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

September 2011
Identifying Policy Impacts

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

1. Difficult to find counterfactuals to estimate causal impacts of federal policy changes [Meyer 1995, Gruber 2008]

We address these challenges by exploiting differences across neighborhoods in knowledge about tax policies.

Key idea: use cities with low levels of information about tax policies as counterfactuals 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.

Proxy for local knowledge about EITC using sharp bunching at kinks via manipulation of reported self-employment income.
Income Distribution for EITC-Eligible Self Employed with Children in 2008

Percent of EITC-Eligible Self-Employed Income Relative to First Kink of EITC Schedule

- Percentages range from 0% to 10%.
- Income levels range from -$10K to $30K.

Graph shows the distribution of income among EITC-eligible self-employed individuals with children in 2008.
Large literature has studied the impacts of EITC on labor supply
Hoynes 2004, Gelber and Mitchell 2011]

Clear evidence of impacts on participation (extensive margin)

But evidence on impacts of EITC on the earnings distribution
(intensive margin) remains mixed

Lack of information has greater impact on intensive margin
because gains from optimization are second-order [Chetty 2009]
Income Distribution for Single Wage Earners with One Child

Percent of EITC-Eligible Wage-Earners

Is the EITC having an effect on this distribution?

Taxable Income

EITC Credit Amount

Percent of EITC-Eligible Wage-Earners

$0K $10K $20K $30K

$0 $1K $2K $3K $4K

$0 $1K $2K $3K $4K

$0 $1K $2K $3K $4K

$0 $1K $2K $3K $4K

$0 $1K $2K $3K $4K
1. Conceptual Framework

2. Data and Institutional Background

3. Neighborhood Effects in Sharp Bunching via Income Manipulation

4. Using Neighborhood Effects to Uncover Wage Earnings Responses

5. Implications for Tax Policy
Workers face a two-bracket income tax system $\tau = (\tau_1, \tau_2)$

- 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 reported income is $K$

Stylized Model: Tax System

Earnings ($1000$) vs. Tax Refund Amount ($1000$)
Workers make two choices: earnings \( z \) and reported income \( \theta \).

- Fraction \( \theta \) of workers face 0 cost of non-compliance \( \Rightarrow \) report \( = K \)
- Remaining workers face infinite cost of non-compliance \( \Rightarrow \) set \( = z_i \)

Workers choose earnings \( z = wI \) to maximize utility \( u(c, l) \).

- Cannot control labor supply perfectly
- Utility maximization therefore produces diffuse “broad bunching” around kink point \( K \) rather than a point mass
- Diffuse response makes it difficult to estimate elasticities using neoclassical non-linear budget set methods (e.g. Hausman 1981)
Cities indexed by $c = 1,\ldots,N$

Cities differ only in one attribute: knowledge of tax code

In city $c$, fraction of workers know about tax subsidy for work

Remaining workers optimize as if tax rates are 0

Firms pay workers fixed wage rate in all cities
Goal: identify how taxes affect earnings distribution $F(z|\tau)$ with average level of knowledge in economy:

Empirical challenge: potential outcome without taxes $F(z|\tau=0)$ unobserved

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

- Need a proxy for degree of knowledge \( \lambda_c \)

- We use degree of sharp bunching at refund-maximizing kink

- Under assumption that \( \theta \) does not vary across cities, fraction who report \( = K \) is proportional to local knowledge:

\[ \rightarrow \text{City with no sharp bunching at kink yields no-tax counterfactual} \]
Stylized model motivates estimating equation of the form

where $\mu_{ic}$ is a measure of “broad bunching” in earnings around $K$ such as size of tax refund

Identification of $\beta$ relies on two assumptions

1. [Measurement error] Differences across cities in $f_c$ due to knowledge $\lambda_c$ and not other determinants of tax compliance $\theta$

2. [Omitted variables] Cities with different levels of knowledge do not have other attributes that affect earnings: $f_c \perp \eta_{ic}$

We use quasi-experimental research designs to address these concerns
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

- Sample restriction: individuals who at least once between 1996-2009:
  1. file a tax return,
  2. have income < $40,000,
  3. claim a dependent

- Sample size after restrictions:
  1. 77.6 million individuals
  2. 1.09 billion person-year observations on income
## Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>$21,175</td>
</tr>
<tr>
<td>Self Employed</td>
<td>9.1%</td>
</tr>
<tr>
<td>Married</td>
<td>24%</td>
</tr>
<tr>
<td>Number of Children</td>
<td>0.78</td>
</tr>
<tr>
<td>Female (among singles)</td>
<td>58%</td>
</tr>
</tbody>
</table>
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
2008 Federal EITC Schedule for a Single Filer with Children

Taxable Income (Real 2010 $)

- One child
- Two children
Income Distribution for EITC-Eligible Households with Children in 2008

Taxable Income (Real 2010 $)

Percent of EITC-Eligible Households

- 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

Reported Income Detected Income

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.
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 for the Self-Employed

Percent of EITC-Eligible Self-Employed

Income Relative to 1st Kink
Income Distribution in Kansas for the Self-Employed

Percent of EITC-Eligible Self-Employed

Income Relative to 1st Kink

- $10K
- $0
- $10K
- $20K
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
EITC Self-Employed Sharp Bunching by State in 2008

The map shows the distribution of EITC self-employed sharp bunching by state in 2008. The scale ranges from 0 to 0.3411. States are color-coded based on the percentage of self-employed individuals who fall into the 'sharp bunching' category, which is defined as earning slightly above the EITC eligibility threshold.

Key:
- 0.0268 – 0.3411
- 0.0187 – 0.0268
- 0.0151 – 0.0187
- 0.0126 – 0.0151
- 0.0110 – 0.0126
- 0.0099 – 0.0110
- 0.0096 – 0.0099
- 0.0084 – 0.0096
- 0 – 0.0084
EITC Elasticities for the Self-Employed in 2008 by 3-Digit Zip Code in Kansas, Louisiana, Oklahoma, and Texas

- Austin
- San Antonio
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: Analyze movers to establish learning as mechanism for differences in sharp bunching across neighborhoods
Variation in elasticities could simply reflect heterogeneity in individual preferences across places.

We evaluate whether variation in sharp bunching across cities is driven by differences in knowledge using four tests:

- **Movers**: do individuals begin to respond when they move to a high response city?
- **Learning**: do individuals continue to respond after leaving a high response city?
- **Spatial diffusion**: does response spread spatially and continue to increase over time?
- **Agglomeration**: response higher in cities with many EITC claimants.
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

-5 0 5

Event Year

0.0% 0.4% 0.8% 1.2%

Self-Emp. Sharp Bunching for Movers

Movers to Lowest Information Areas
Movers to Medium Information Areas
Movers to Highest Information Areas

$\Delta \varepsilon = 0.41\% (0.05\%)$
Movers’ Income Distributions: Before Move

- Percent of Movers
- Income Relative to 1st Kink

Movers to Lowest Information Areas
Movers to Medium Information Areas
Movers to Highest Information Areas
Movers’ Income Distributions: After Move

Percent of Movers

Income Relative to 1st Kink

Movers to Lowest Information Areas
Movers to Medium Information Areas
Movers to Highest Information Areas
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-Information Neighborhoods

Memory: old neighborhood matters when moving to lowest-information areas
Post-Move Distributions for Movers to **Highest-Information** Neighborhoods

→ **Learning:** Old neighborhood does not matter when moving to **highest-information** areas.
Asymmetric Impact of Neighborhoods on Bunching

<table>
<thead>
<tr>
<th></th>
<th>Move Up (1)</th>
<th>Move Down (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{old}$</td>
<td>0.252</td>
<td>$0.496$</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.046)</td>
</tr>
<tr>
<td>$\beta_{new}$</td>
<td>$0.822$</td>
<td>0.354</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.046)</td>
</tr>
</tbody>
</table>

p Value for Relative Change in Coefficients Across Columns: $p < 0.001$
Spatial Diffusion

- Macro-level implication of learning is that degree of sharp bunching should increase over time and diffuse spatially.

- Evaluate by examining evolution of bunching by year across states.
Self-Employed Sharp Bunching in 2005

Map of the United States showing the distribution of self-employed sharp bunching in 2005. The map uses a color scale ranging from light yellow (0 – 0.0084) to dark red (0.0268 – 0.3411). States with higher values are shaded in deeper red, while those with lower values are shaded in lighter colors. The map highlights the variation across different regions of the country.
Self-Employed Sharp Bunching in 2008

0.0268 – 0.3411
0.0187 – 0.0268
0.0151 – 0.0187
0.0126 – 0.0151
0.0110 – 0.0126
0.0110 – 0.0126
0.0099 – 0.0110
0.0099 – 0.0110
0.0084 – 0.0099
0.0084 – 0.0099
0 – 0.0084

$\beta = 0.00164 \pm 0.0000408$

Graph showing the relationship between Log Population per Square Mile and 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: Analyze movers to establish learning as 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 Distributions for Single Wage Earners with One Child

Percent of EITC-Eligible Wage-Earners

Is the EITC having an effect on this distribution?
Wage Earnings Distributions in High vs. Low Information Areas
Single Individuals with One Child

Percent of EITC-Eligible Wage-Earners

- High Information Neighborhoods
- Low Information Neighborhoods

Income:
- $0
- $10K
- $20K
- $30K
Wage Earnings Distributions in High vs. Low Information Areas
Single Individuals with Two Children

Income Relative to First Kink in EITC Schedule

Difference in Income Densities

All Firms

EITC Amount

$0

$1000

$2000

$3000

$4000

$5000

$-10K

$0

$10K

$20K

$30K

$10K

$20K

$30K
Wage Earnings Distributions in High vs. Low Information Areas
Single Individuals with Two Children

Income Relative to First Kink in EITC Schedule

Difference in Income Densities

EITC Amount

-0.1
-0.05
0

$0
$10K
$20K
$30K

-$10K
$0
$10K
$20K
$30K

$5000
$4000
$3000
$2000
$1000

All Firms
>100 Employees
EITC Credit Amount for Single Wage Earners with Two Children vs. Neighborhood Bunching

<table>
<thead>
<tr>
<th>EITC Credit Amount</th>
<th>Neighborhood Self-Emp. Sharp Bunching</th>
</tr>
</thead>
<tbody>
<tr>
<td>$3200</td>
<td>0.0%</td>
</tr>
<tr>
<td>$3250</td>
<td>0.8%</td>
</tr>
<tr>
<td>$3300</td>
<td>1.6%</td>
</tr>
<tr>
<td>$3350</td>
<td>2.4%</td>
</tr>
<tr>
<td>$3400</td>
<td>3.2%</td>
</tr>
</tbody>
</table>

Graph showing the relationship between EITC Credit Amount and Neighborhood Self-Emp. 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: Analyze movers to establish learning as 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 changes in EITC subsidies on earnings across low vs. high knowledge nbhds. to account for omitted variables
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?
To identify causal impacts of EITC, need variation in tax incentives

- Birth of first child $\rightarrow$ 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 Distributions in the Year Before First Child Birth for Wage Earners

Percent of Households

Income

Lowest Information Neighborhoods

Medium Information Neighborhoods

Highest Information Neighborhoods

Percent of Households

$0

$10K

$20K

$30K

$40K

$0

$10K

$20K

$30K

$40K
Earnings Distributions in the Year of First Child Birth for Wage Earners

Percent of Households

Lowest Information Neighborhoods

Medium Information Neighborhoods

Highest Information Neighborhoods

Income

Percent of Households

Lowest Information Neighborhoods

Medium Information Neighborhoods

Highest Information Neighborhoods

$0 $10K $20K $30K $40K
Earnings Distributions in the Year of First Child Birth for Wage Earners
Individuals Working at Firms with More than 100 Employees

Percent of Households

Lowest Information Neighborhoods

Medium Information Neighborhoods

Highest Information Neighborhoods

Income

$0 $10K $20K $30K $40K

Percent of Households

Lowest Information Neighborhoods

Medium Information Neighborhoods

Highest Information Neighborhoods

$0 $10K $20K $30K $40K

Lowest Information Neighborhoods

Medium Information Neighborhoods

Highest Information Neighborhoods

$0 $10K $20K $30K $40K
Simulated EITC Credit Amount for Wage Earners Around First Child Birth Individuals Working at Firms with More than 100 Employees

Simulated EITC Credit

Age of Child

Lowest Information Neighborhoods

Medium Information Neighborhoods

Highest Information Neighborhoods

-4
-2
0
2
4

Age of Child

Simulated EITC Credit

$1400
$1300
$1200
$1100
$1000
$900

$1400
$1300
$1200
$1100
$1000
$900

Lowest Information Neighborhoods

Medium Information Neighborhoods

Highest Information Neighborhoods
Increase in Simulated EITC Credit around Births for Wage Earners

Percent Increase in Simulated EITC Credit

Neighborhood Self-Emp. Sharp Bunching

-5% 0% 5% 10% 15% 20%

β = 7.25

(0.644)

0% 1% 2% 3% 4%
Increase in Simulated EITC Credit around Births for Wage Earners

- 0 to 1 Children
- 2 to 3 Children

- \( \beta = 0.214 \) (0.334)
- \( \beta = 7.25 \) (0.644)
Composition of Wage Earnings Responses

- Where is the excess mass in the plateau coming from?
  - Phase-In
  - Phase-Out
  - Extensive Margin

- Important for understanding welfare implication of EITC
Change in Fraction on Plateau around First Births

Log Change in Fraction on Plateau

\[ \beta = 0.109 \pm 0.007 \]

Neighborhood Self-Emp. Sharp Bunching

0% 1% 2% 3% 4%
Change in Fraction on Plateau around First Births

Log Change in Fraction on Plateau

\[ \beta = 0.026 \quad (0.012) \]

\[ \beta = 0.109 \quad (0.007) \]

Neighborhood Self-Emp. Sharp Bunching

0% 1% 2% 3% 4%

Starts in “Phase-in”  
Starts in “Phase-out”
Extensive Margin: Changes in Probability of Working around First Birth

\[
\beta = 1.46 \\
(0.045)
\]
Response to the EITC varies across cities for wage earners

Our hypothesis is that this is because of differences in knowledge

To verify the causal effect of neighborhoods, we again use movers

Do EITC-eligible individuals who move to high response cities have higher concentration of earnings near plateau?
Income Distributions Before Move for Wage Earners

Percent of EITC-Eligible Households

Income Relative to 1st Kink

Movers to Lowest Information Areas
Movers to Medium Information Areas
Movers to Highest Information Areas

Income Distributions Before Move for Wage Earners

Percent of EITC-Eligible Households

Income Relative to 1st Kink

Movers to Lowest Information Areas
Movers to Medium Information Areas
Movers to Highest Information Areas
Event Study of EITC Amount for Wage-Earners by Destination Area

<table>
<thead>
<tr>
<th>Event Year</th>
<th>EITC Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Movers to Lowest Information Areas</td>
</tr>
<tr>
<td>-5</td>
<td>$1500</td>
</tr>
<tr>
<td>0</td>
<td>$1650</td>
</tr>
<tr>
<td>5</td>
<td>$1500</td>
</tr>
</tbody>
</table>
Tax Policy Implications

- Our estimates can be used to characterize impact of EITC on income distribution taking into account behavioral responses

- Use neighborhoods with little self-employment bunching as counterfactual for earnings distribution without EITC
Impact of EITC on Income Distribution for Single Earners with 2+ Children

Percent of EITC-Eligible Wage-Earners

Total Income

$0 $10K $20K $30K $40K

No EITC

Counterfactual
Impact of EITC on Income Distribution for Single Earners with 2+ Children

Percent of EITC-Eligible Wage-Earners

Total Income

- $0
- $10K
- $20K
- $30K
- $40K

- No EITC
- EITC, No Behavioral Response

Counterfactual
Impact of EITC on Income Distribution for Single Earners with 2+ Children

Percent of EITC-Eligible Wage-Earners

Total Income

- No EITC
- EITC, No Behavioral Response
- EITC with Behavioral Response

Income Levels:
- $0
- $10K
- $20K
- $30K
- $40K
Our estimates imply that average EITC refund amount for wage-earners is 7% ($140) larger due to behavioral responses. 40% of aggregate response from the top 10% of neighborhoods. Response primarily due to an intensive-margin increase in earnings coming from the phase-in region. In neoclassical model, generating an increase of 7% in refund amount would require an intensive margin elasticity of 0.2.
Neighborhood effects could be used to uncover impacts of many policies

Example: Saver’s Credit

Saver’s Credit provides up to a 100% subsidy to save in an IRA for low-income households

Eligibility based on discontinuous income thresholds

Previous work has documented modest impacts of saver’s credit on IRA contributions in aggregate [Duflo et al. 2006, 2007; Ramnath 2011]
IRA Take-Up Rates by Income Bin

<table>
<thead>
<tr>
<th>Income</th>
<th>% Take-up of IRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>-$5K</td>
<td>1.6%</td>
</tr>
<tr>
<td>-$4K</td>
<td>1.8%</td>
</tr>
<tr>
<td>-$3K</td>
<td>2.0%</td>
</tr>
<tr>
<td>-$2K</td>
<td>2.2%</td>
</tr>
<tr>
<td>-$1K</td>
<td>2.4%</td>
</tr>
<tr>
<td>$0</td>
<td>2.6%</td>
</tr>
<tr>
<td>$1K</td>
<td>2.8%</td>
</tr>
<tr>
<td>$2K</td>
<td>3.0%</td>
</tr>
<tr>
<td>$3K</td>
<td></td>
</tr>
<tr>
<td>$4K</td>
<td></td>
</tr>
<tr>
<td>$5K</td>
<td></td>
</tr>
</tbody>
</table>
Savers Credit Response, 2002-2008
Saver’s Credit Response by 3-Digit Zip, 2002-2008 in Illinois, Indiana, Michigan, and Wisconsin

- Chicago
- Detroit
- Indianapolis
Future work could use neighborhood effects in response to saver’s tax credit to analyze impacts of IRAs’ on behavior.

Compare effect of IRA eligibility change in areas with high vs. low saver’s credit response.

Neighborhood effects could also be used to analyze other tax policies, e.g. impacts of social security on retirement.

Classify areas based on response to a policy such as earnings test, as in Friedberg (1999).

Use low-response areas as a counterfactual to study the impact of changes in social security policies on retirement.