis

- Step 1: Develop a proxy for knowledge about the EITC in each neighborhood using sharp bunching among self-employed

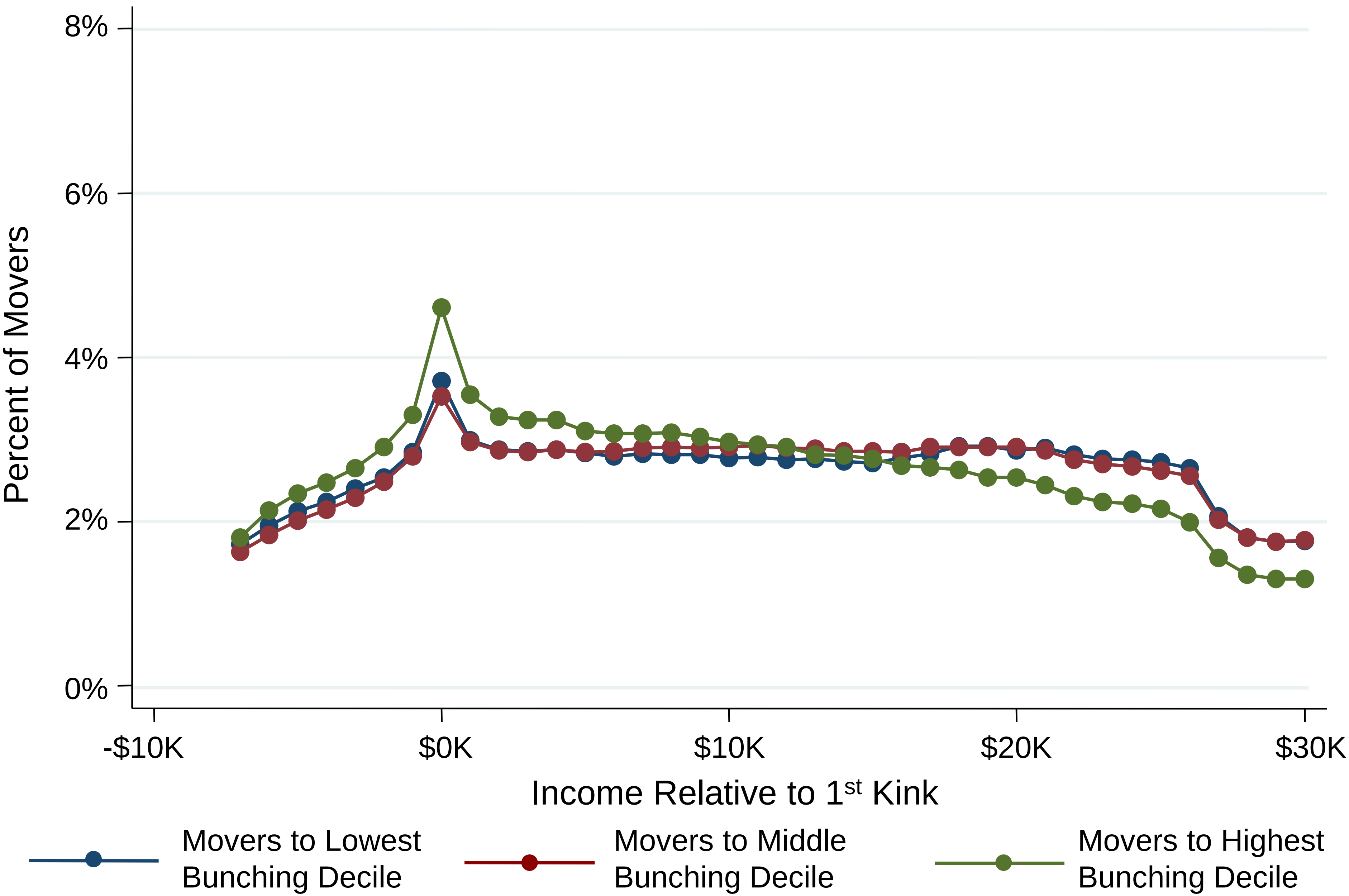
- Step 2: Establish learning as a mechanism for differences in sharp bunching across neighborhoods

Movers: Neighborhood Changes

- Look at individuals who move across neighborhoods to isolate causal impacts of neighborhoods on elasticities
  - 54 million observations in panel data on cross-zip movers
- Define "neighborhood sharp bunching" as degree of bunching for *stayers*
  - Classify movers based on deciles of neighborhood response of original neighborhood and new neighborhood


Event Study of Bunching for Movers, by Destination Area
Self-Emp. Sharp Bunching for Movers
5%
4%
3%
2%
1%
-4
-2
0
2
4
Event Year
Movers to Lowest Bunching Decile
Movers to Middle Bunching Decile
Movers to Highest Bunching Decile

Movers' Income Distributions: Before Move

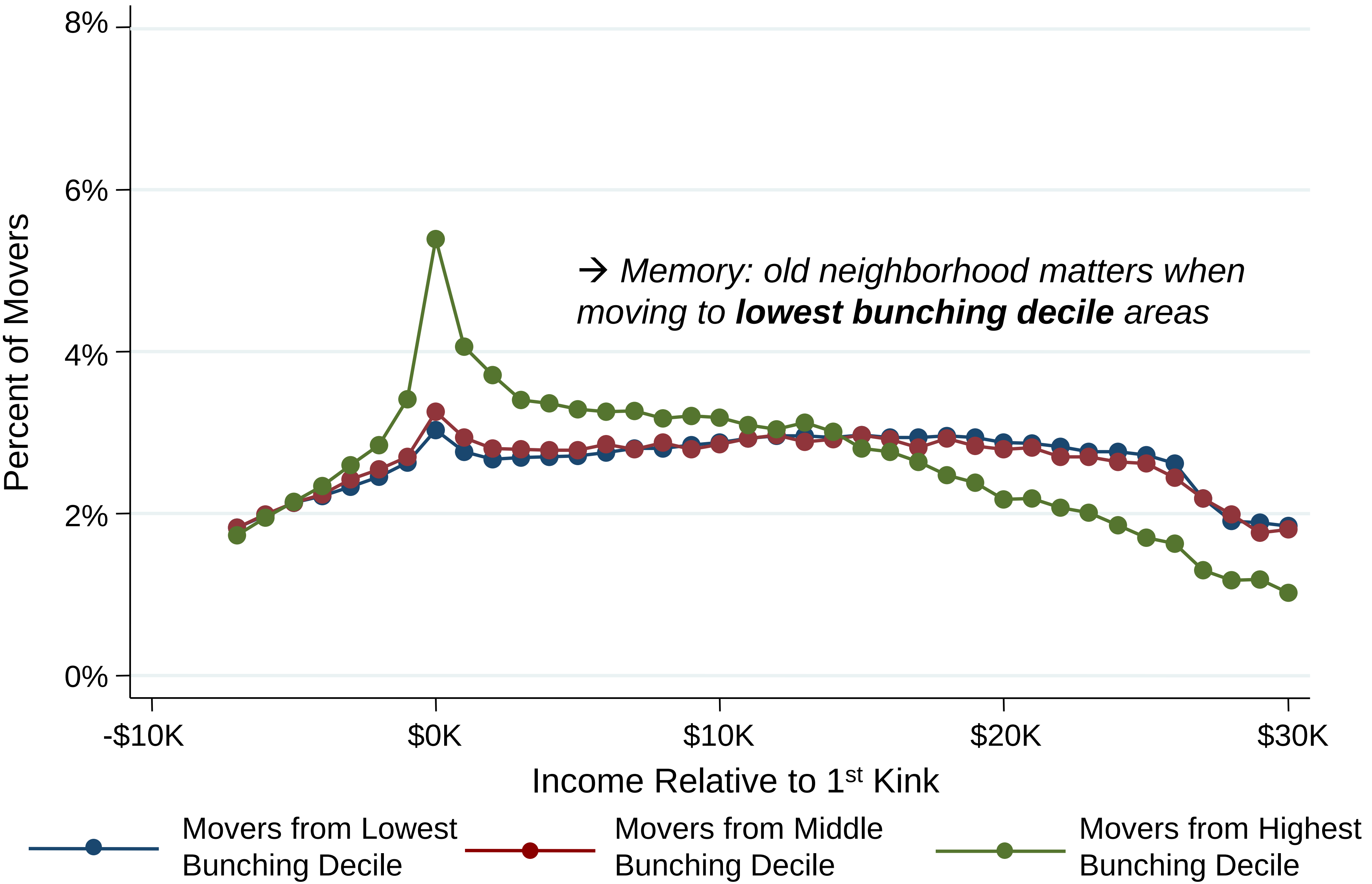



Movers' Income Distributions: After Move
Percent of Movers
8%
6%
4%
2%
0%
-$10K
$0K
$10K
$20K
$30K
Income Relative to 1st Kink
Movers to Lowest Bunching Decile
Movers to Middle Bunching Decile
Movers to Highest Bunching Decile

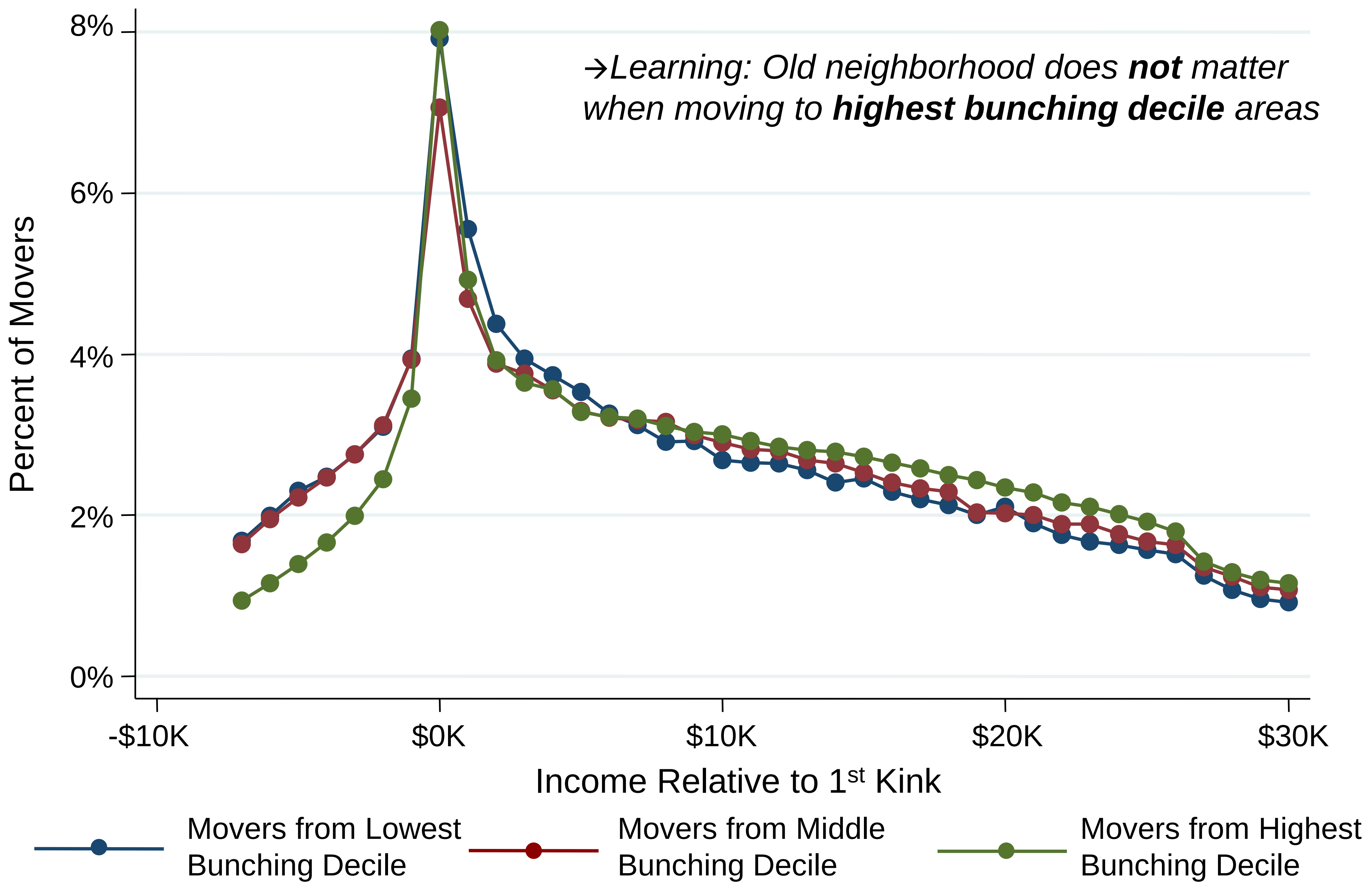
Learning and Asymmetry

- Knowledge model makes strong prediction about asymmetry of effects:
  - Memory: level of response in prior neighborhood should continue to matter for those who move to a low-EITC-response neighborhood
  - Learning: prior neighborhood matters less when moving to a high-EITC-response neighborhood

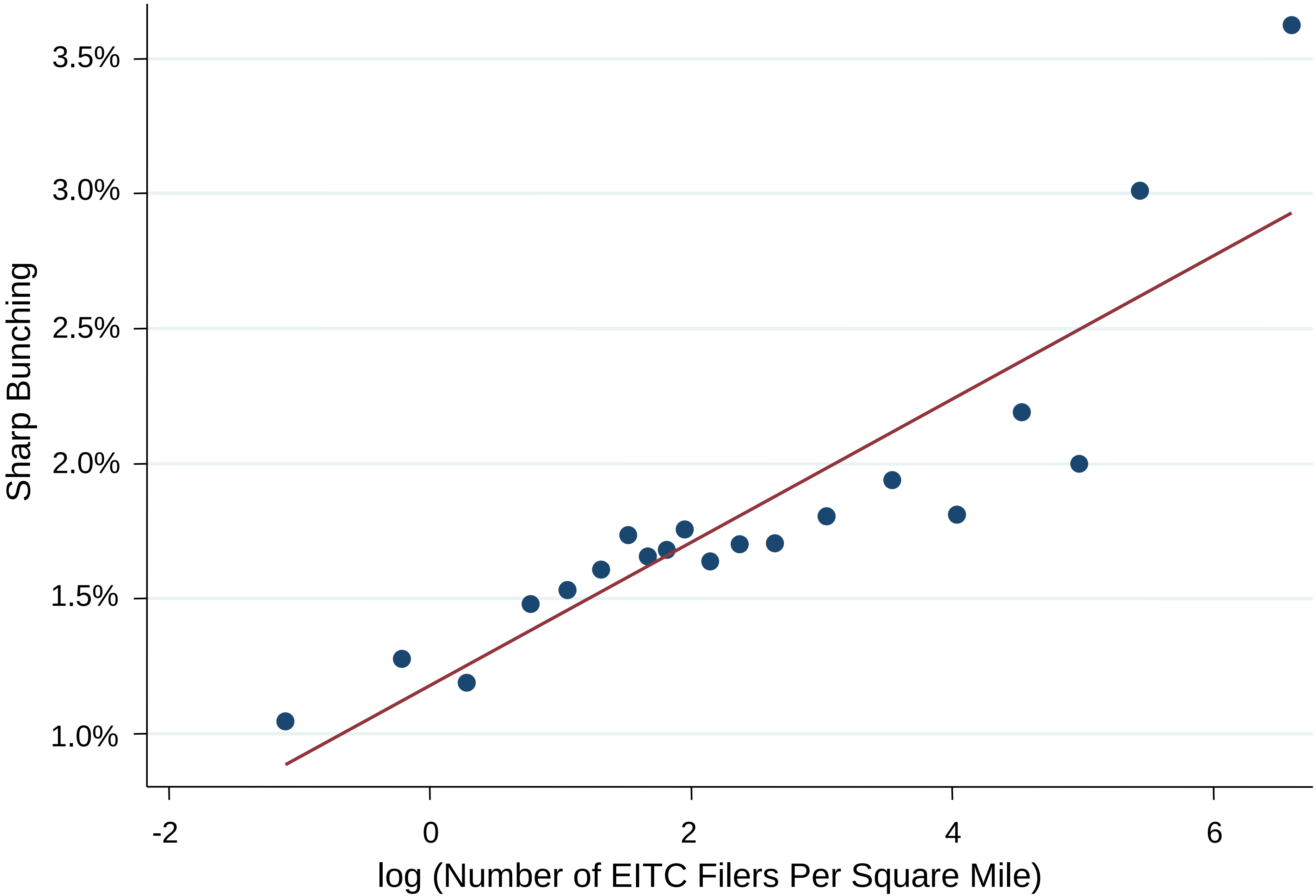
Post-Move Distributions for Movers to Lowest Bunching Decile Neighborhoods

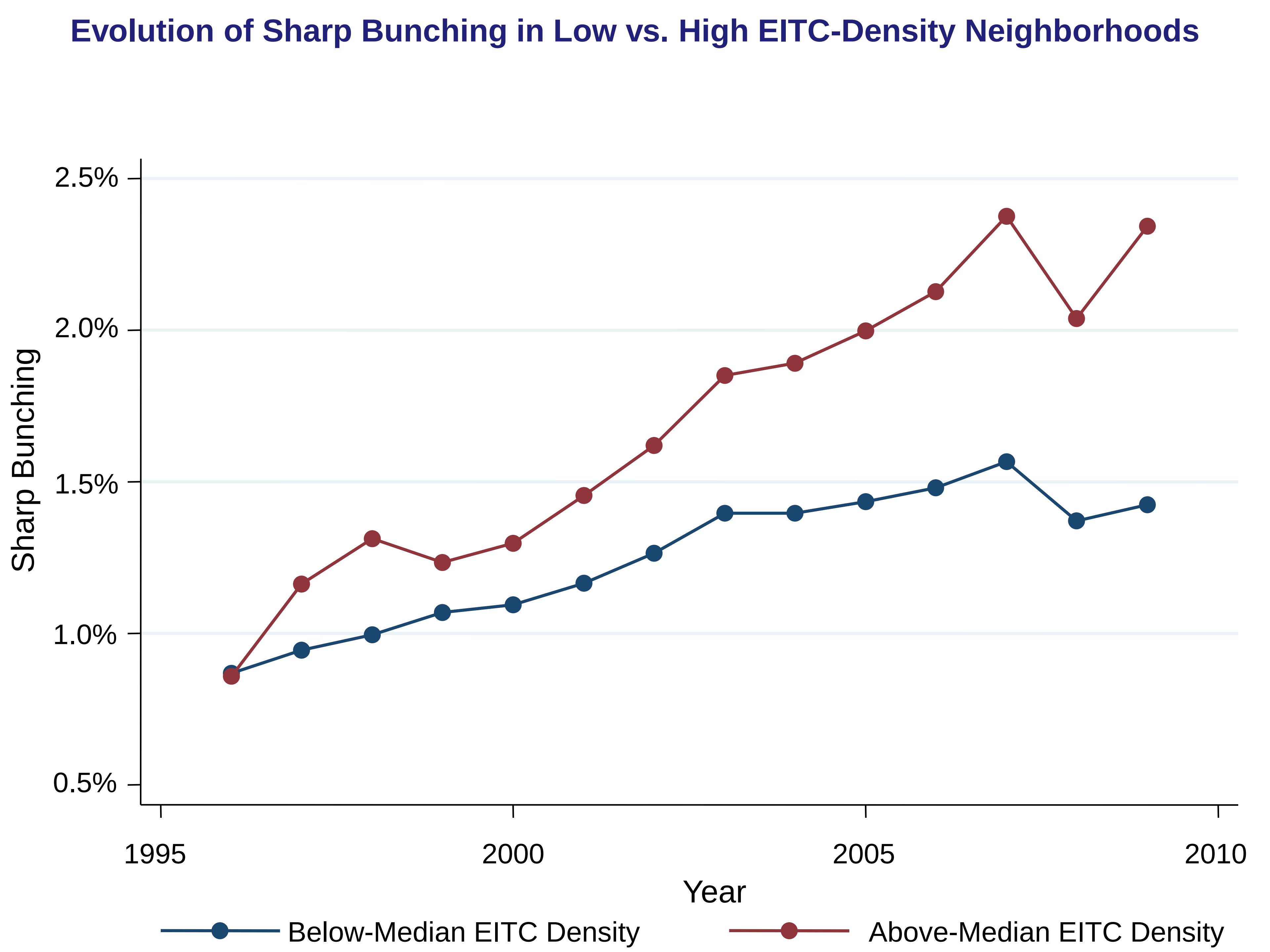


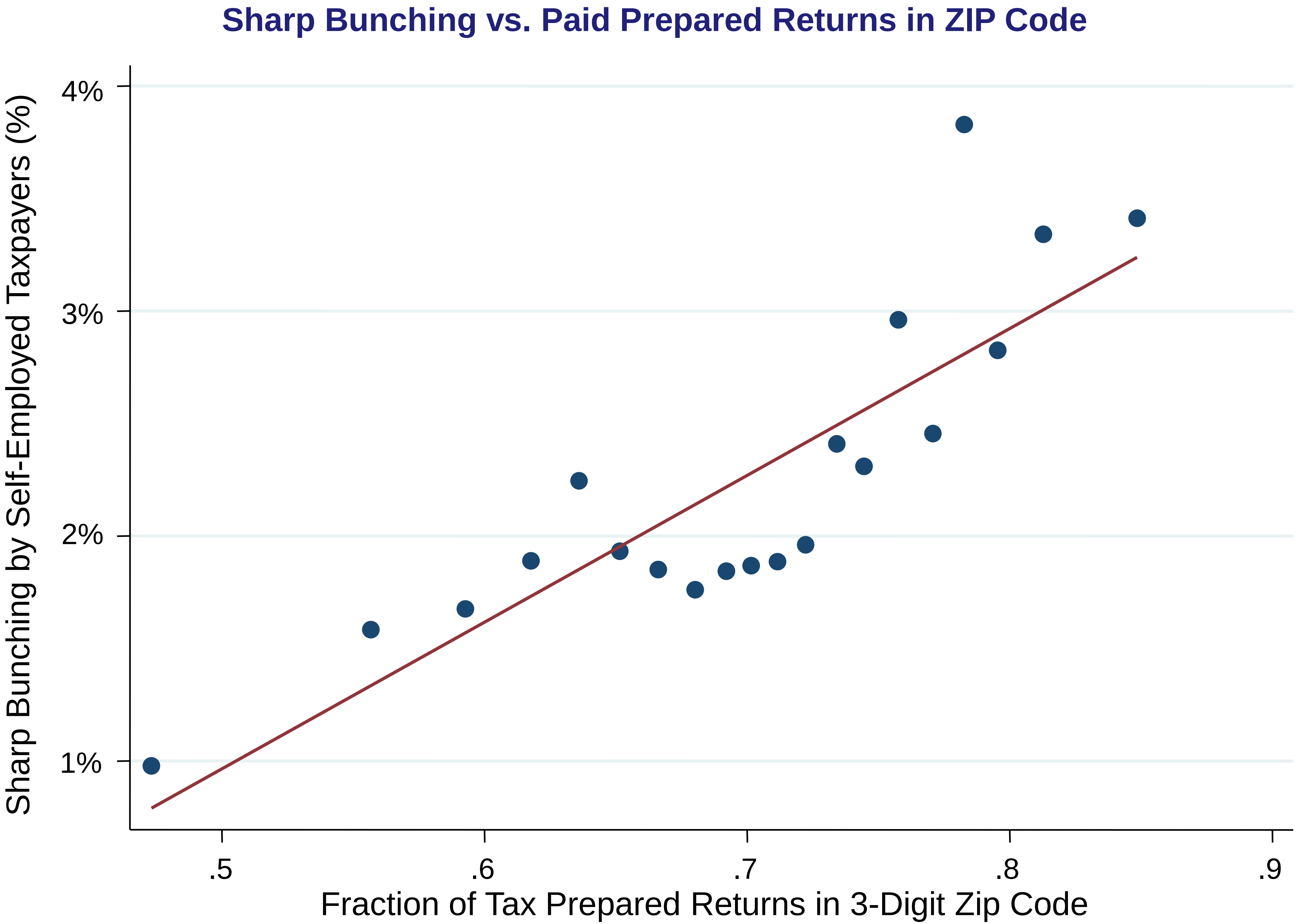
Post-Move Distributions for Movers to Highest Bunching Decile Neighborhoods

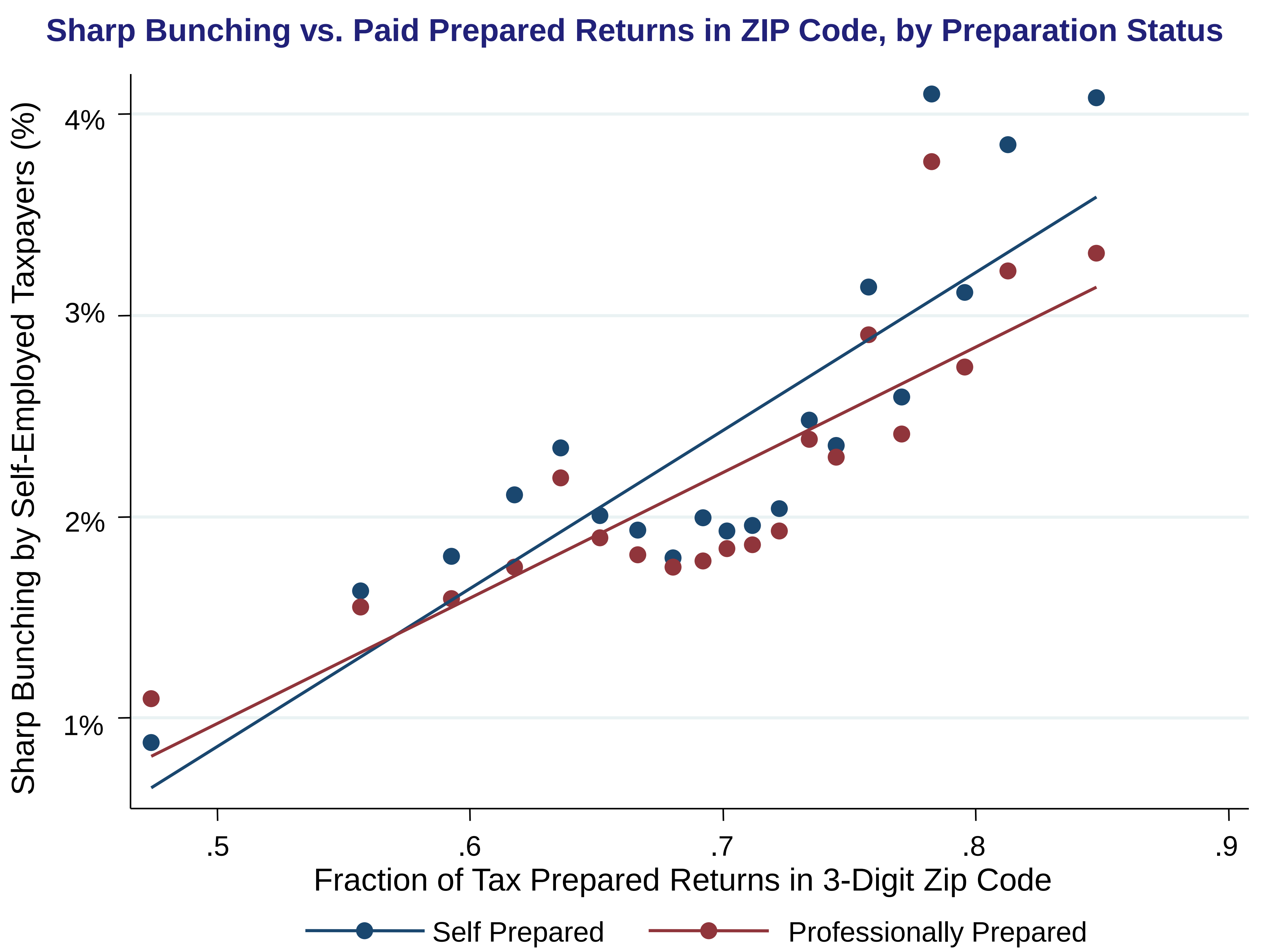


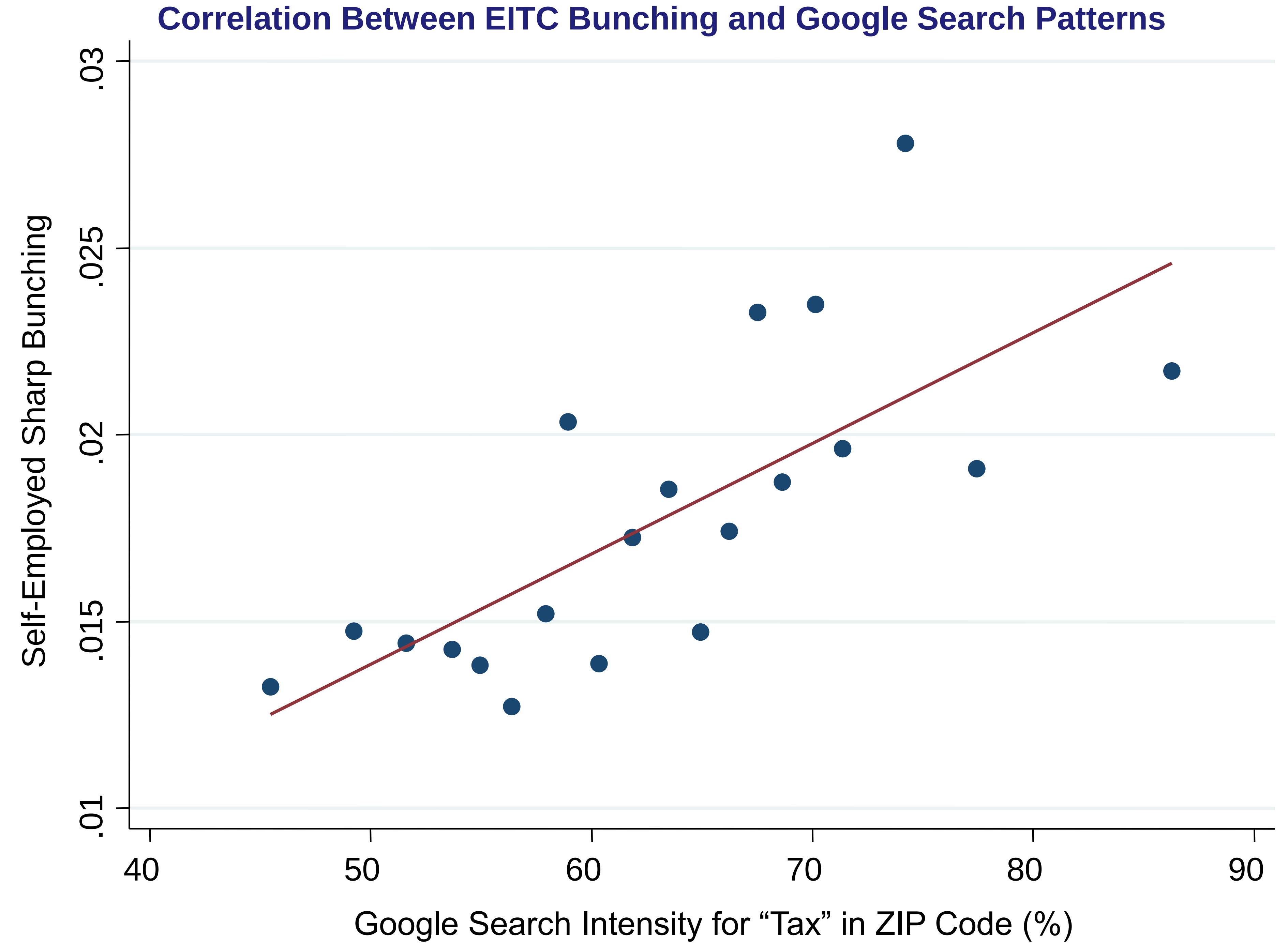
Agglomeration: Sharp Bunching vs. EITC Filer Density by ZIP Code




Evolution of Sharp Bunching in Low vs. High EITC-Density Neighborhoods
Sharp Bunching
2.5%
2.0%
1.5%
1.0%
0.5%
1995
2000
2005
2010
Year
Below-Median EITC Density
Above-Median EITC Density

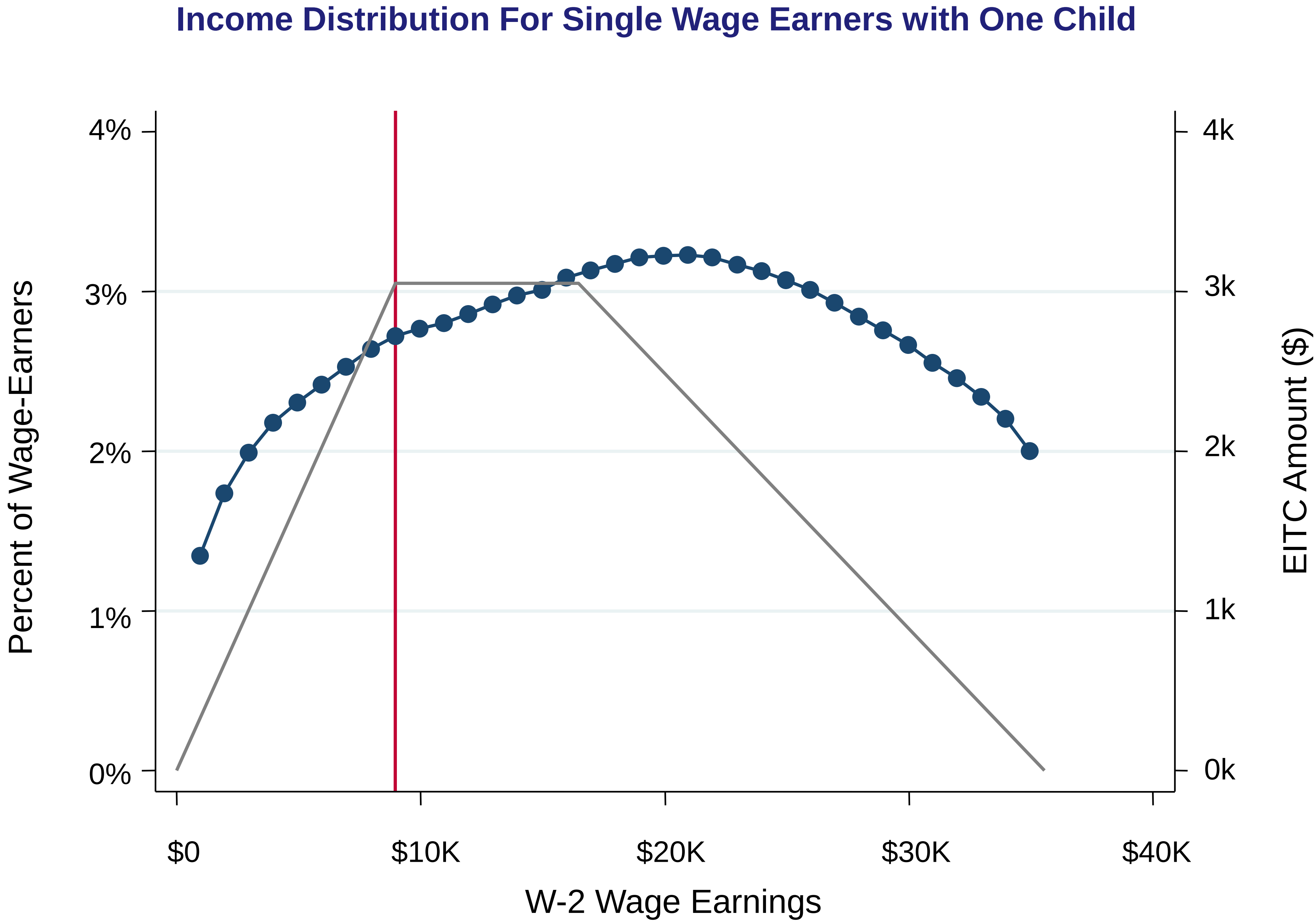

Sharp Bunching vs. Paid Prepared Returns in ZIP Code
Sharp Bunching by Self-Employed Taxpayers (%)
4%
3%
2%
1%
.5
.6
.7
.8
.9
Fraction of Tax Prepared Returns in 3-Digit Zip Code


Sharp Bunching vs. Paid Prepared Returns in ZIP Code, by Preparation Status
Sharp Bunching by Self-Employed Taxpayers (%)
4%
3%
2%
1%
.5
.6
.7
.8
.9
Fraction of Tax Prepared Returns in 3-Digit Zip Code
Self Prepared
Professionally Prepared

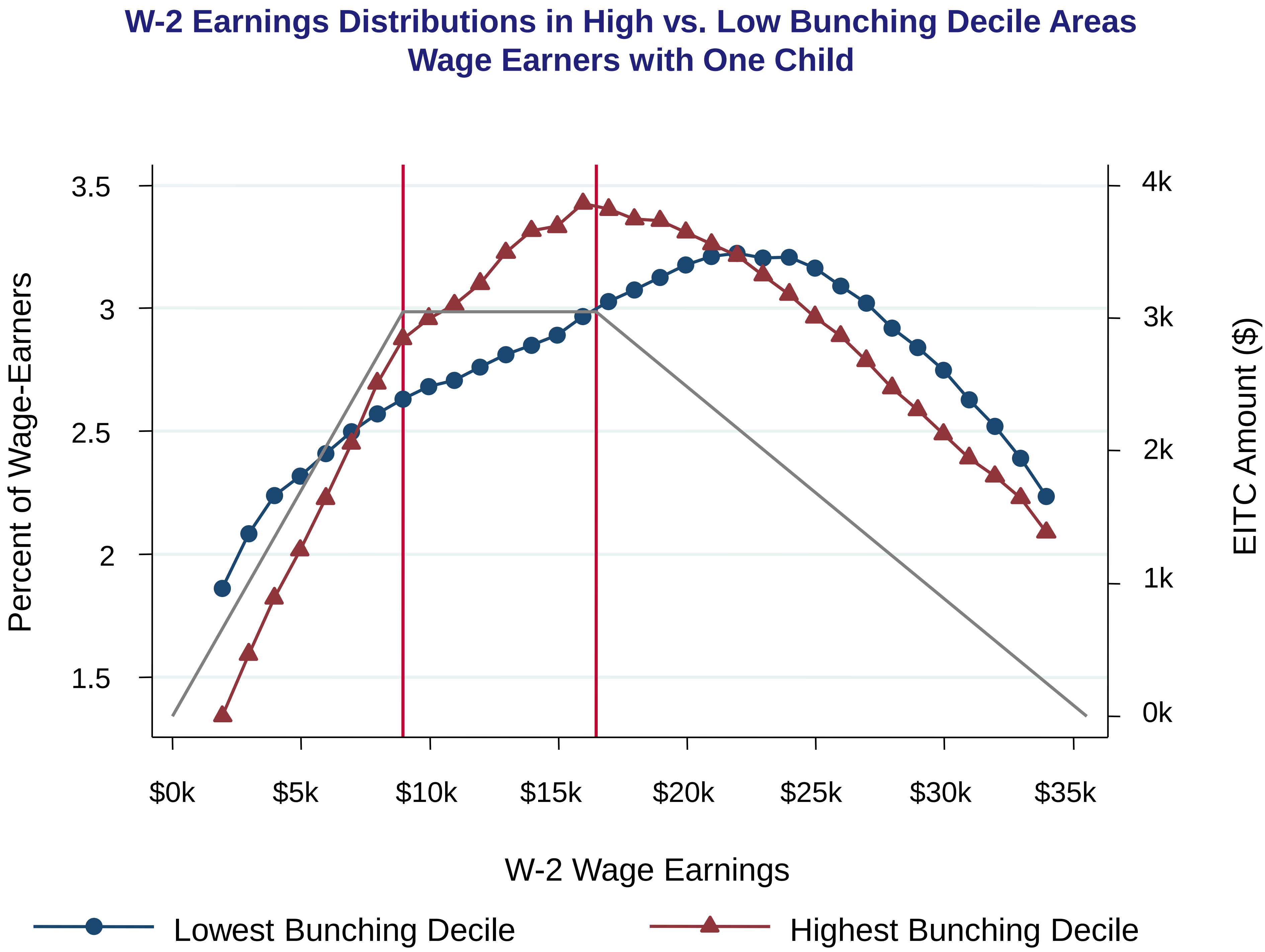

Correlation Between EITC Bunching and Google Search Patterns
Self-Employed Sharp Bunching
.01
.015
.02
.025
.03
40
50
60
70
80
90
Google Search Intensity for “Tax” in ZIP Code (%)

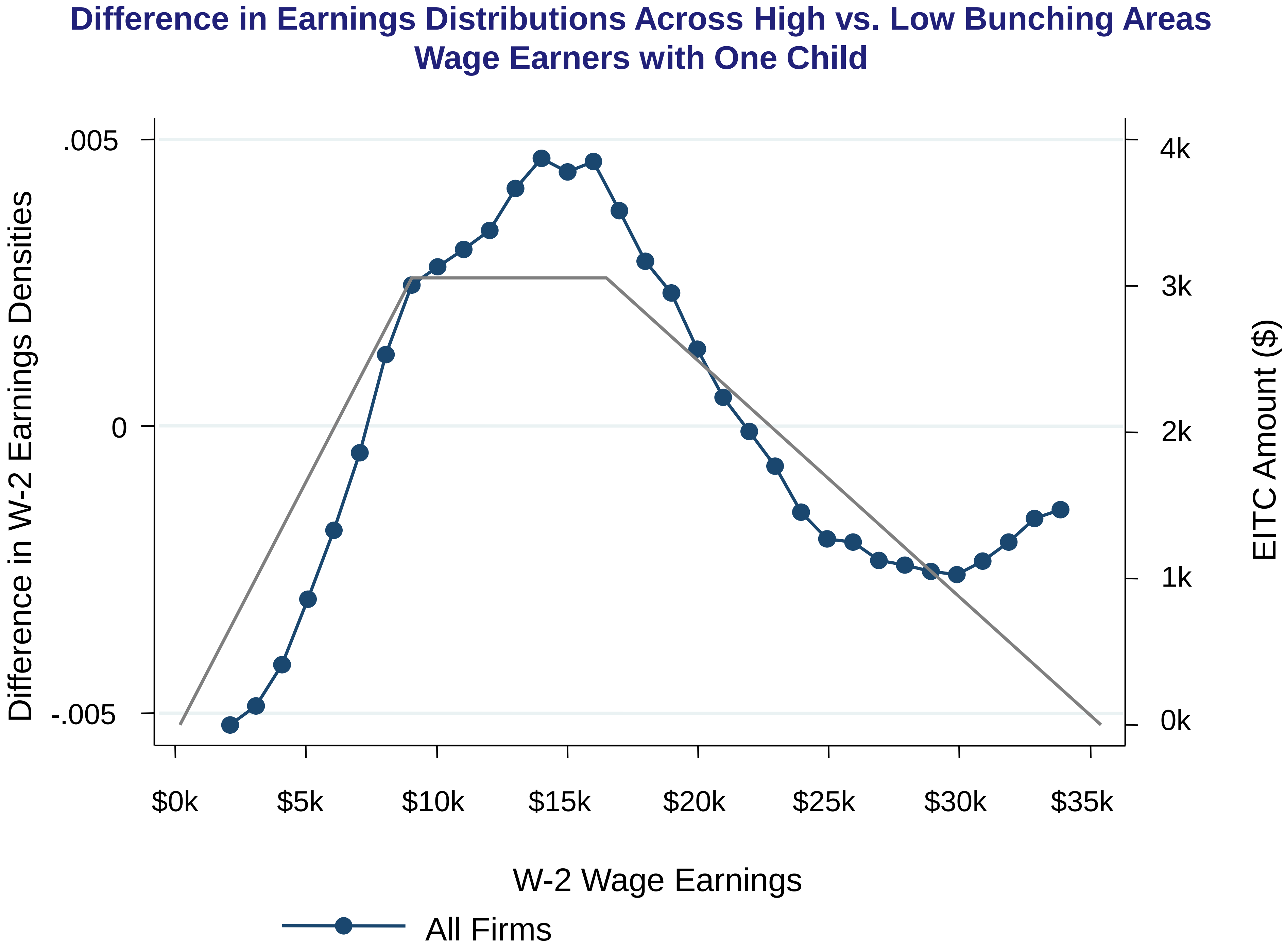
Outline of Empirical Analysis

- Step 1: Develop a proxy for knowledge about the EITC in each neighborhood using sharp bunching among self-employed
- Step 2: Establish learning as a mechanism for differences in sharp bunching across neighborhoods
- Step 3: Compare wage earnings distributions across low- and high-knowledge neighborhoods to uncover impacts of EITC on earnings

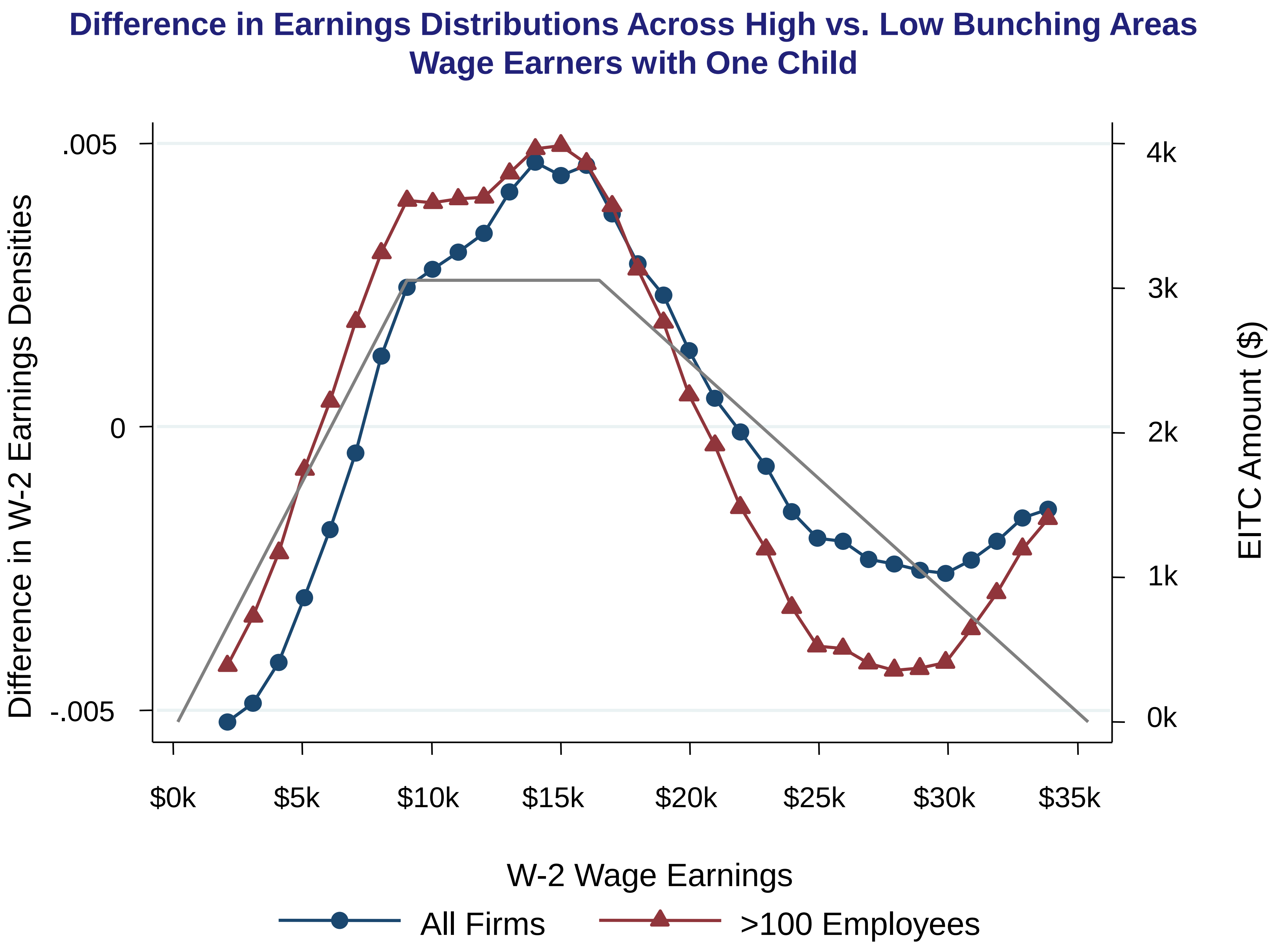


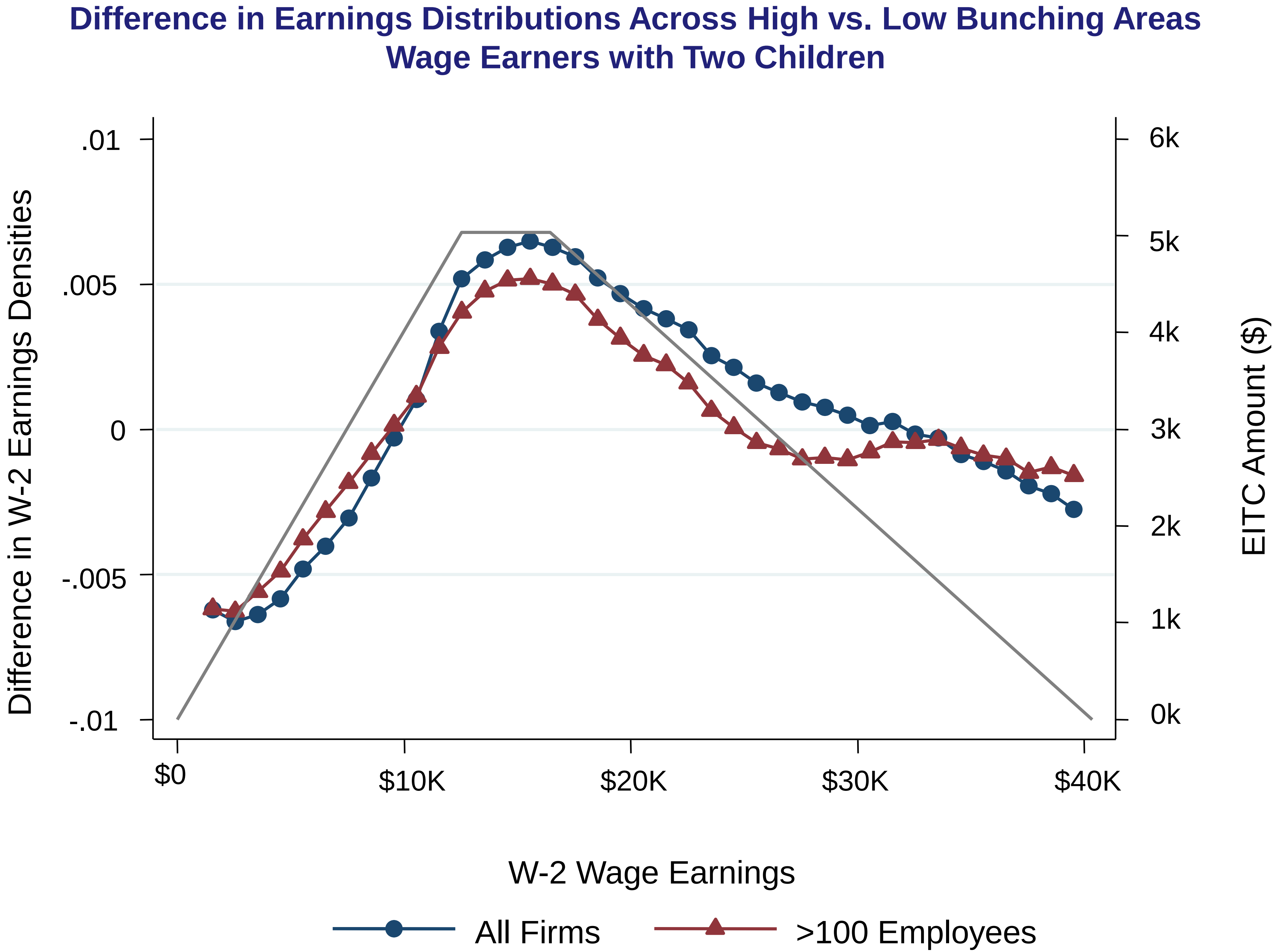
Income Distribution For Single Wage Earners with One Child
Percent of Wage-Earners
EITC Amount ($)
W-2 Wage Earnings
4%
3%
2%
1%
0%
4k
3k
2k
1k
0k
$0
$10K
$20K
$30K
$40K

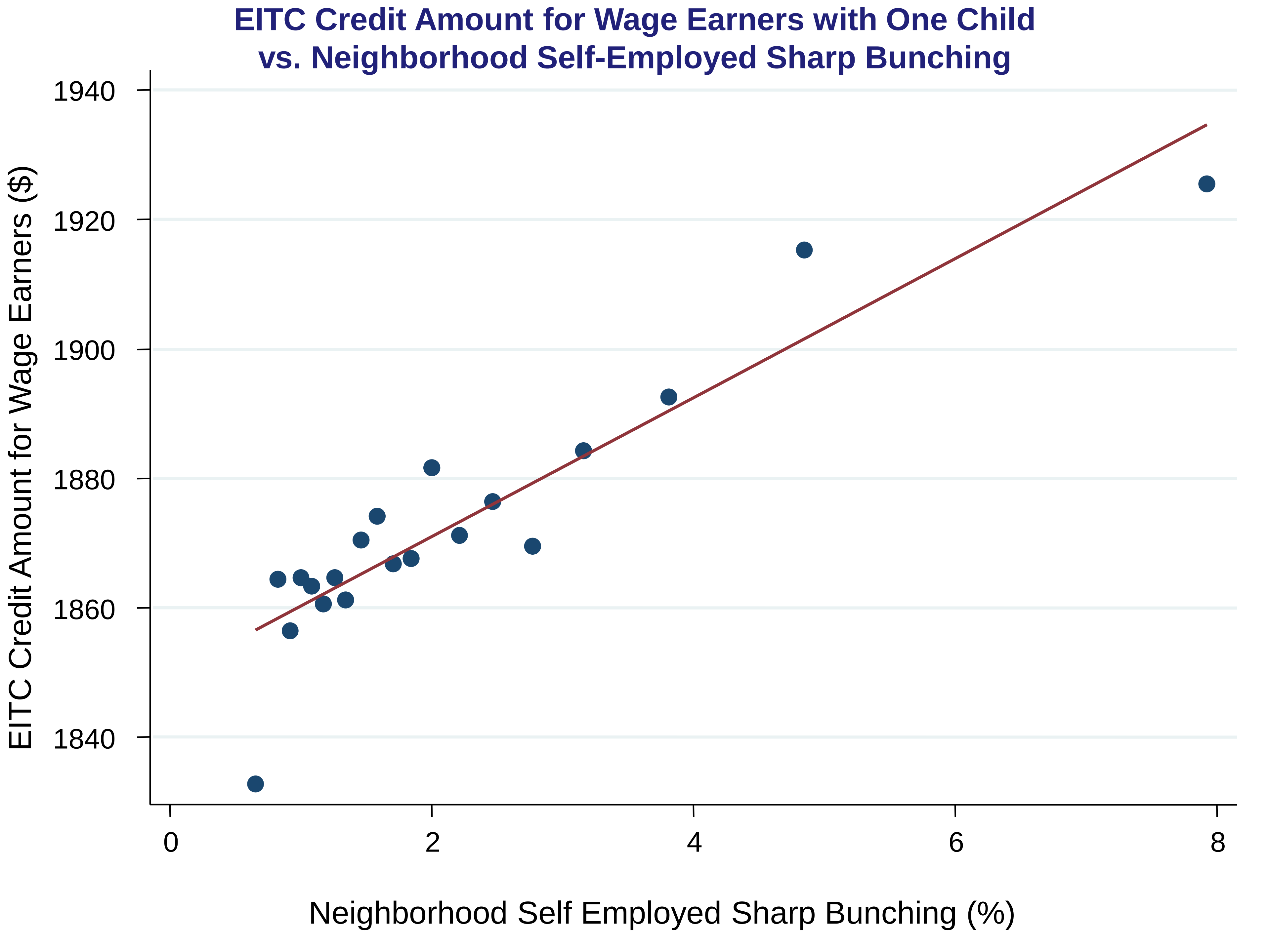

W-2 Earnings Distributions in High vs. Low Bunching Decile Areas
Wage Earners with One Child
Percent of Wage-Earners
3.5
3
2.5
2
1.5
EITC Amount ($)
4k
3k
2k
1k
0k
$0k
$5k
$10k
$15k
$20k
$25k
$30k
$35k
W-2 Wage Earnings
Lowest Bunching Decile
Highest Bunching Decile



Difference in Earnings Distributions Across High vs. Low Bunching Areas
Wage Earners with One Child
Difference in W-2 Earnings Densities
.005
0
-.005
EITC Amount ($)
4k
3k
2k
1k
0k
$0k
$5k
$10k
$15k
$20k
$25k
$30k
$35k
W-2 Wage Earnings
All Firms


Difference in Earnings Distributions Across High vs. Low Bunching Areas
Wage Earners with One Child
.005
0
-.005
Difference in W-2 Earnings Densities
4k
3k
2k
1k
0k
EITC Amount ($)
$0k
$5k
$10k
$15k
$20k
$25k
$30k
$35k
W-2 Wage Earnings
All Firms
>100 Employees


Difference in Earnings Distributions Across High vs. Low Bunching Areas
Wage Earners with Two Children
Difference in W-2 Earnings Densities
.01
.005
0
-.005
-.01
EITC Amount ($)
6k
5k
4k
3k
2k
1k
0k
$0
$10K
$20K
$30K
$40K
W-2 Wage Earnings
All Firms
>100 Employees


EITC Credit Amount for Wage Earners with One Child
vs. Neighborhood Self-Employed Sharp Bunching
EITC Credit Amount for Wage Earners ($)
1940
1920
1900
1880
1860
1840
0
2
4
6
8
Neighborhood Self Employed Sharp Bunching (%)

Outline of Empirical Analysis

- Step 1: Develop a proxy for knowledge about the EITC in each neighborhood using sharp bunching among self-employed
- Step 2: Establish learning as a mechanism for differences in sharp bunching across neighborhoods
- Step 3: Compare wage earnings distributions across low- and high-knowledge neighborhoods to uncover impacts of EITC on earnings
- Step 4: Compare impacts of *changes* in EITC subsidies on earnings across low vs. high knowledge nbhds. to account for omitted variables

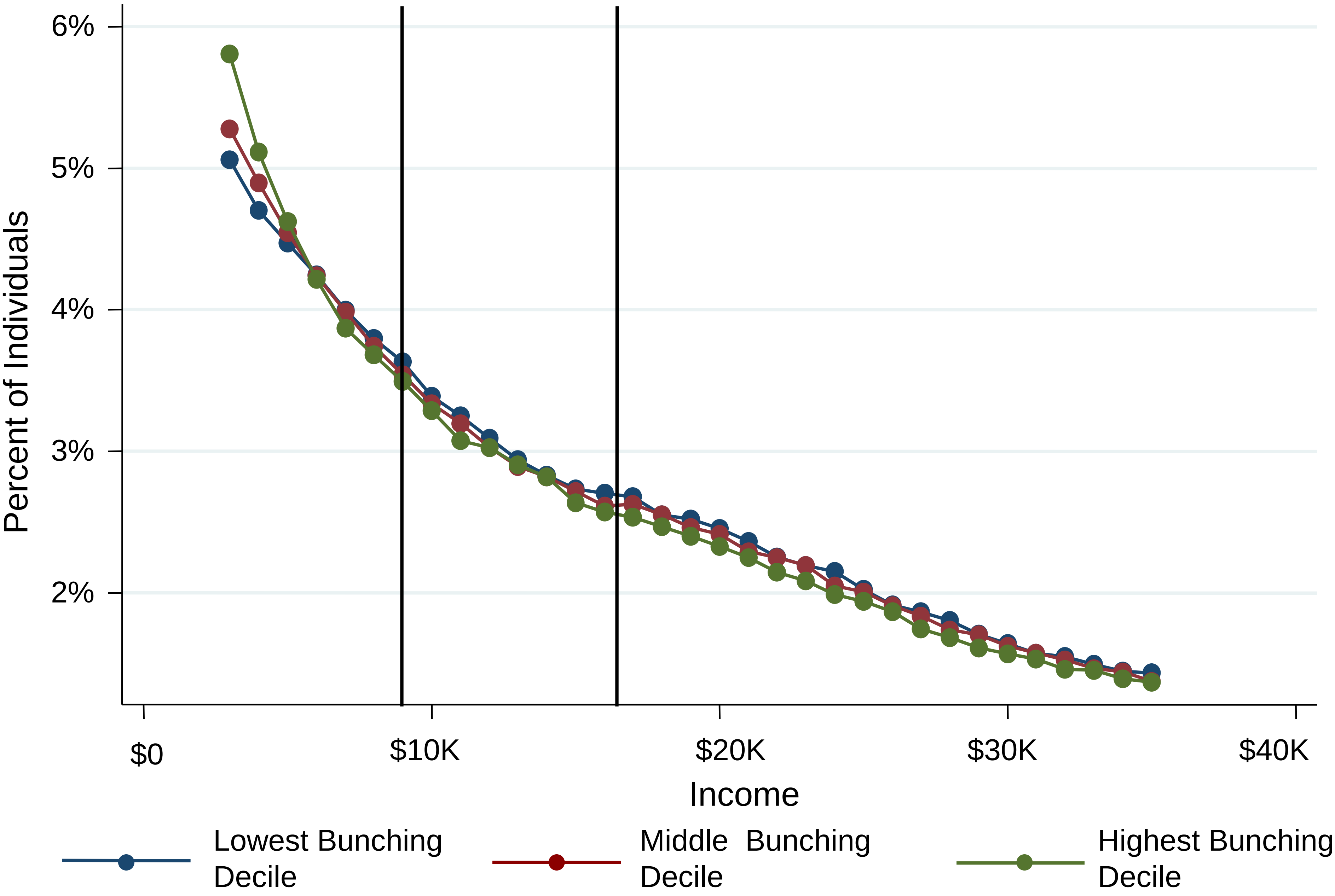
Accounting for Omitted Variables: Tax Changes

- Cross-sectional differences in income distributions could be biased by omitted variables
    - City effects: differences in industry structure or labor demand
    - Individual sorting: preferences may vary across cities
- We account for these omitted variables by analyzing impacts of changes in EITC subsidy
    - Do EITC changes affect earnings more in high knowledge cities?

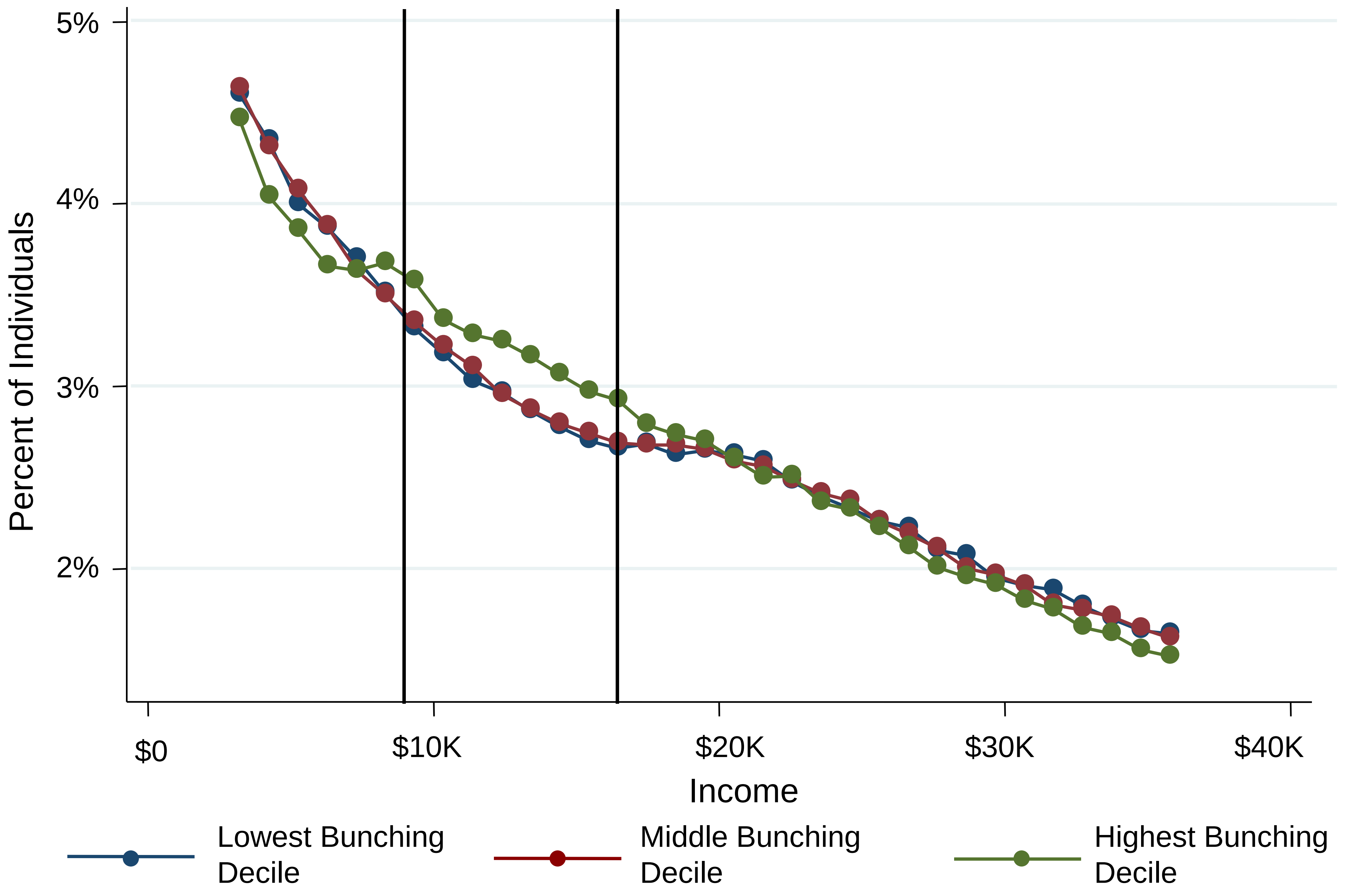
Child Birth as a Source of Tax Variation

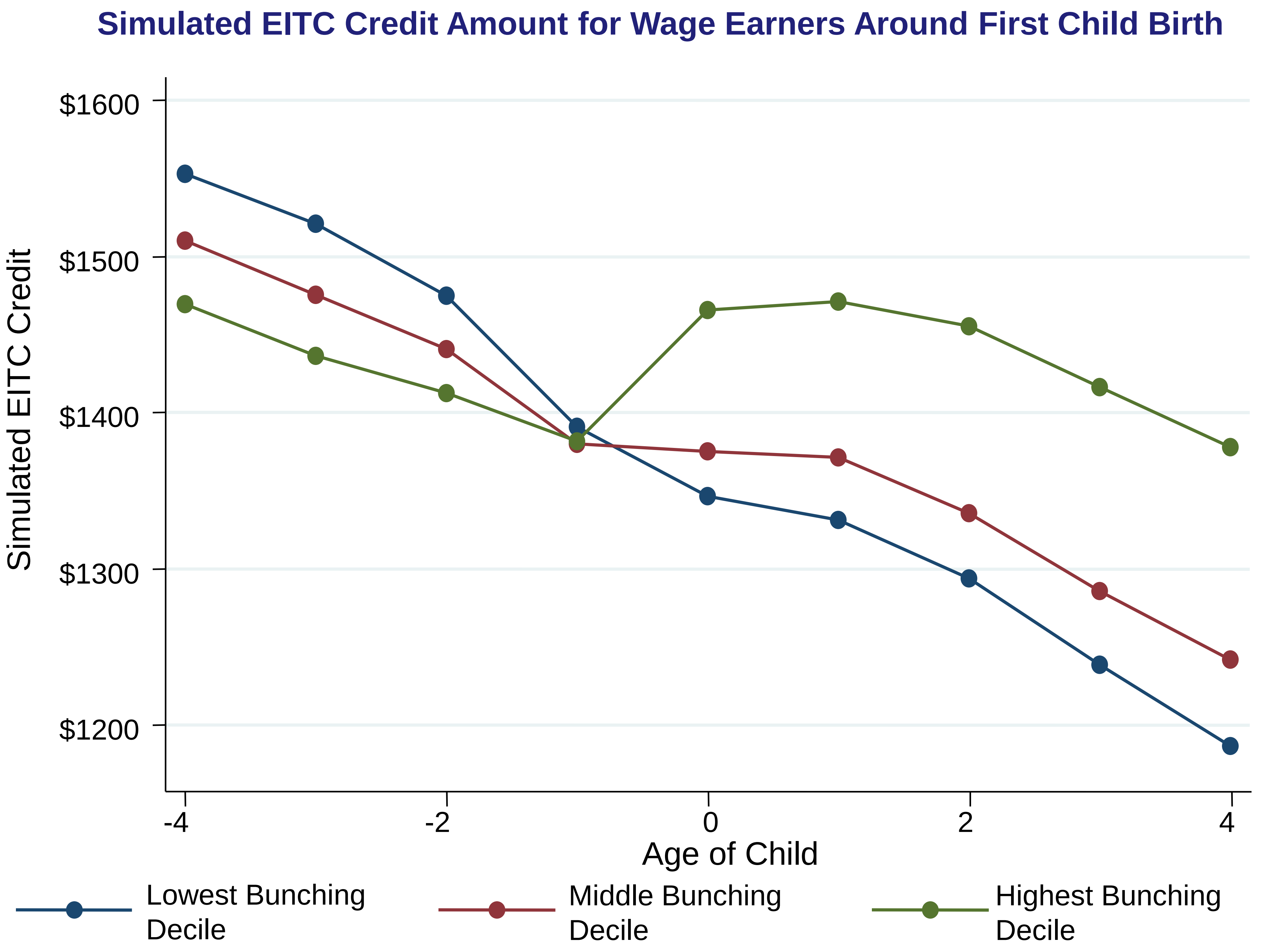
- To identify causal impacts of EITC, need variation in tax incentives
  - Birth of first child → substantial change in EITC incentives
  - Although birth affects labor supply directly, cross-neighborhood comparisons provide good counterfactuals
- 12 million EITC-eligible individuals give birth within our sample

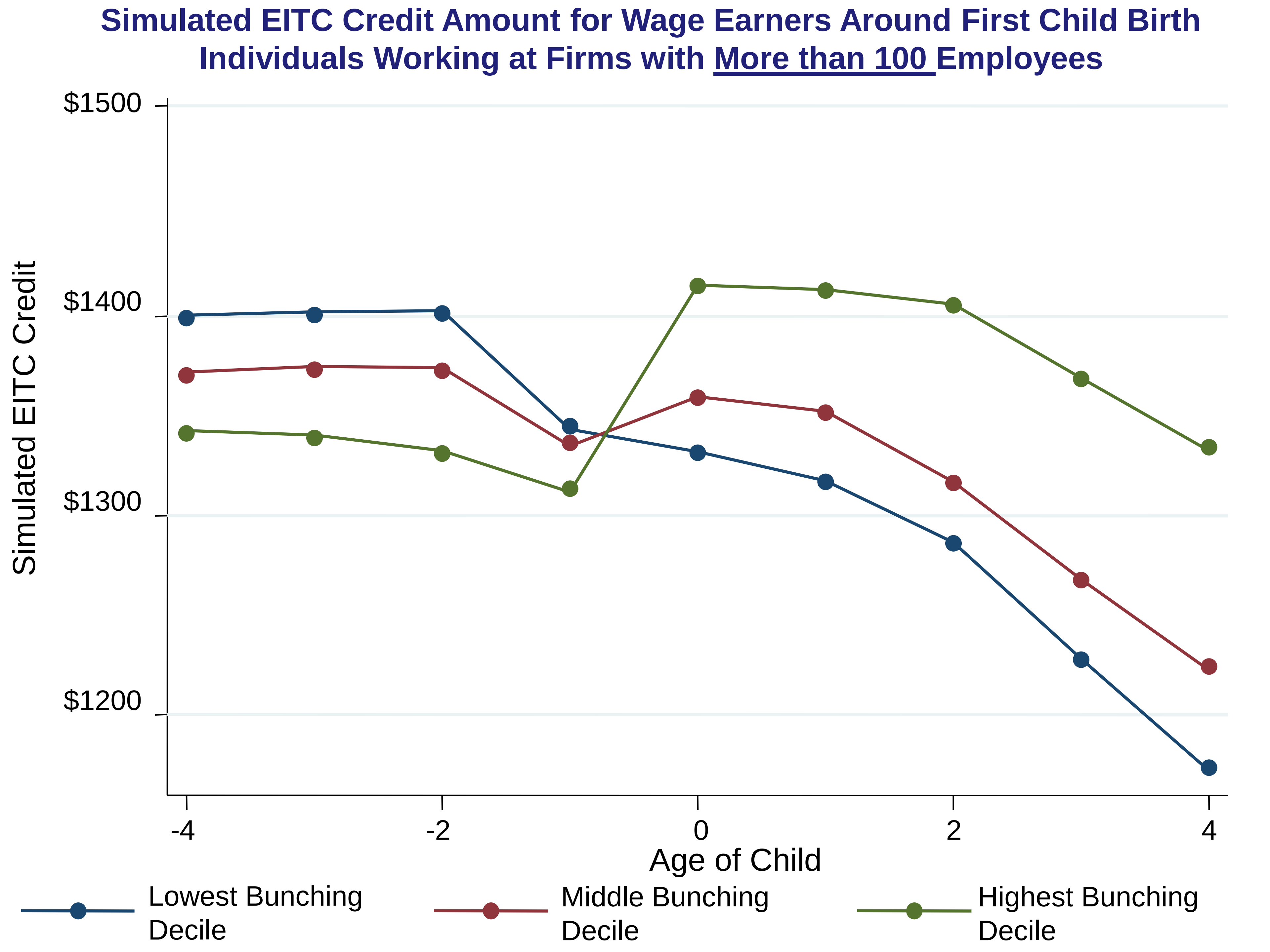
Earnings Distribution in the Year Before First Child Birth for Wage Earners

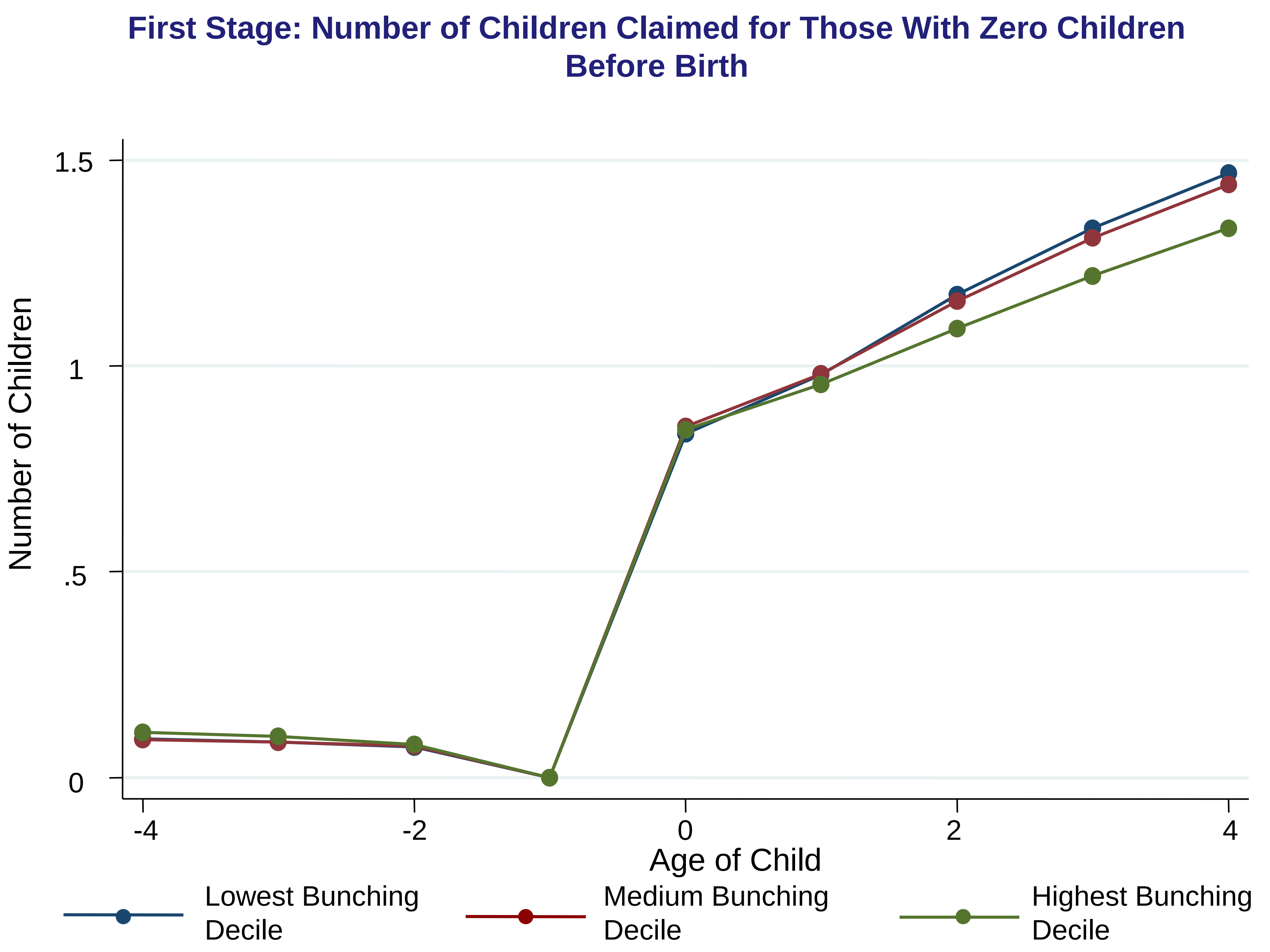


Earnings Distribution in the Year of First Child Birth for Wage Earners




Simulated EITC Credit Amount for Wage Earners Around First Child Birth
Simulated EITC Credit
$1600
$1500
$1400
$1300
$1200
-4
-2
0
2
4
Age of Child
Lowest Bunching Decile
Middle Bunching Decile
Highest Bunching Decile

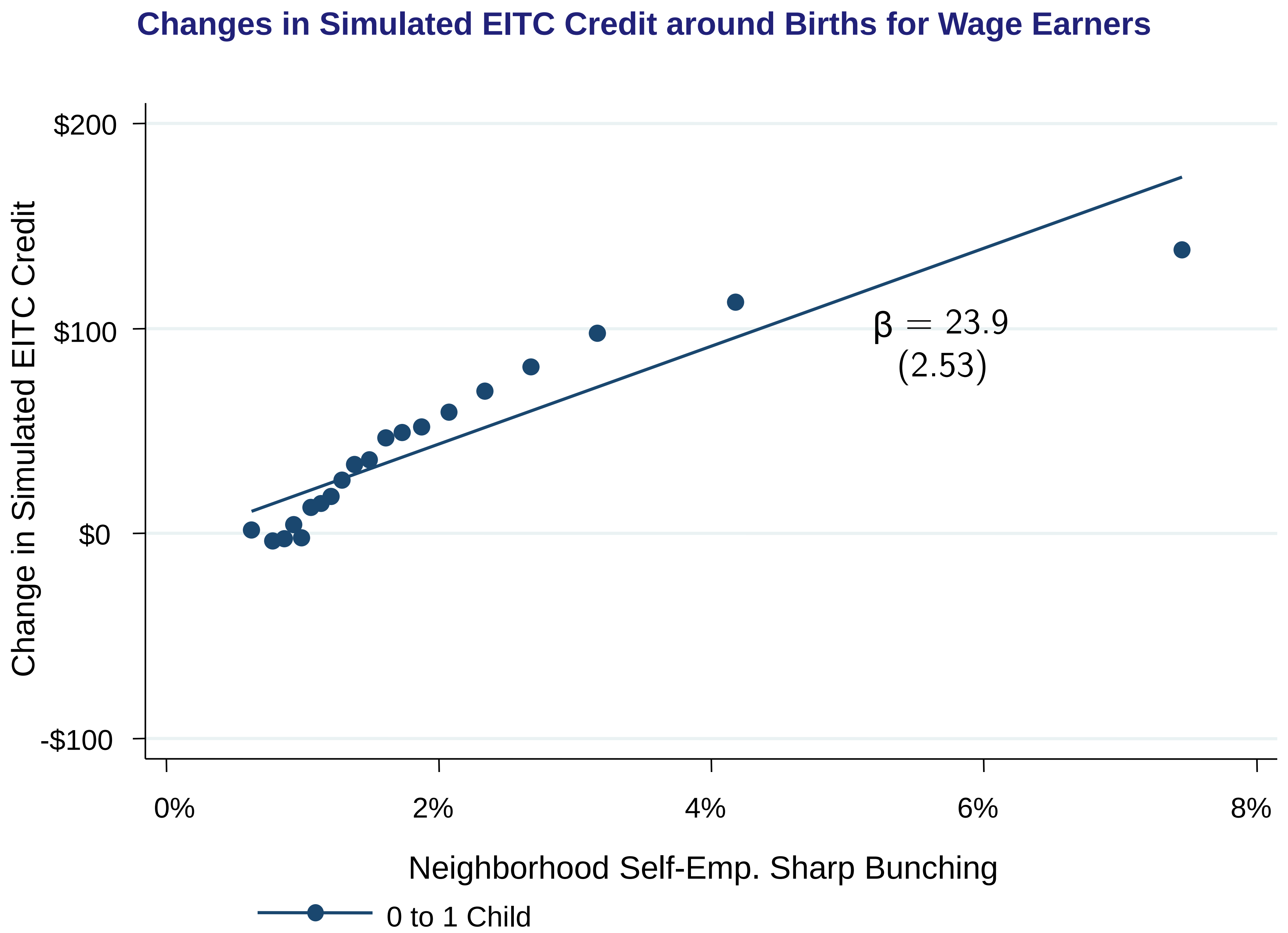

Simulated EITC Credit Amount for Wage Earners Around First Child Birth
Individuals Working at Firms with More than 100 Employees
$1500
$1400
$1300
$1200
Simulated EITC Credit
-4
-2
0
2
4
Age of Child
Lowest Bunching Decile
Middle Bunching Decile
Highest Bunching Decile



First Stage: Number of Children Claimed for Those With Zero Children Before Birth
Number of Children
1.5
1
.5
0
-4
-2
0
2
4
Age of Child
Lowest Bunching Decile
Medium Bunching Decile
Highest Bunching Decile

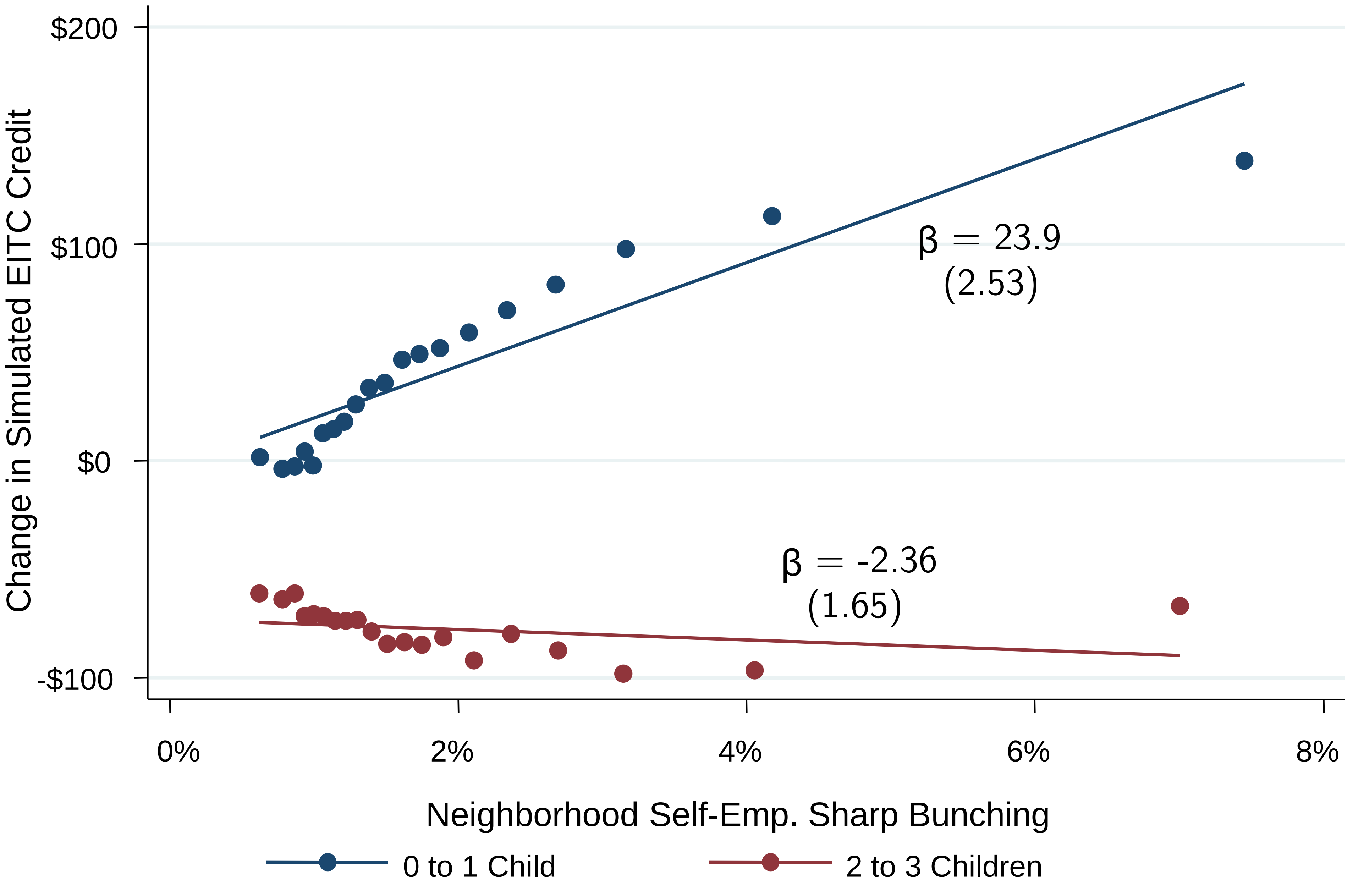
Composition of Wage Earnings Responses

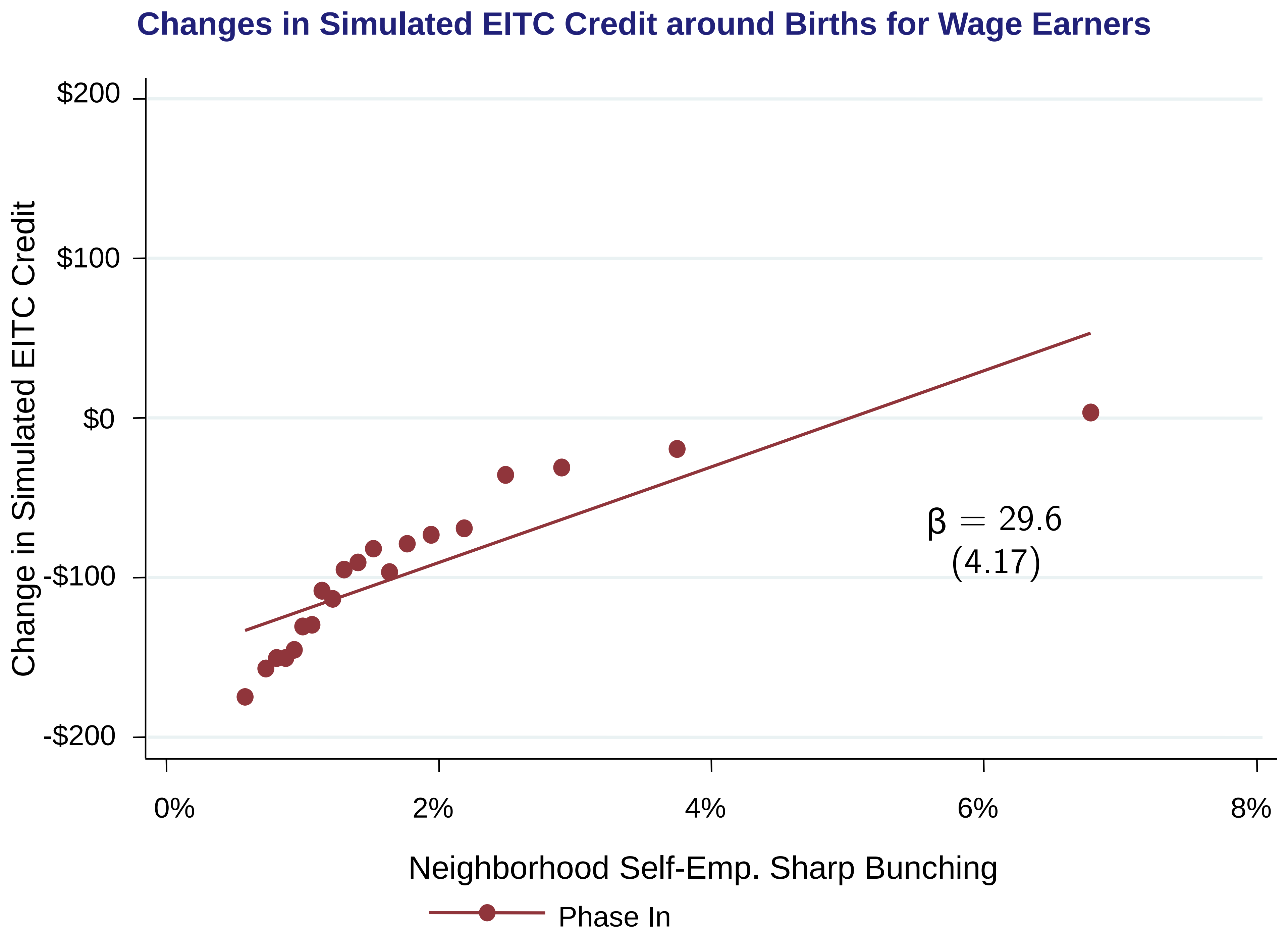
- Where is the increase in EITC refunds coming from?
  - Phase-in, phase-out, or extensive margin?
  - Important for understanding welfare consequences of EITC
- Calculate change in EITC amounts from year -1 to 0
  - Compare across low and high information areas to recover causal impact of EITC



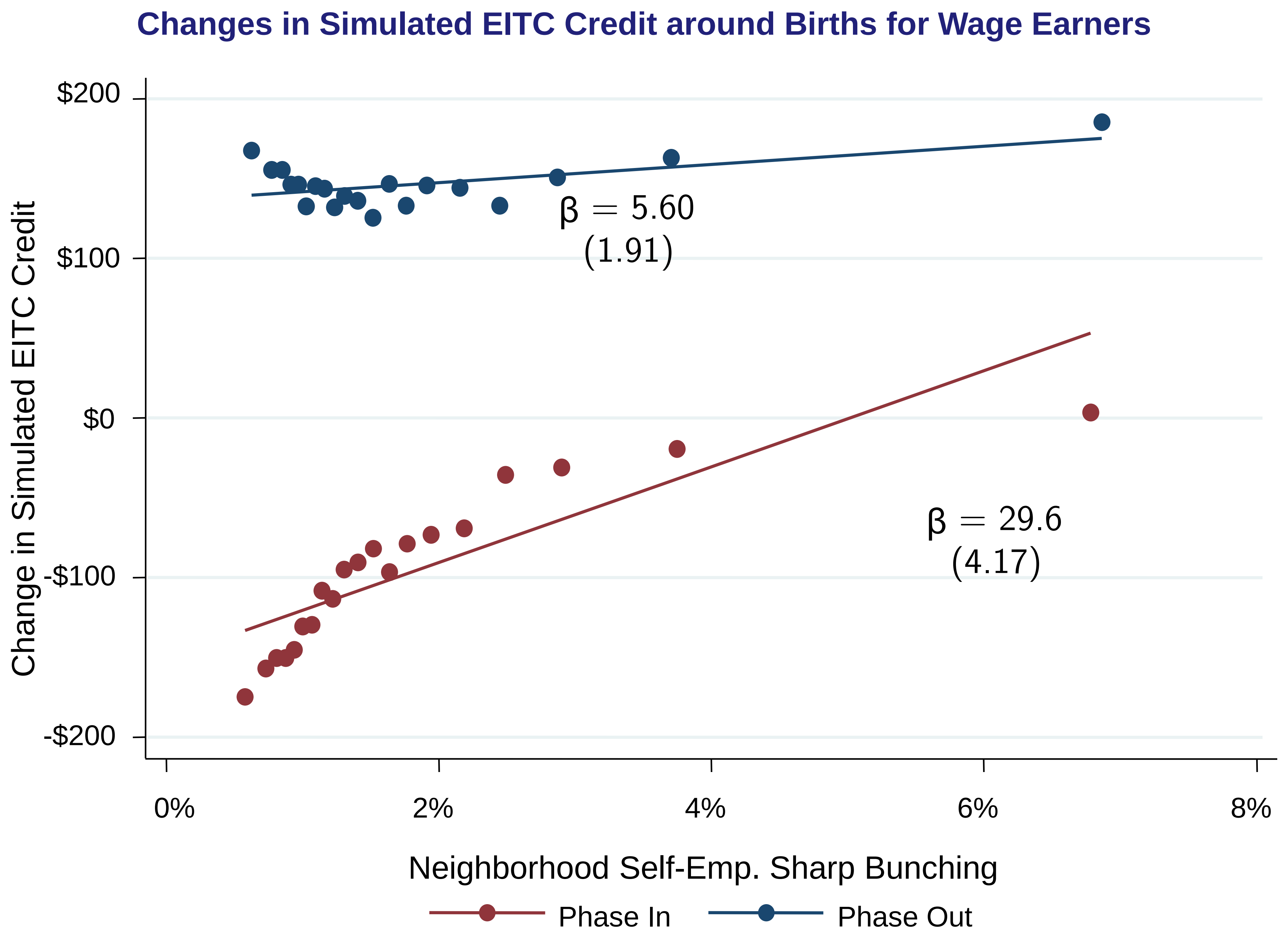
Changes in Simulated EITC Credit around Births for Wage Earners
$200
$100
$0
-$100
Change in Simulated EITC Credit
β = 23.9
(2.53)
0%
2%
4%
6%
8%
Neighborhood Self-Emp. Sharp Bunching
0 to 1 Child

Changes in Simulated EITC Credit around Births for Wage Earners

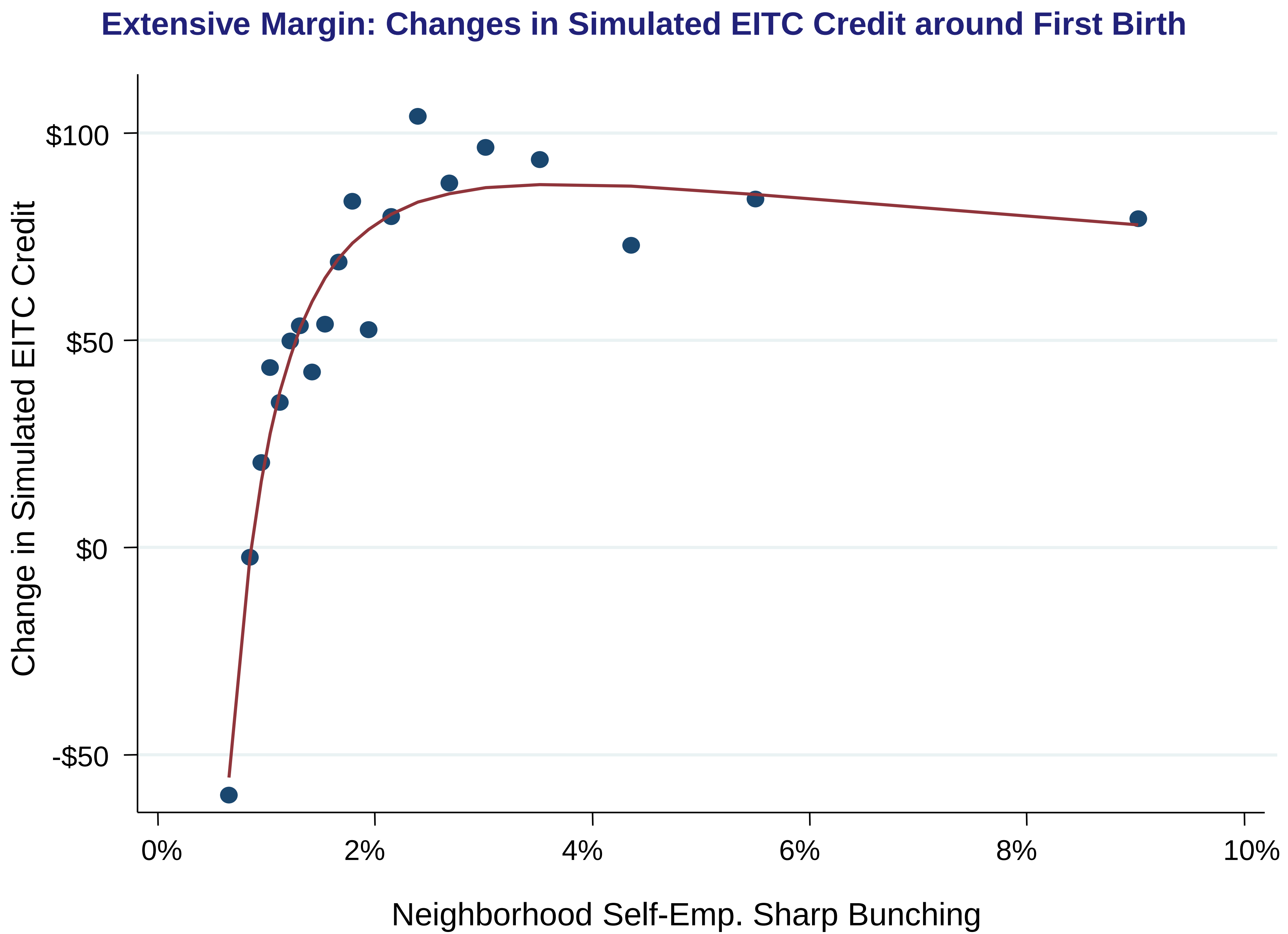




Changes in Simulated EITC Credit around Births for Wage Earners
$200
$100
$0
-$100
-$200
Change in Simulated EITC Credit
β = 29.6
(4.17)
0%
2%
4%
6%
8%
Neighborhood Self-Emp. Sharp Bunching
Phase In



Changes in Simulated EITC Credit around Births for Wage Earners
$200
$100
$0
-$100
-$200
Change in Simulated EITC Credit
β = 5.60
(1.91)
β = 29.6
(4.17)
0%
2%
4%
6%
8%
Neighborhood Self-Emp. Sharp Bunching
Phase In
Phase Out



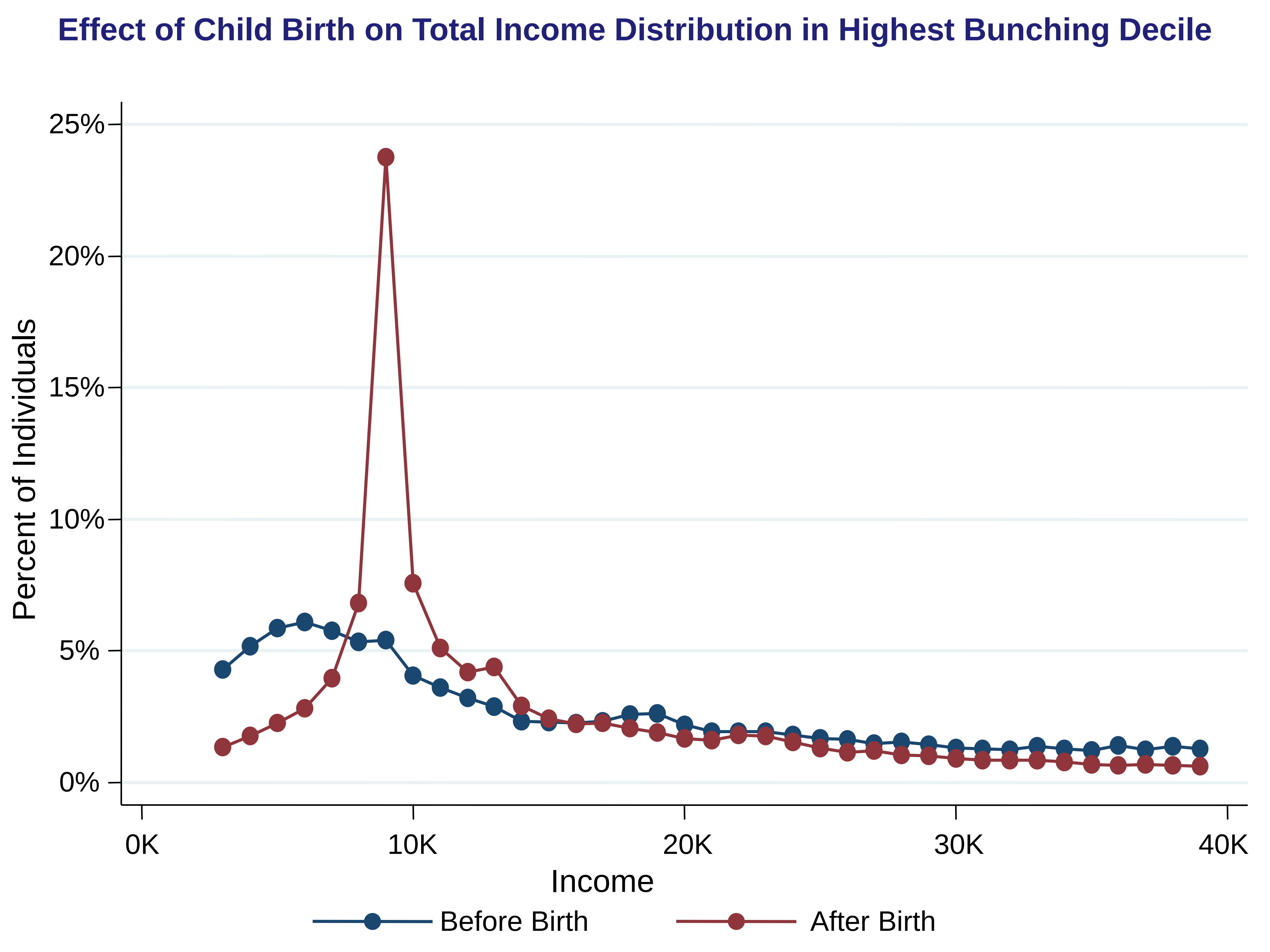
Extensive Margin: Changes in Simulated EITC Credit around First Birth
$100
$50
$0
-$50
Change in Simulated EITC Credit
0%
2%
4%
6%
8%
10%
Neighborhood Self-Emp. Sharp Bunching

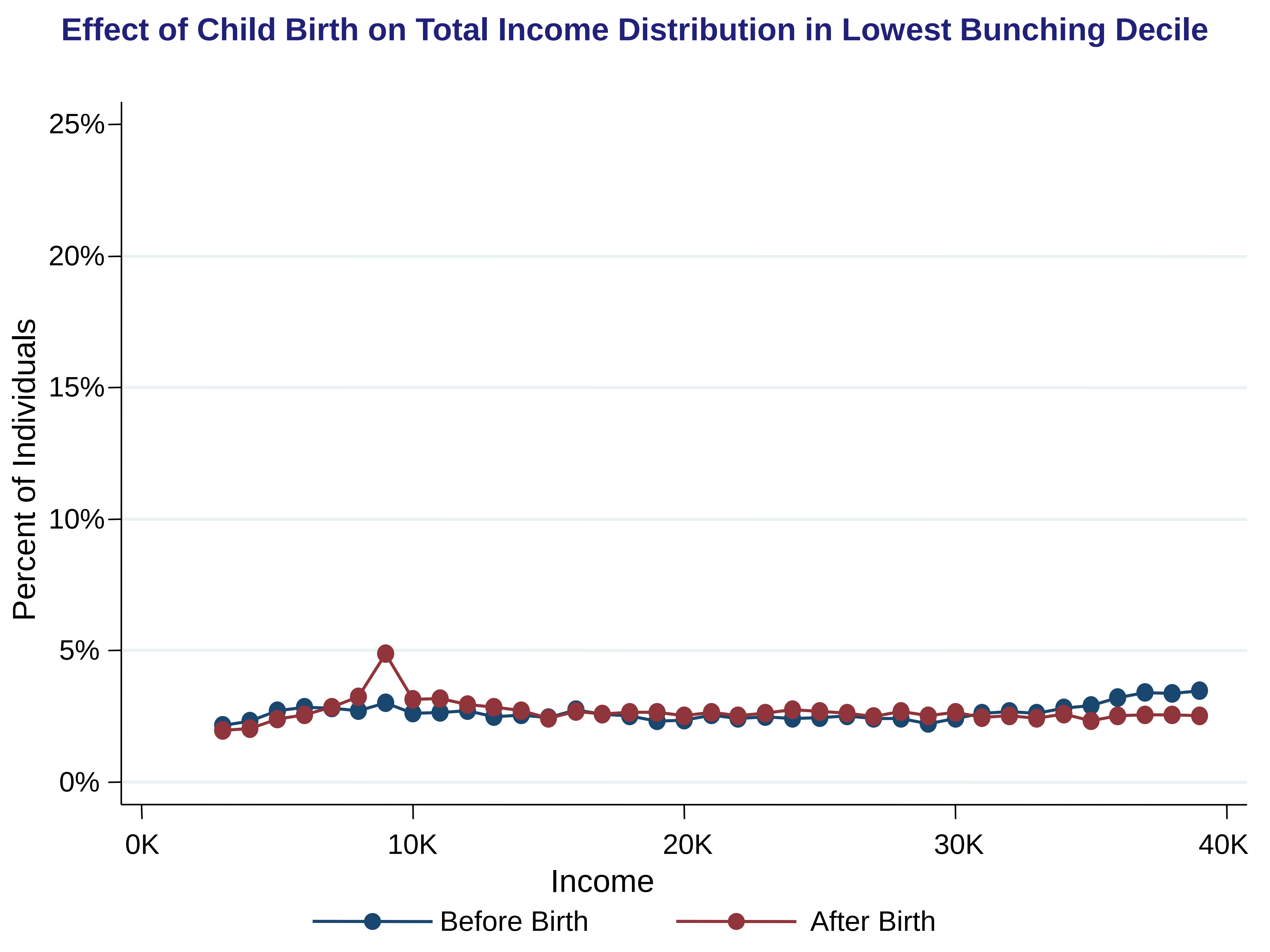
Composition of Wage Earnings Responses

- Assume that extensive margin entrants obtain average EITC refund of \$1,300
- Where is the increase in EITC refunds coming from?
  - Phase-In: 50%
  - Phase-Out: 14%
  - Zero earnings (extensive margin): 17%
  - Plateau: 19%

Tax Policy Implications

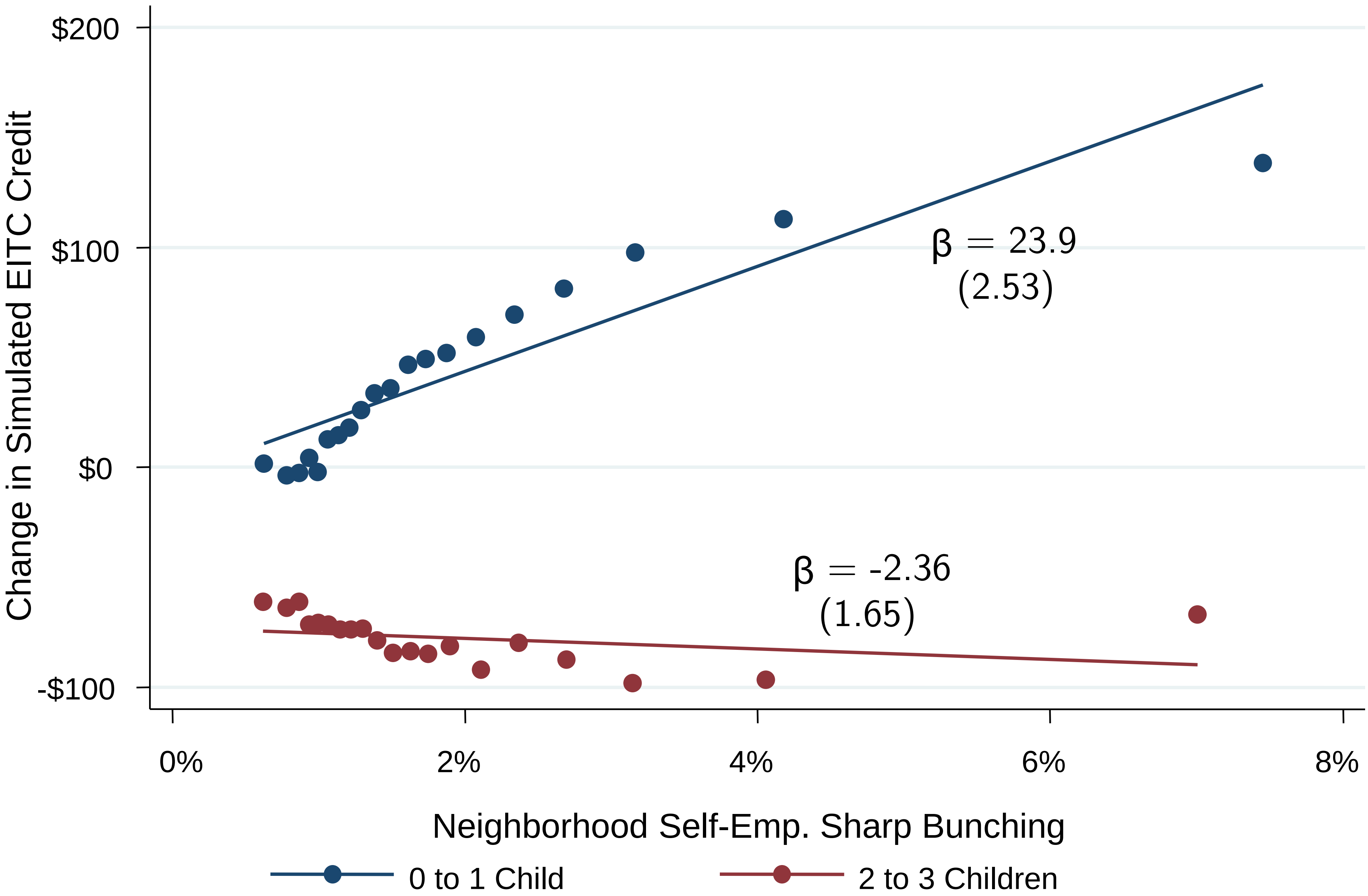
- Our estimates can be used to characterize impact of EITC on income distribution taking into account behavioral responses
- Use neighborhoods in bottom decile of self-employment bunching as counterfactual for earnings distribution without EITC
- Recall key assumption: neighborhoods with no self-employment bunching are places where people perceive marginal tax rates as zero
- Now present two pieces of evidence supporting this assumption


Effect of Child Birth on Total Income Distribution in Highest Bunching Decile
25%
20%
15%
10%
5%
0%
Percent of Individuals
0K
10K
20K
30K
40K
Income
Before Birth
After Birth



Effect of Child Birth on Total Income Distribution in Lowest Bunching Decile
Percent of Individuals
25%
20%
15%
10%
5%
0%
0K
10K
20K
30K
40K
Income
Before Birth
After Birth

Changes in Simulated EITC Credit around Births for Wage Earners



Impact of EITC on Income Distribution

	Percent of EITC Recipients with 2+ Kids Below:			
	1/2 Poverty Line	1 x Poverty Line	1.5 x Poverty Line	2 x Poverty Line
No EITC Counterfactual	17.75	49.93	75.82	93.77
EITC, No Behavioral Response	11.33	35.40	69.81	92.60
EITC, with Behavioral Response	10.02	34.81	69.91	92.72

Tax Policy Implications

- Average EITC refund amount for wage-earners is **7%** ($140) larger due to behavioral responses, primarily from increases in earnings
  - 40% of aggregate response from the top 10% of neighborhoods
  - In neoclassical model, generating an increase of 7% in refund amount requires an intensive margin taxable income elasticity of 0.2
- Information and learning via networks are central determinants of impacts of tax policy
  - Differences in knowledge can be used to identify causal impacts of other policies where traditional counterfactuals are unavailable
  - Ex: impacts of social security on retirement behavior