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
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

- $5K: 10%
- $4K: 8%
- $3K: 4%
- $2K: 2%
- $1K: 0%
- $0: 0%
- -$10K: 0%

Income Relative to First Kink of EITC Schedule

Percent of EITC-Eligible Self-Employed

- 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

**Is the EITC having an effect on this distribution?**

- **EITC Credit Amount**
  - $0
  - $1K
  - $2K
  - $3K
  - $4K

- **TAXABLE INCOME**
  - $0
  - $10K
  - $20K
  - $30K

**Percent of EITC-Eligible Wage-Earners**

- 0%
- 1%
- 2%
- 3%
- 4%
Outline

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
Conceptual Framework

- Heterogeneity in EITC knowledge across cities
- In each city, exogenous fraction of individuals willing to misreport income using self-employment earnings (SE) to maximize EITC refund
  - Fraction EITC claimants bunching at 1st kink with SE proxies the level of EITC knowledge in population
- Honest EITC claimants with EITC knowledge respond with real labor supply responses
  - Cannot bunch exactly at kink point but will change their earnings to get bigger EITC
  - Compare low vs. high knowledge places to uncover labor supply responses
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 $K$
Workers make two choices: earnings \( z \) and reported income \( \hat{z}_i \).

- Fraction \( \theta \) of workers face 0 cost of non-compliance \( \rightarrow \text{report} \hat{z}_i = K \).
- Remaining workers face infinite cost of non-compliance \( \rightarrow \text{set} \ z_i = \hat{z}_i \).

Workers choose earnings \( z = wl \) 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 $\lambda_c$ 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
Identifying Tax Policy Impacts

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

$$\Delta F(z|\tau) = [F(z|\tau > 0, \lambda_c) - F(z|\tau = 0, \lambda_c)]$$

- 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

$$F(z|\tau > 0, \lambda_c = 0) = F(z|\tau = 0, \lambda_c)$$
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 $\hat{z}_i = K$ is proportional to local knowledge:

  $$f_c = \theta \lambda_c$$

$\rightarrow$ City with no sharp bunching at kink yields no-tax counterfactual

$$F(z|\tau > 0, f_c = 0) = F(z|\tau = 0, \lambda_c)$$
Empirical Implementation

- Stylized model motivates estimating equation of the form

\[ \mu_{ic} = \alpha + \beta f_c + \eta_{ic} \]

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:
  - 77.6 million individuals
  - 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>
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
2008 Federal EITC Schedule for a Single Filer with Children

Taxable Income (Real 2010 $) vs. EITC Credit

- **One child**: Blue line
- **Two children**: Red line

Income Thresholds:
- $0 - $10K
- $10K - $20K
- $20K - $30K
- $30K - $40K

EITC Credit Amounts:
- $0 - $5K
- $5K - $10K
- $10K - $15K
- $15K - $20K
- $20K - $25K
- $25K - $30K
- $30K - $35K
- $35K - $40K

Explanation:
- The graph shows the EITC credit as a function of taxable income for single filers with children, distinguishing between one child and two children.
- The credit amount increases with income up to a certain point, then decreases as income rises further.
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.
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

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

15%
10%
5%
0%
Income Distribution in Kansas for the Self-Employed

Percent of EITC-Eligible Self-Employed

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
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
Are Neighborhood Effects Driven by Knowledge?

- 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

\[ \Delta \varepsilon = 0.41\% (0.05\%) \]

<table>
<thead>
<tr>
<th>Event Year</th>
<th>Self-Emp. Sharp Bunching for Movers</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5</td>
<td>Movers to Lowest Information Areas</td>
</tr>
<tr>
<td>-4</td>
<td>Movers to Medium Information Areas</td>
</tr>
<tr>
<td>-3</td>
<td>Movers to Highest Information Areas</td>
</tr>
<tr>
<td>-2</td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
Movers’ Income Distributions: Before Move

Income Relative to 1st Kink

Percent of Movers

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

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

Income Relative to 1st Kink

Percent of Movers

-10K 0 10K 20K 30K

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

Income Relative to 1st Kink

Movers from Lowest Information Areas
Movers from Medium Information Areas
Movers from Highest 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

\[ b_{\text{mover}} = \alpha + \beta_{\text{old}} b_{\text{old \ neighborhood}} + \beta_{\text{new}} b_{\text{new \ neighborhood}} \]

<table>
<thead>
<tr>
<th>Dependent variable: ( b ) for movers</th>
<th>Move Up</th>
<th>Move Down</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move Up</td>
<td>Move Down</td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td></td>
</tr>
<tr>
<td>( \beta_{\text{old}} )</td>
<td>0.252</td>
<td>0.496</td>
</tr>
<tr>
<td>(0.058)</td>
<td>(0.046)</td>
<td></td>
</tr>
<tr>
<td>( \beta_{\text{new}} )</td>
<td>0.822</td>
<td>0.354</td>
</tr>
<tr>
<td>(0.058)</td>
<td>(0.046)</td>
<td></td>
</tr>
</tbody>
</table>

p Value for Relative Change in Coefficients Across Columns: \( p < 0.001 \)
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 1999

The map shows the distribution of self-employed workers in 1999, with different states shaded to indicate the percentage of self-employed workers. The scale on the right indicates the percentage range for each color:

- Dark red: 0.0268 – 0.3411
- Maroon: 0.0187 – 0.0268
- Deep red: 0.0151 – 0.0187
- Red: 0.0126 – 0.0151
- Orange: 0.0110 – 0.0126
- Yellow-orange: 0.0099 – 0.0110
- Yellow: 0.0096 – 0.0099
- Light yellow: 0.0084 – 0.0096
- Pale yellow: 0 – 0.0084

States with the highest percentage of self-employed workers in 1999 are highlighted in dark red, while those with the lowest are in pale yellow. The map provides a visual representation of the geographical distribution of self-employment across the United States.
Self-Employed Sharp Bunching in 2002

The map illustrates the distribution of self-employed workers across the United States in 2002, with shades indicating the range of values from 0 to 0.3411. States with darker shades have higher percentages of self-employed workers, while lighter shades represent lower percentages.

Key:
- 0.3411 – 0.0268
- 0.0187 – 0.0151
- 0.0126 – 0.0110
- 0.0096 – 0.0084
- 0.0084 – 0.0096
- 0 – 0.0084
Self-Employed Sharp Bunching in 2005

![Map showing self-employed sharp bunching in 2005 across the United States. The map uses a color scale to represent different ranges of self-employment rates. The color key indicates rates ranging from 0.0268 to 0.3411. States with higher self-employment rates are shown in darker shades of red, while those with lower rates are shown in lighter shades of yellow and green.](image-url)
Self-Employed Sharp Bunching in 2008

\[ \beta = 0.00164 \pm 0.0000408 \]

Graph showing the relationship between self-employment and log population density per square mile.
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

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

Income

High Information Neighborhoods

Low Information Neighborhoods
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

All Firms
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

-0.01
-0.005
0
0.005
0.01

$0
$1000
$2000
$3000
$4000
$5000

EITC Amount

$0
$10K
$20K
$30K

-100 Employees
All Firms
>100 Employees

$4000
$5000

Amount

Come

Dence in Income

Earnings

Distributions

High
vs. Low
Information
Areas

Single
Individuals
with Two
Children

$0
$10K
$20K
$30K

>100 Employees
All Firms
>100 Employees

Difference in Income Densities

-0.01
-0.005
0
0.005
0.01

$0
$1000
$2000
$3000
$4000
$5000

EITC Amount

$0
$10K
$20K
$30K

-100 Employees
All Firms
>100 Employees

Wage Earnings Distributions in High vs. Low Information Areas
Single Individuals with Two Children
EITC Credit Amount for Single Wage Earners with Two Children vs. Neighborhood Bunching

![Graph showing the relationship between EITC credit amount and neighborhood self-employment sharp bunching. The x-axis represents neighborhood self-employment sharp bunching percentage (0.0% to 4.0%), and the y-axis represents EITC credit amount ($3200 to $3350). The data points are scattered, with a trend line indicating a positive correlation.]
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.
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?
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 of First Child Birth for Wage Earners

![Graph showing earnings distributions by income and information neighborhoods.](chart.png)
Earnings Distributions in the Year of First Child Birth for Wage Earners Individuals Working at Firms with More than 100 Employees

The graph shows the earnings distributions in the year of first child birth for wage earners individuals working at firms with more than 100 employees. The graph is divided into three categories: Lowest Information Neighborhoods, Medium Information Neighborhoods, and Highest Information Neighborhoods. The x-axis represents income levels ranging from $0 to $40K, while the y-axis represents the percent of households.
Simulated EITC Credit Amount for Wage Earners Around First Child Birth
Individuals Working at Firms with More than 100 Employees
Increase in Simulated EITC Credit around Births for Wage Earners

\[ \beta = 7.25 \quad (0.644) \]
Increase in Simulated EITC Credit around Births for Wage Earners

- β = 7.25 (0.644)
- β = 0.214 (0.334)

Graph showing the percent increase in simulated EITC credit for different neighborhood self-employment sharp bunching percentages, with separate lines for 0 to 1 children and 2 to 3 children.
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

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

Neighborhood Self-Emp. Sharp Bunching

Log Change in Fraction on Plateau

Started in “Phase-in”  Started in “Phase-out”
Change in Fraction on Plateau around First Births

\[ \beta = 0.026 \pm 0.012 \]

\[ \beta = 0.109 \pm 0.007 \]
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 After Move for Wage Earners

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

Income Relative to 1st Kink

Percent of EITC-Eligible Households

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

- 4%
- 3%
- 2%
- 1%
- 0%

-$10K

$30K
Event Study of EITC Amount for Wage-Earners by Destination Area

Event Year

-5 0 5

EITC Amount

Movers to Lowest Information Areas

Movers to Medium Information Areas

Movers to Highest Information Areas
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

No EITC
Counterfactual
Impact of EITC on Income Distribution for Single Earners with 2+ Children

The graph shows the distribution of total income for single earners with 2+ children, comparing scenarios with and without the Earned Income Tax Credit (EITC). The x-axis represents total income, ranging from $0 to $40K, while the y-axis shows the percentage of EITC-eligible wage-earners.

Two lines are depicted: one for the counterfactual scenario (No EITC) and another for the scenario with the EITC, but without behavioral responses. The graph illustrates how the EITC shifts the income distribution, particularly at lower income levels, with a noticeable increase in the percentage of EITC-eligible wage-earners at income levels below $20K compared to the counterfactual scenario.
Impact of EITC on Income Distribution for Single Earners with 2+ Children

- **No EITC Counterfactual**
- **EITC, No Behavioral Response**
- **EITC with Behavioral Response**

The chart shows the percentage of EITC-eligible wage-earners across different income levels.

- The income scale ranges from $0 to $40K.
- The y-axis represents the percentage of EITC-eligible wage-earners.
- The x-axis represents total income in thousands of dollars.
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

% Take-up of IRA

Income

-$5K  -$4K  -$3K  -$2K  -$1K  $0  $1K  $2K  $3K  $4K  $5K
Saver’s Credit Response by 3-Digit Zip, 2002-2008
in Illinois, Indiana, Michigan, and Wisconsin
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