Lecture 4: Labor Supply Responses to Taxation

Stefanie Stantcheva

Fall 2016
GOALS OF THIS LECTURE

1) Cover empirical studies of labor supply responses to taxation going historically from earlier to more recent papers. Contributes to our highly important “internal paper wikipedia” (IPW).

2) Understand key methodologies such as non-linear budget sets and “bunching at the kinks,” which are useful for a wide range of empirical work.

3) Critically discuss papers’ methodologies and results so as to practice our research skills.
MOTIVATION

1) Labor supply responses to taxation are of fundamental importance for income tax policy [efficiency costs and optimal tax formulas]

2) Labor supply responses along many dimensions:
   (a) Intensive: hours of work on the job, intensity of work, occupational choice [including education]
   (b) Extensive: whether to work or not [e.g., retirement and migration decisions]

3) Reported earnings for tax purposes can also vary due to (a) tax avoidance [legal tax minimization], (b) tax evasion [illegal under-reporting]

4) Different responses in short-run and long-run: long-run response most important for policy but hardest to estimate
STATIC MODEL: SETUP

Baseline model: (a) static, (b) linearized tax system, (c) pure intensive margin choice, (d) single hours choice, (e) no frictions

Let $c$ denote consumption and $l$ hours worked, utility $u(c, l)$ increases in $c$, and decreases in $l$

Individual earns wage $w$ per hour (net of taxes) and has $y$ in non-labor income

Key example: pre-tax wage rate $w^p$ and linear tax system with tax rate $\tau$ and demogrant $G \Rightarrow c = w^p(1 - \tau)l + G$

Individual solves

$$\max_{c,l} u(c, l) \quad \text{subject to} \quad c = wl + y$$
LABOR SUPPLY BEHAVIOR

FOC: \( wu_c + u_l = 0 \) defines uncompensated (Marshallian) labor supply function \( l^u(w, y) \)

Uncompensated elasticity of labor supply: \( \varepsilon^u = (w/l) \frac{\partial l^u}{\partial w} \) [\% change in hours when net wage \( w \) ↑ by 1%]

Income effect parameter: \( \eta = w \frac{\partial l}{\partial y} \leq 0 \): $ increase in earnings if person receives $1 extra in non-labor income

Compensated (Hicksian) labor supply function \( l^c(w, u) \) which minimizes cost \( wl - c \) st to constraint \( u(c, l) \geq u \).

Compensated elasticity of labor supply: \( \varepsilon^c = (w/l) \frac{\partial l^c}{\partial w} > 0 \)

Slutsky equation: \( \frac{\partial l}{\partial w} = \frac{\partial l^c}{\partial w} + l \frac{\partial l}{\partial y} \Rightarrow \varepsilon^u = \varepsilon^c + \eta \)
Labor Supply Theory

\[ c = \text{consumption} \]

slope = \( w \)

Marshallian Labor Supply

\[ l(w, R) \]

Indifference Curve

\[ c = w l + R \]

0 labor supply \( l \)
Labor Supply Theory

\[ c = \text{consumption} \]

Hicksian Labor Supply

\[ l^c(w, u) \]

utility \( u \)

slope = \( w \)
Labor Supply Income Effect

\[ \eta = w \frac{\partial l}{\partial R} \leq 0 \]
Labor Supply Substitution Effect

\[ \text{slope} = w \]

utility \( u \)

\[ \text{slope} = w + \Delta w \]

\[ l^c(w, u) \]

\[ l^c(w + \Delta w, u) \]

\[ \varepsilon^c = \frac{w \partial l^c}{l \partial w} > 0 \]
Uncompensated Labor Supply Effect

\[ \varepsilon^u = \varepsilon^c + \eta \]

Substitution effect: \( \varepsilon^c > 0 \)

Income effect:

\[ \eta \leq 0 \]

\[ \varepsilon^u = \varepsilon^c + \eta \]
BASIC CROSS SECTION ESTIMATION

Data on hours or work, wage rates, non-labor income started becoming available in the 1960s when first micro surveys and computers appeared:

Simple OLS regression:

\[ l_i = \alpha + \beta w_i + \gamma y_i + X_i \delta + \epsilon_i \]

\( w_i \) is the net-of-tax wage rate

\( y_i \) measures non-labor income [including spousal earnings for couples]

\( X_i \) are demographic controls [age, experience, education, etc.]

\( \beta \) measures uncompensated wage effects, and \( \gamma \) income effects [can be converted to \( \varepsilon^u, \eta \)]
BASIC CROSS SECTION RESULTS

   a) Small effects $\varepsilon^u = 0$, $\eta = -0.1$, $\varepsilon^c = 0.1$ with some variation across estimates (sometimes $\varepsilon^c < 0$).

2. Female workers [secondary earners when married] (Killingsworth and Heckman, 1986):
   Much larger elasticities on average, with larger variations across studies. Elasticities go from zero to over one. Average around 0.5. Significant income effects as well

   Female labor supply elasticities have declined overtime as women become more attached to labor market (Blau-Kahn JOLE’07)
PROBLEMS WITH OLS ESTIMATION OF LABOR SUPPLY EQUATION

1) Econometric issues
   a) Unobserved heterogeneity [tax instruments]
   b) Measurement error in wages and division bias [tax instruments]
   c) Selection into labor force [selection models]
   d) Endogenous tax rates [non-linear budget set methods]

2) Extensive vs. intensive margin responses [participation models]

3) Non-hours responses [taxable income]
Non-Linear Budget Set Estimation: Virtual Incomes

Source: Hausman (Hbk 1985)
NON-LINEAR BUDGET SETS

Actual tax system is not linear but piece-wise linear with varying marginal tax rate $\tau$ due to (a) means-tested transfer programs, (b) progressive individual income tax.

Same theory applies when considering the linearized tax system $c = wl + y$ with $w = w^p(1 - T')$ and $y$ defined as virtual income (intercept of budget with x-axis when setting $l = 0$).

Main complications:

(a) $w$ [and $y$] become endogenous to choice of $l$

(b) FOC may not hold if individual bunches at a kink

(c) FOC may not characterize the optimum choice
What is the problem?

What is the issue with NLBS and the simple OLS regression?

\[ l_i = \alpha + \beta w_i + \gamma y_i + X_i \delta + \epsilon_i \]

\( w_i \) is the net-of-tax wage rate

\( y_i \) measures non-labor income [including spousal earnings for couples]

\( X_i \) are demographic controls [age, experience, education, etc.]
NON-LINEAR BUDGET SETS

Non-linear budget set creates two problems:

1) Model mis-specification: OLS regression no longer recovers structural elasticity parameter $\varepsilon$ of interest

Two reasons: (a) underestimate response because people pile up at kink and (b) mis-estimate income effects

2) Econometric bias: $\tau_i$ depends on income $w_i/l_i$ and hence on $l_i$

Tastes for work are positively correlated with $\tau_i \rightarrow$ downward bias in OLS regression of hours worked on net-of-tax rates

Solution to problem #2: only use reform-based variation in tax rates. But problem #1 requires fundamentally different estimation method
OLD NON-LINEAR BUDGET SET METHOD

Issue addressed by non linear budget set studies pioneered by Hausman in late 1970s (Hausman, 1985 PE handbook chapter)

Method uses a structural model of labor supply to derive and estimate labor supply function fully consistent with theory

Key point: the method still uses the standard cross-sectional variation in pre-tax wages $w^p$ for identification. Taxes are seen as a problem to deal with rather than an opportunity for identification.

New literature identifying labor supply elasticities using tax changes has a totally different perspective: taxes are seen as an opportunity to identify labor supply
KEY ISSUE: $w$ correlated with tastes for work

\[ l_i = \alpha + \beta w_i + \gamma y_i + \epsilon_i \]

Identification is based on cross-sectional variation in $w_i$: comparing hours of work of highly skilled individuals (high $w_i$) to hours of work of low skilled individuals (low $w_i$)

If highly skilled workers have more taste for work (independent of the wage effect), then $\epsilon_i$ is positively correlated with $w_i$ leading to an upward bias in OLS

Plausible scenario: hard workers acquire better education and hence have higher wages

Controlling for $X_i$ can help but can never be sure that we have controlled for all the factors correlated with $w_i$ and tastes for work: Omitted variable bias

⇒ Tax changes provide more compelling identification
Natural Experiment Labor Supply Literature

Literature exploits variation in taxes/transfers to estimate Hours and Participation Elasticities

1) Large literature in labor/Public economics estimates effects of taxes and wages on hours worked and participation

2) Now discuss some estimates from this older literature
Negative Income Tax (NIT) Experiments

1) Best way to resolve identification problems: exogenously change taxes/transfers with a randomized experiment

2) NIT experiment conducted in 1960s/70s in Denver, Seattle, and other cities

3) First major social experiment in U.S. designed to test proposed transfer policy reform

4) Provided lump-sum welfare grants $G$ combined with a steep phaseout rate $\tau$ (50%-80%) [based on family earnings]


6) Several groups, with randomization within each; approx. $N = 75$ households in each group
Table 1
Parameters of the 11 Negative Income Tax Programs

<table>
<thead>
<tr>
<th>Program Number</th>
<th>G ($)</th>
<th>τ</th>
<th>Declining Tax Rate</th>
<th>Break-even Income ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3,800</td>
<td>.5</td>
<td>No</td>
<td>7,600</td>
</tr>
<tr>
<td>2</td>
<td>3,800</td>
<td>.7</td>
<td>No</td>
<td>5,429</td>
</tr>
<tr>
<td>3</td>
<td>3,800</td>
<td>.7</td>
<td>Yes</td>
<td>7,367</td>
</tr>
<tr>
<td>4</td>
<td>3,800</td>
<td>.8</td>
<td>Yes</td>
<td>5,802</td>
</tr>
<tr>
<td>5</td>
<td>4,800</td>
<td>.5</td>
<td>No</td>
<td>9,600</td>
</tr>
<tr>
<td>6</td>
<td>4,800</td>
<td>.7</td>
<td>No</td>
<td>6,857</td>
</tr>
<tr>
<td>7</td>
<td>4,800</td>
<td>.7</td>
<td>Yes</td>
<td>12,000</td>
</tr>
<tr>
<td>8</td>
<td>4,800</td>
<td>.8</td>
<td>Yes</td>
<td>8,000</td>
</tr>
<tr>
<td>9</td>
<td>5,600</td>
<td>.5</td>
<td>No</td>
<td>11,200</td>
</tr>
<tr>
<td>10</td>
<td>5,600</td>
<td>.7</td>
<td>No</td>
<td>8,000</td>
</tr>
<tr>
<td>11</td>
<td>5,600</td>
<td>.8</td>
<td>Yes</td>
<td>10,360</td>
</tr>
</tbody>
</table>

Source: Ashenfelter and Plant (1990), p. 403

How would you estimate the effect of the program if you cannot see hours worked?

Would you compare the payments made to the treatment group (post tax income) to post-tax income of control group?
1) Present non-parametric evidence of labor supply effects

2) Compare actual benefit payments to treated household vs. hypothetical benefit payments to control households

3) Difference in benefit payments reflects aggregates hours and participation responses

4) This is the relevant parameter for expenditure calculations and for welfare analysis

5) Shortcoming: approach does not decompose estimates into income and substitution effects and intensive vs. extensive margin

⇒ Hard to identify the key elasticity relevant for policy purposes and predict labor supply effect of other programs
<table>
<thead>
<tr>
<th>$G$ ($$)</th>
<th>$\tau$</th>
<th>Declining Tax Rate</th>
<th>Preexperimental Payment ($$)</th>
<th>Payments for Year of Experiment ($$)</th>
<th>Postexperimental Payment ($$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,800</td>
<td>.5</td>
<td>No</td>
<td>193.78</td>
<td>248.46, 368.95*, 389.24*</td>
<td>138.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(143.45)</td>
<td>(149.58), (170.75), (182.99)</td>
<td>(188.20)</td>
</tr>
<tr>
<td>3,800</td>
<td>.7</td>
<td>No</td>
<td>124.96</td>
<td>185.18, 317.28, 218.37</td>
<td>-47.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(223.77)</td>
<td>(237.91), (252.99), (325.57)</td>
<td>(314.66)</td>
</tr>
<tr>
<td>3,800</td>
<td>.7</td>
<td>Yes</td>
<td>-33.37</td>
<td>68.94, 158.44, 324.84</td>
<td>29.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(178.05)</td>
<td>(176.07), (213.59), (230.50)</td>
<td>(222.42)</td>
</tr>
<tr>
<td>3,800</td>
<td>.8</td>
<td>Yes</td>
<td>75.40</td>
<td>336.06, 221.54, 160.83</td>
<td>91.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(229.44)</td>
<td>(237.18), (245.92), (264.53)</td>
<td>(261.84)</td>
</tr>
<tr>
<td>4,800</td>
<td>.5</td>
<td>No</td>
<td>52.02</td>
<td>85.17, 294.55, 337.23</td>
<td>70.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(192.31)</td>
<td>(184.85), (201.73), (221.73)</td>
<td>(219.58)</td>
</tr>
<tr>
<td>4,800</td>
<td>.7</td>
<td>No</td>
<td>220.76</td>
<td>288.33, 496.85*, 543.25*</td>
<td>178.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(160.04)</td>
<td>(169.04), (197.88), (204.50)</td>
<td>(194.03)</td>
</tr>
<tr>
<td>4,800</td>
<td>.7</td>
<td>Yes</td>
<td>136.99</td>
<td>281.98*, 423.30*, 348.03*</td>
<td>23.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(127.36)</td>
<td>(137.19), (157.51), (162.38)</td>
<td>(140.58)</td>
</tr>
<tr>
<td>4,800</td>
<td>.8</td>
<td>Yes</td>
<td>-16.87</td>
<td>305.09, 417.90, 317.39</td>
<td>121.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(175.54)</td>
<td>(209.24), (234.32), (274.11)</td>
<td>(239.59)</td>
</tr>
<tr>
<td>5,600</td>
<td>.5</td>
<td>No</td>
<td>-163.12</td>
<td>200.75, 664.41*, 717.15*</td>
<td>124.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(252.05)</td>
<td>(258.13), (283.28), (280.65)</td>
<td>(287.04)</td>
</tr>
<tr>
<td>5,600</td>
<td>.7</td>
<td>No</td>
<td>-59.97</td>
<td>23.34, 386.12, 744.94*</td>
<td>267.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(164.95)</td>
<td>(156.41), (200.59), (263.80)</td>
<td>(259.45)</td>
</tr>
<tr>
<td>5,600</td>
<td>.8</td>
<td>Yes</td>
<td>-27.64</td>
<td>-51.03, 117.85, 273.44</td>
<td>121.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(121.47)</td>
<td>(126.67), (138.52), (157.96)</td>
<td>(169.26)</td>
</tr>
</tbody>
</table>

**Note.**—Terms are explained in text.

* Denotes mean is more than twice its standard error.
<table>
<thead>
<tr>
<th>G ($)</th>
<th>τ</th>
<th>Declining Tax Rate</th>
<th>Preexperimental Payment ($)</th>
<th>Payment for Year of Experiment ($)</th>
<th>Postexperimental Payment ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,800</td>
<td>.5</td>
<td>No</td>
<td>102.24</td>
<td>345.68</td>
<td>526.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(185.55)</td>
<td>(221.42)</td>
<td>(241.53)</td>
</tr>
<tr>
<td>3,800</td>
<td>.7</td>
<td>No</td>
<td>81.16</td>
<td>23.30</td>
<td>-99.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(309.85)</td>
<td>(316.06)</td>
<td>(330.14)</td>
</tr>
<tr>
<td>3,800</td>
<td>.7</td>
<td>Yes</td>
<td>6.99</td>
<td>490.00</td>
<td>176.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(234.01)</td>
<td>(288.13)</td>
<td>(272.87)</td>
</tr>
<tr>
<td>3,800</td>
<td>.8</td>
<td>Yes</td>
<td>-130.30</td>
<td>349.73</td>
<td>189.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(271.23)</td>
<td>(286.56)</td>
<td>(280.63)</td>
</tr>
<tr>
<td>4,800</td>
<td>.5</td>
<td>No</td>
<td>-23.66</td>
<td>30.15</td>
<td>160.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(183.73)</td>
<td>(208.90)</td>
<td>(199.26)</td>
</tr>
<tr>
<td>4,800</td>
<td>.7</td>
<td>No</td>
<td>-129.98</td>
<td>25.71</td>
<td>-4.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(185.46)</td>
<td>(208.14)</td>
<td>(211.44)</td>
</tr>
<tr>
<td>4,800</td>
<td>.7</td>
<td>Yes</td>
<td>75.66</td>
<td>224.96</td>
<td>387.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(234.21)</td>
<td>(280.43)</td>
<td>(367.56)</td>
</tr>
<tr>
<td>4,800</td>
<td>.8</td>
<td>Yes</td>
<td>467.89*</td>
<td>325.17</td>
<td>599.43*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(252.40)</td>
<td>(276.31)</td>
<td>(274.39)</td>
</tr>
<tr>
<td>5,600</td>
<td>.5</td>
<td>No</td>
<td>-224.97</td>
<td>560.51</td>
<td>723.08*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(286.39)</td>
<td>(298.21)</td>
<td>(306.90)</td>
</tr>
<tr>
<td>5,600</td>
<td>.7</td>
<td>No</td>
<td>-158.74</td>
<td>500.18</td>
<td>1194.68*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(239.17)</td>
<td>(311.24)</td>
<td>(416.25)</td>
</tr>
<tr>
<td>5,600</td>
<td>.8</td>
<td>Yes</td>
<td>-6.48</td>
<td>193.54</td>
<td>617.29*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(175.15)</td>
<td>(199.51)</td>
<td>(255.89)</td>
</tr>
</tbody>
</table>

**Note:** Terms are explained in text.
* Denotes mean is more than twice its standard error.

Source: Ashenfelter and Plant (1990), p. 407
NIT Experiments: Findings

1) Significant labor supply response but small overall

2) Implied earnings elasticity for males around 0.1

3) Implied earnings elasticity for women around 0.5

4) Academic literature not careful to decompose response along intensive and extensive margin

5) Response of women is concentrated along the extensive margin (can only be seen in official govt. report)

6) Earnings of treated women who were working before the experiment did not change much
From true experiment to “natural experiments”

True experiments are costly to implement and hence rare

However, real economic world (nature) provides variation that can be exploited to estimate behavioral responses ⇒ “Natural Experiments”

Natural experiments sometimes come very close to true experiments: Imbens, Rubin, Sacerdote AER ’01 did a survey of lottery winners and non-winners matched to Social Security administrative data to estimate income effects

Lottery generates random assignment conditional on playing

Find significant but relatively small income effects: $\eta = w \partial l / \partial y$ between -0.05 and -0.10

Identification threat: differential response-rate among groups
Figure 2. Proportion with Positive Earnings for Nonwinners, Winners, and Big Winners

Note: Solid line = nonwinners; dashed line = winners; dotted line = big winners.

On average the individuals in our basic sample won yearly prizes of $26,000 (averaged over the $55,000 for winners and zero for nonwinners). Typically they won 10 years prior to completing our survey in 1996, implying they are on average halfway through their 20 years of lottery payments when they responded in 1996. We asked all individuals how many tickets they bought in a typical week in the year they won the lottery. As expected, the number of tickets bought is considerably higher for winners than for nonwinners. On average, the individuals in our basic sample are 50 years old at the time of winning, which, for the average person was in 1986; 35 percent of the sample was over 55 and 15 percent was over 65 years old at the time of winning; 63 percent of the sample was male. The average number of years of schooling, calculated as years of high school plus years of college plus 8, is equal to 13.7; 64 percent claimed at least one year of college.

We observe, for each individual in the basic sample, Social Security earnings for six years preceding the time of winning the lottery, for the year they won (year zero), and for six years following winning. Average earnings, in terms of 1986 dollars, rise over the pre-winning period from $13,930 to $16,330, and then decline back to $13,290 over the post-winning period. For those with positive Social Security earnings, average earnings rise over the entire 13-year period from $20,180 to $24,300. Participation rates, as measured by positive Social Security earnings, gradually decline over the 13 years, starting at around 70 percent before going down to 56 percent. Figures 1 and 2 present graphs for average earnings and the proportion of individuals with positive earnings for the three groups, nonwinners, winners, and big winners. One can see a modest decline in earnings and proportion of individuals with positive earnings for the full winner sample compared to the nonwinners after winning the lottery, and a sharp and much larger decline for big winners at the time of winning. A simple difference-in-differences type estimate of the marginal propensity to earn out of unearned income (mpe) can be based on the ratio of the difference in the average change in earnings before and after winning the lottery for two groups and the difference in the average prize for the same two groups. For the winners, the difference in average earnings over the six post-lottery years and the six pre-lottery years is -$1,877 and for the nonwinners the average change is $448. Given a difference in average prize of $55,000 for the winner/nonwinners comparison, the estimated mpe is \((-1,877 - 448)/(55,000 - 0) = -0.042\) (SE 0.016). For the big-winners/small-winners comparison, this estimate is -0.059 (SE 0.018). In Section IV we report estimates for this quantity using more sophisticated analyses.

On average the value of all cars was $18,200. For housing the average value was $166,300, with an average mortgage of $44,200. We aggregated the responses to financial wealth into two categories. The first concerns retirement.
1) Identify elasticities from 1980-2000 using grouping instrument

a) Define cells (year/age/education) and compute mean wages

b) Instrument for actual wage with mean wage in cell

2) Identify purely from group-level variation, which is less contaminated by individual endogenous choice

3) Results: (a) total hours elasticity for married women (including int + ext margin) shrank from 0.4 in 1980 to 0.2 today, (b) effect of husband earnings ↓ overtime

4) Interpretation: elasticities shrink as women become more attached to the labor force
Summary of Static Labor Supply Literature

1) Small elasticities for prime-age males

Probably institutional restrictions, need for at least one income, etc. prevent a short-run response

2) Larger responses for workers who are less attached to labor force: Married women, low income earners, retirees

3) Responses driven primarily by extensive margin

a) Extensive margin (participation) elasticity around 0.2-0.5

b) Intensive margin (hours) elasticity smaller
Responses to Low-Income Transfer Programs

1) Particular interest in treatment of low incomes in a progressive tax system: are they responsive to incentives?

2) Complicated set of transfer programs in US
   a) In-kind: food stamps, Medicaid, public housing, job training, education subsidies
   b) Cash: TANF, EITC, SSI

3) See Gruber undergrad textbook for details on institutions
Overall Costs of Anti Poverty Programs

1) US government (fed+state and local) spent $800bn in 2013 on income-tested programs

a) About 4% of GDP but 15% of $5 Trillion govt budget (fed+state+local).

b) About 50% is health care (Medicaid)

2) Only $200 billion in cash (1% of GDP, or 25% of transfer spending)
1996 US Welfare Reform

1) Largest change in welfare policy

2) Reform modified AFDC (Aid to Families with Dependent Children) cash welfare program to provide more incentives to work (renamed TANF)
   a) Requiring recipients to go to job training or work
   b) Limiting the duration for which families able to receive welfare
   c) Reducing phase-out rate of benefits

3) Variation across states because Fed govt. gave block grants with guidelines (no categorical grants).

4) EITC also expanded during this period: general shift from welfare to “workfare”
The landscape providing assistance to poor families with children has changed substantially.
Annual Employment Rates for Women
By Marital Status and Presence of Children, 1980-2009

Welfare Reform: Two Empirical Questions

1) Incentives: did welfare reform actually increase labor supply?
   a) Test whether EITC expansions affect labor supply
   b) Use state welfare randomized experiments implemented before reform to assess effects of switch from AFDC to TANF

2) Benefits: did removing many people from transfer system reduce their welfare? How did consumption change?
   Focus on single mothers, who were most impacted by reform
Earned Income Tax Credit (EITC) program

Hotz-Scholz ’04, Eissa-Hoynes ’06, Nichols-Rothstein ’15 provide detailed surveys

1) EITC started small in the 1970s but was expanded in 1986-88, 1994-96, 2008-09: today, largest means-tested cash transfer program [$60bn in 2012, 25m families recipients]

2) Eligibility: families with kids and low earnings.

3) Refundable Tax credit: administered as annual tax refund received in Feb-April, year $t+1$ (for earnings in year $t$)

4) EITC has flat pyramid structure with phase-in (negative MTR), plateau, (0 MTR), and phase-out (positive MTR)

5) States have added EITC components to their income taxes [in general a percentage of the Fed EITC, great source of natural experiments, understudied bc CPS too small]
Source: Federal Govt
Figure 1: Earned Income Tax Credit by Number of Children and Filing Status, 2013

Figure 2. Maximum credit over time, constant 2013 dollars, by number of children

Source: Nichols and Rothstein (2015)
Theoretical Behavioral Responses to the EITC

**Extensive margin**: positive effect on Labor Force Participation

**Intensive margin**: earnings conditional on working, mixed effects

1) Phase in: (a) Substitution effect: work more due to wage subsidy, (b) Income effect: work less $\Rightarrow$ Net effect: ambiguous; probably work more

2) Plateau: Pure income effect (no change in net wage) $\Rightarrow$ Net effect: work less

3) Phase out: (a) Substitution effect: work less, (b) Income effect: also work less $\Rightarrow$ Net effect: work less

Should expect bunching at the EITC kink points
Eissa and Liebman 1996

1) Pioneering study of labor force participation of single mothers before/after 1986-7 EITC expansion using CPS data

2) Limitation: this expansion was relatively small

3) Diff-in-Diff strategy:
   a) Treatment group: single women with kids
   b) Control group: single women without kids
LABOR SUPPLY RESPONSE TO THE EITC

FIGURE IV
1986 and 1988 Earned Income Tax Credit

Diff-in-Diff (DD) Methodology:

Step 1: Simple Difference

Outcome:  *LFP* (labor force participation)

Two groups: Treatment group (T) which faces a change [single women with kids] and control group (C) which does not [single women without kids]

Simple Difference estimate: \( D = LFP^{T} - LFP^{C} \) captures treatment effect if absent the treatment, *LFP* equal across 2 groups

Note: this assumption always holds when *T* and *C* status is randomly assigned

Test for this assumption: Compare *LFP* before treatment happened \( D_{B} = LFP_{B}^{T} - LFP_{B}^{C} \)
Diff-in-Diff (DD) Methodology:

Step 2: Diff-in-Difference (DD)

If \( D_B \neq 0 \), can estimate DD:

\[
DD = D_A - D_B = LFP_A^T - LFP_A^C - [LFP_B^T - LFP_B^C]
\]

(A = after reform, B = before reform)

DD is unbiased if parallel trend assumption holds:

Absent the change, difference across \( T \) and \( C \) would have stayed the same before and after.

OLS Regression estimation of DD:

\[
LFP_{it} = \beta_0 AFTER + \beta_1 TREAT + \gamma AFTER \cdot TREAT + \varepsilon
\]

\[
\hat{\gamma}_{OLS} = LFP_A^T - LFP_A^C - [LFP_B^T - LFP_B^C]
\]
TABLE II
LABOR FORCE PARTICIPATION RATES OF UNMARRIED WOMEN

<table>
<thead>
<tr>
<th></th>
<th>Pre-TRA86 (1)</th>
<th>Post-TRA86 (2)</th>
<th>Difference (3)</th>
<th>Difference-in-differences (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Treatment group:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With children</td>
<td>0.729 (0.004)</td>
<td>0.753 (0.004)</td>
<td>0.024 (0.006)</td>
<td></td>
</tr>
<tr>
<td>[20,810]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Control group:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without children</td>
<td>0.952 (0.001)</td>
<td>0.952 (0.001)</td>
<td>0.000 (0.002)</td>
<td>0.024 (0.006)</td>
</tr>
<tr>
<td>[46,287]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B. Treatment group:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school, with children</td>
<td>0.479 (0.010)</td>
<td>0.497 (0.010)</td>
<td>0.018 (0.014)</td>
<td></td>
</tr>
<tr>
<td>[5396]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Control group 1:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than high school, without children</td>
<td>0.784 (0.010)</td>
<td>0.761 (0.009)</td>
<td>-0.023 (0.013)</td>
<td>0.041 (0.019)</td>
</tr>
<tr>
<td>[3958]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Control group 2:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beyond high school, with children</td>
<td>0.911 (0.005)</td>
<td>0.920 (0.005)</td>
<td>0.009 (0.007)</td>
<td>0.009 (0.015)</td>
</tr>
<tr>
<td>[5712]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>C. Treatment group:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school, with children</td>
<td>0.764 (0.006)</td>
<td>0.787 (0.006)</td>
<td>0.023 (0.008)</td>
<td></td>
</tr>
<tr>
<td>[9702]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Control group 1:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school, without children</td>
<td>0.945 (0.002)</td>
<td>0.943 (0.003)</td>
<td>-0.002 (0.004)</td>
<td>0.025 (0.009)</td>
</tr>
<tr>
<td>[16,527]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Control group 2:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beyond high school, with children</td>
<td>0.911 (0.005)</td>
<td>0.920 (0.005)</td>
<td>0.009 (0.007)</td>
<td>0.014 (0.011)</td>
</tr>
<tr>
<td>[5712]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Diff-in-Diff (DD) Methodology

DD most convincing when groups are very similar to start with [closer to randomized experiment]

Should always test DD using data from more periods and plot the two time series to check parallel trend assumption

Use alternative control groups [not as convincing as potential control groups are many]

In principle, can create a DDD as the difference between actual DD and $DD_{Placebo}$ (DD between 2 control groups). However, DDD of limited interest in practice because

(a) if $DD_{Placebo} \neq 0$, DD test fails, hard to believe DDD removes bias

(b) if $DD_{Placebo} = 0$, then DD=DDD but DDD has higher s.e.
All Unmarried Females

![Graph showing the proportion of all unmarried females over time, with lines for those without children and with children.](image-url)

Unmarried Males With Less Than High School Education

![Graph showing labor force participation rates for unmarried males with less than high school education from 1981 to 1991. The graph indicates a general decline in participation rates over the years.]

**Figure II**


Diff-in-Diff (DD) Methodology

1) DD sensitive to functional form (e.g. log vs levels) when $D_{\text{before}} \neq 0$.

Example: $T \uparrow$ from 40% to 50% and $C \uparrow$ from 15% to 20%:

$\text{DD}_{\text{level}} = [50 - 40] - [20 - 15] = 5$ but $\text{DD}_{\text{log}} = \log[50/40] - \log[20/15] = -.06$

2) To obtain elasticity estimate, need to take ratio of $\text{DD}_{\text{outcome}}$ to $\text{DD}_{\text{policy change}}$ to form the Wald estimate:

$$\hat{e} = \frac{[\log \text{LFP}_T^A - \log \text{LFP}_A^C] - [\log \text{LFP}_B^T - \log \text{LFP}_B^C]}{\log(1 - \tau_T^A) - \log(1 - \tau_A^C)] - [\log(1 - \tau_T^B) - \log(1 - \tau_B^C)]}$$

$\text{DD}_{\text{policy change}}$ is the 1st stage, $\text{DD}_{\text{outcome}}$ is the reduced form effect, the ratio is the 2nd stage estimate

Wald estimated with 2SLS regression:

$$\text{LFP}_{it} = \beta_0 \text{AFTER} + \beta_1 \text{TREAT} + e \cdot \log(1 - \tau) + \varepsilon$$

where $\log(1 - \tau)$ is instrumented with interaction $\text{AFTER} \cdot \text{TREAT}$
Eissa and Liebman 1996: Results

1) Find a small but significant DD effect: 2.4% (larger DD effect 4% among women with low education) \( \Rightarrow \) Translates into substantial participation elasticities above 0.5

2) Note the labor force participation for women with/without children are not great comparison groups (70% LFP vs. +90%): time series evidence is only moderately convincing

3) Subsequent studies have used much bigger EITC expansions of the mid 1990s and also find positive effects on labor force participation of single women/single mothers (but contaminated by AFDC to TANF transition)

4) Conventional standard errors probably overstate precision
Show that conventional standard errors in fixed effects regressions with state reform variation are too low.

Randomly generated placebo state laws: half the states pass law at random date. $I_{st}$ is one if state $s$ has law in place at time $t$.

Use female wages $w_{ist}$ in CPS data and run OLS:

$$\log w_{ist} = A_s + B_t + bI_{st} + \varepsilon_{ist}$$

$\hat{b}$ significant (at 5% level) in 65% of cases $\Rightarrow \varepsilon_{ist}$ are not iid.

Clustering by state*year cells is not enough (significant 45% of the time).

Need to cluster at state level to obtain reasonable s.e. because of strong serial correlation within states.
Welfare Reform Effects on Consumption

Meyer and Sullivan ’04 examine consumption of single mothers and their families from 1984–2000 using CEX data

1) Material conditions of single mothers did not decline in 1990s, either in absolute terms or relative to single childless women or married mothers

2) In most cases, evidence suggests that the material conditions of single mothers have improved slightly

3) Question: is this because economy was booming in 1990s?

4) Is workfare approach more problematic in current economy? [SNAP households surged from 12M in ’07 to 20M in ’10 while TANF households increased slightly from 1.7M in ’07 to 1.85M in ’10]
Key prediction of standard labor supply model: individuals should bunch at (convex) kink points of the budget set

1) The only non-parametric source of identification for intensive elasticity in a single cross-section of earnings is amount of bunching at kinks creating by tax/transfer system

2) Saez ‘10 develops method of using bunching at kinks to estimate the compensated income elasticity

Formula for elasticity: \( \varepsilon^c = \frac{dz/z^*}{dt/(1-t)} = \frac{\text{excess mass at kink}}{\text{change in NTR}} \)

⇒ Amount of bunching proportional to compensated elasticity
elasticity would no longer be a pure compensated elasticity, but a mix of the compensated elasticity and the uncompensated elasticity. Four points should be noted. First, the larger the behavioral elasticity, the more bunching we should expect. Unsurprisingly, if there are no behavioral responses to marginal tax rates, there

Panel A. Indifference curves and bunching

Before tax income $z$

After-tax income $c = z - T(z)$

Panel B. Density distributions and bunching

Before reform density

After reform density

P re-reform incomes between $z^*$ and $z^* + dz^*$ bunch at $z^*$ after reform

Figure 1. Bunching Theory

Notes: Panel A displays the effect on earnings choices of introducing a (small) kink in the budget set by increasing the tax rate $t$ by $dt$ above income level $z^*$. Individual $L$ who chooses $z^*$ before the reform stays at $z^*$ after the reform. Individual $H$ chooses $z^*$ after the reform and was choosing $z^* + dz^*$ before the reform. Panel B depicts the effects of introducing the kink on the earnings density distribution. The pre-reform density is smooth around $z^*$. After the reform, all individuals with income between $z^*$ and $z^* + dz^*$ before the reform, bunch at $z^*$, creating a spike in the density distribution. The density above $z^* + dz^*$ shifts to $z^*$ (so that the resulting density and is no longer smooth at $z^*$).

Source: Saez (2010), p. 184
elasticity would no longer be a pure compensated elasticity, but a mix of the compensated elasticity and the uncompensated elasticity. Four points should be noted.

First, the larger the behavioral elasticity, the more bunching we should expect. Unsurprisingly, if there are no behavioral responses to marginal tax rates, there

Panel A . Indifference curves and bunching

Before tax income $z$

Slope 1 $-t$

$z^*$

$dz^*$

Slope 1 $-t-\Delta t$

Individual L chooses $z^*$ before and after reform

Individual H chooses $z^* + dz^*$ before and after reform

$dz^*/z^* = e \Delta t/(1-t)$ with compensated elasticity

Individually H indifference curves

Panel B. Density distributions and bunching

Pre-reform incomes between $z^*$ and $z^* + dz^*$ bunch at $z^*$ after reform

Density distribution

Before reform density

After reform density

Before tax income $z$

Source: Saez (2010), p. 184
Bunching at Kinks (Saez AEJ-EP’10)

1) Uses individual tax return micro data (IRS public use files) from 1960 to 2004

2) Advantage of dataset over survey data: very little measurement error

3) Finds bunching around:
   a) First kink point of the Earned Income Tax Credit (EITC), especially for self-employed
   b) At threshold of the first tax bracket where tax liability starts, especially in the 1960s when this point was very stable

4) However, no bunching observed around all other kink points
EITC Amount as a Function of Earnings

- **Subsidy: 40%**
  - Married, 2+ kids
  - Single, 2+ kids
- **Subsidy: 34%**
  - Single, 1 kid
  - Married, 1 kid
- **Phase-out tax: 16%**
  - No kids
- **Phase-out tax: 21%**

Source: Federal Govt
indexes earnings to 2008 using the IRS inflation parameters, so that the EITC kinks are perfectly aligned for all years.

Two elements are worth noting in Figure 3. First, there is a clear clustering of tax filers around the first kink point of the EITC. In both panels, the density is maximum exactly at the first kink point. The fact that the location of the first kink point differs between EITC recipients with one child, versus those with two or more children, constitutes strong evidence that the clustering is driven by behavioral responses to the EITC as predicted by the standard model. Second, however, we cannot discern any

Figure 3. Earnings Density Distributions and the EITC

Notes: The figure displays the histogram of earnings (by $500 bins) for tax filers with one dependent child (panel A) and tax filers with two or more dependent children (panel B). The histogram includes all years 1995–2004 and inflates earnings to 2008 dollars using the IRS inflation parameters (so that the EITC kinks are aligned for all years).

Earnings are defined as wages and salaries plus self-employment income (net of one-half of the self-employed payroll tax). The EITC schedule is depicted in dashed line and the three kinks are depicted with vertical lines. Panel A is based on 57,692 observations (representing 116 million tax returns), and panel B on 67,038 observations (representing 115 million returns).

Source: Saez (2010), p. 191
Indexing earnings to 2008 using the IRS inflation parameters, so that the EITC kinks are perfectly aligned for all years.

Two elements are worth noting in Figure 3. First, there is a clear clustering of tax filers around the first kink point of the EITC. In both panels, the density is maximum exactly at the first kink point. The fact that the location of the first kink point differs between EITC recipients with one child, versus those with two or more children, constitutes strong evidence that the clustering is driven by behavioral responses to the EITC as predicted by the standard model. Second, however, we cannot discern any

Panel A. One child

Earnings density ($500 bins)

Earnings (2008 $)

Density  EIC Amount

Source: Saez (2010), p. 191
systematic clustering around the second kink point of the EITC. Similarly, we cannot discern any gap in the distribution of earnings around the concave kink point where the EITC is completely phased-out. This differential response to the first kink point, versus the other kink points, is surprising in light of the standard model predicting that any convex (concave) kink should produce bunching (gap) in the distribution of earnings.

In Figure 4, we break down the sample of earners into those with nonzero self-employment income versus those with zero self-employment income (and hence whose

\[
\begin{align*}
\text{Earnings density} & \quad \text{Earnings (2008 $)} \\
0 & \quad 5,000 \quad 10,000 \quad 15,000 \quad 20,000 \quad 25,000 \quad 30,000 \quad 35,000 \quad 40,000 \quad 45,000 \quad 50,000 \\
\end{align*}
\]

Panel A. One child

Panel B. Two or more children

Wage earners
Self-employed
EIC amount

Figure 4. Earnings Density and the EITC: Wage Earners versus Self-Employed

Notes: The figure displays the kernel density of earnings for wage earners (those with no self-employment earnings) and for the self-employed (those with nonzero self-employment earnings). Panel A reports the density for taxpayers with one dependent child and panel B for taxpayers with two or more dependent children. The charts include all years 1995–2004. The bandwidth is $400 in all kernel density estimations. The fraction self-employed in 16.1 percent and 20.5 percent in the population depicted on panels A and B (in the data sample, the unweighted fraction self-employed is 32 percent and 40 percent). We display in dotted vertical lines around the first kink point the three bands used for the elasticity estimation with \(\delta = $1,500\).

Source: Saez (2010), p. 192
systematic clustering around the second kink point of the EITC. Similarly, we cannot discern any gap in the distribution of earnings around the concave kink point where the EITC is completely phased-out. This differential response to the first kink point, versus the other kink points, is surprising in light of the standard model predicting that any convex (concave) kink should produce bunching (gap) in the distribution of earnings.

In Figure 4, we break down the sample of earners into those with nonzero self-employment income versus those zero self-employment income (and hence whose

Panel A. One child

<table>
<thead>
<tr>
<th>Earnings density</th>
<th>Earnings (2008 $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5,000</td>
<td>10,000</td>
</tr>
<tr>
<td>10,000</td>
<td>15,000</td>
</tr>
<tr>
<td>15,000</td>
<td>20,000</td>
</tr>
<tr>
<td>20,000</td>
<td>25,000</td>
</tr>
<tr>
<td>25,000</td>
<td>30,000</td>
</tr>
<tr>
<td>30,000</td>
<td>35,000</td>
</tr>
<tr>
<td>35,000</td>
<td>40,000</td>
</tr>
<tr>
<td>40,000</td>
<td>45,000</td>
</tr>
<tr>
<td>45,000</td>
<td>50,000</td>
</tr>
</tbody>
</table>

Panel B. Two or more children

<table>
<thead>
<tr>
<th>Earnings density</th>
<th>Earnings (2008 $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5,000</td>
<td>10,000</td>
</tr>
<tr>
<td>10,000</td>
<td>15,000</td>
</tr>
<tr>
<td>15,000</td>
<td>20,000</td>
</tr>
<tr>
<td>20,000</td>
<td>25,000</td>
</tr>
<tr>
<td>25,000</td>
<td>30,000</td>
</tr>
<tr>
<td>30,000</td>
<td>35,000</td>
</tr>
<tr>
<td>35,000</td>
<td>40,000</td>
</tr>
<tr>
<td>40,000</td>
<td>45,000</td>
</tr>
<tr>
<td>45,000</td>
<td>50,000</td>
</tr>
</tbody>
</table>

Figure 4. Earnings Density and the EITC: Wage Earners versus Self-Employed

Notes: The figure displays the kernel density of earnings for wage earners (those with no self-employment earnings) and for the self-employed (those with nonzero self-employment earnings). Panel A reports the density for taxpayers with one dependent child and panel B for taxpayers with two or more dependent children. The charts include all years 1995–2004. The bandwidth is $400 in all kernel density estimations. The fraction self-employed in 16.1 percent and 20.5 percent in the population depicted on panels A and B (in the data sample, the unweighted fraction self-employed is 32 percent and 40 percent). We display in dotted vertical lines around the first kink point the three bands used for the elasticity estimation with $\delta = $1,500.

Source: Saez (2010), p. 192
Why not more bunching at kinks?

1) True intensive elasticity of response may be small

2) Randomness in income generation process: Saez (1999) shows that year-to-year income variation too small to erase bunching if elasticity is large

3) Frictions: Adjustment costs and institutional constraints (Chetty, Friedman, Olsen, and Pistaferri QJE’11)

4) Information and salience:

Chetty-Friedman-Saez AER’13 show how information about EITC affects bunching at kink point
Chetty, Friedman, Olsen, and Pistaferri QJE’11

1) If workers face adjustment costs, may not reoptimize in response to tax changes of small size and scope in short run

a) Search costs, costs of acquiring information about taxes

b) Institutional constraints imposed by firms (e.g. 40 hour week) that does not apply to the self-employed or workers with more flexibility (e.g. secondary earners)

2) Question: How much are elasticity estimates affected by frictions?
Chetty et al. 2011: Administrative data

Matched employer-employee panel data with admin tax records for full population of Denmark matching employee-employer information

Sample restriction: Wage-earners aged 15-70, 1994-2001

Approximately 2.42 million people per year

Important development in empirical micro in recent years: shift from survey data to administrative data (Card-Chetty-Feldstein-Saez ’10 and Einav and Levin NBER’13)
Value of Administrative data

Key advantages of admin data (in most advanced countries such as Scandinavia):

1) Size (often full population available)

2) Longitudinal structure (can follow individual across years)

3) Ability to match wide variety of data (tax records, earnings records, family records, health records, education records)

US is lagging behind in terms of admin data access [hard to match across agencies]

Private sector also generates valuable **big data** (Google, Credit Bureaus, Personnel/health data from large companies)
Marginal Tax Rates in Denmark in 1995

Source: Chetty et al. (2009)

Note: $1 \approx 6$ DKr
Income Distribution for Wage Earners Around Top Kink (1994-2001)

- Excess mass = 5.97%
- Standard error = 0.38%

Source: Chetty et al. (2009)
Single Men

Excess mass = 1.83%
Standard error = 0.34%

Source: Chetty et al. (2009)
Married Women

Excess mass = 14.1%
Standard error = 0.90%

Source: Chetty et al. (2009)
Married Women at the Middle Tax: 10% Tax Kink

Excess mass = 2.24%
Standard error = 0.46%

Source: Chetty et al. (2009)
Observed Elasticity vs. Size of Tax Change
Married Female Wage Earners

Chetty et al. 2009
Self Employed: Top Kink

Excess mass = 150.0%
Standard error = 1.4%

Chetty et al. 2009
Self-Employed: Middle Kink

Excess mass = 11.2%
Standard error = 0.72%

Chetty et al. 2009
Chetty et al. 2011: Results

1) Search costs attenuate observed behavioral responses substantially: find larger elasticities around large kink points

2) Groups with more flexibility respond more (secondary earners, self-employed)

3) Overall elasticities estimated from bunching are small in magnitude perhaps because frictions prevent full response

⇒ Bunching methods are good to detect behavioral responses but not necessarily to pin down magnitude of a long-run response to a large tax reform
EITC Behavioral Studies

Strong evidence of response along extensive margin, little evidence of response along intensive margin (except for self-employed) ⇒ Possibly due to lack of understanding of the program

Qualitative surveys show that:

Low income families know about EITC and understand that they get a tax refund if they work

However very few families know whether tax refund ↑ or ↓ with earnings

Such confusion might be good for the government as the EITC induces work along participation margin without discouraging work along intensive margin
Use US population wide tax return data since 1996 (through IRS special contract)

1) Substantial heterogeneity in fraction of EITC recipients bunching (using self-employment) across geographical areas

⇒ Information on EITC varies across areas and grows overtime

2) Places with high self-employment EITC bunching display wage earnings distribution more concentrated around plateau

3) Omitted variable test: use birth of first child to test causal effect of EITC on wage earnings

⇒ Evidence of wage earnings response to EITC along intensive margin
Earnings Distributions in Lowest and Highest Bunching Deciles

Source: Chetty, Friedman, and Saez NBER'12
Fraction of Tax Filers Who Report SE Income that Maximizes EITC Refund in 1996

Source: Chetty, Friedman, and Saez NBER'12
Fraction of Tax Filers Who Report SE Income that Maximizes EITC Refund in 1999

Source: Chetty, Friedman, and Saez NBER'12
Fraction of Tax Filers Who Report SE Income that Maximizes EITC Refund in 2002

Source: Chetty, Friedman, and Saez NBER’12
Fraction of Tax Filers Who Report SE Income that Maximizes EITC Refund in 2005

Source: Chetty, Friedman, and Saez NBER'12
Fraction of Tax Filers Who Report SE Income that Maximizes EITC Refund in 2008

Source: Chetty, Friedman, and Saez NBER’12
Income Distribution For Single Wage Earners with One Child

Is the EITC having an effect on this distribution?

Source: Chetty, Friedman, and Saez NBER'12
Income Distribution For Single Wage Earners with One Child
High vs. Low Bunching Areas

Source: Chetty, Friedman, and Saez NBER'12
Earnings Distribution in the Year of First Child Birth for Wage Earners

Source: Chetty, Friedman, and Saez NBER'12
IMPLICATIONS OF ROLE OF INFORMATION

Empirical work:

Information should be a key explanatory variable in estimation of behavioral responses to govt programs

When doing empirical project, always ask the question: did people affected understand incentives?

Cannot identify structural parameters of preferences without modeling information and salience

Normative analysis:

Information is a powerful and inexpensive policy tool to affect behavior

Should be incorporated into optimal policy design problems
Bunching at Notches

Taxes and transfers sometimes also generate notches (=discontinuities) in the budget set.

Such discontinuities should create bunching (and gaps) in the resulting distributions.

Example: Pakistani income tax creates notches because average tax rate jumps ⇒ Bunching below the notch and gap in density just above the notch.

Empirically: Kleven and Waseem QJE’13 find evidence of bunching (primarily among self-employed) but size of the response is quantitatively small.

Large fraction of taxpayers are unresponsive to notch likely due to lack of information.
FIGURE 3
Personal Income Tax Schedules in Pakistan

Notes: the figure shows the statutory (average) tax rate as a function of annual taxable income in the personal income tax schedules for wage earners (red dashed line) and self-employed individuals and unincorporated firms (blue solid line), respectively. Taxable income is shown in thousands of Pakistani Rupees (PKR), and the PKR-USD exchange rate is around 85 as of April 2011. The schedule for the self-employed applies to the full period of this study (2006-08), while the schedule for wage earners applies only to 2006-07 and was changed by a tax reform in 2008. The tax system classifies individuals as either wage earners or self-employed in the year of assessment.

Source: Kleven and Waseem '11
FIGURE 1
Effect of Notch on Taxpayer Behavior

Panel A: Bunching at the Notch

Source: Kleven and Waseem '11
FIGURE 2
Effect of Notch on Density Distribution

Panel A: Theoretical Density Distributions

Source: Kleven and Waseem '11
FIGURE 5
Density Distribution around Middle Notches:
Self-Employed Individuals and Firms (Sophisticated Filers)

Panel A: Notch at 300k

Panel B: Notch at 400k

Panel C: Notch at 500k

Panel D: Notch at 600k

Source: Kleven and Waseem '11
Kleven and Waseem QJE’13 notch analysis

With optimization frictions (lack of information, costs of adjustment), a fraction of individuals fail to respond to notch.

Kleven-Waseem use empirical density in the theoretical gap area to measure the fraction of unresponsive individuals.

This allows them to back up the frictionless elasticity (i.e. the elasticity among responsive individuals).

The frictionless elasticity is much higher than the reduced form elasticity but remains still relatively modest.

Additional notch studies: Best and Kleven ’14 on UK housing purchase tax (stamp duty), Kopczuk-Munroe AEJ’15 on NY-NJ Mansion tax [also find evidence of bunching responses]
Many Recent Bunching Studies

- Individual tax (Bastani-Selin ’14 on Sweden)
- Payroll tax (Tazhidinova ’15 on UK)
- Corporate tax (Devereux-Liu-Loretz ’13)
- Health spending (Einav-Finkelstein-Schrimpf ’13 on Medicare Part D)
- Retirement savings (401(k) matches)
- Retirement age (Brown ’13 on California Teachers)
- Housing transactions (Best and Kleven, 2014)

General findings:

1. clear bunching when information is salient and outcome easily manipulable
2. bunching is almost always small relative to conventional elasticity estimates
Intertemporal substitution: Tax Holiday in Iceland

In 1987, Iceland transitioned from paying taxes on previous year’s income to current income.

To avoid double taxation during transition, no tax charged over 1987 incomes.

Average tax rate of 14.5% in 1986, 0% in 1987, 8% in 1988.

Reform announced in late 1986 ⇒ unanticipated temporary tax change.

Temporary change in incentives ⇒ ideal quasi-experiment to intertemporal substitution elasticity (work hard in 1987, take a break in 1986 or 1988).

Bianchi et al. AER’01 look at employment effects [hard to know what counterfactual is]
Figure 1a: 1987 Tax Holiday in Iceland
Intertemporal Substitution: High Frequency Studies

1) Recent literature focuses on high frequency substitution

2) Focus on groups with highly flexible and well measured labor supply such as:

   a) cab drivers [Camerer et al. QJE’97, Farber JPE’05, AER-PP’08, Crawford-Meng ’09]: debate on whether cab drivers are rational or have a daily income target

   b) stadium vendors [Oettinger JPE’99]

   c) cycling messengers randomized experiment [Fehr-Goette AER’07]
Macro Long-Run Evidence

1) Macroeconomists also estimate elasticities by examining long-term trends/cross-country comparisons

2) Identification more questionable but estimates perhaps more relevant to long-run policy questions of interest

3) Use aggregate hours data and aggregate measures of taxes (average tax rates)

4) Highly influential in calibration of macroeconomic models
Trend-based Estimates and Macro Evidence

**Long-Run:** US real wage rates multiplied by about 5 from 1900 to present due to economic growth.

Aged 25-54 male hours have fallen 25% and then stabilized (Ramey and Francis AEJ-macro ’09).

⇒ Uncompensated hours of work elasticity is small (< .1).

However, taxes are rebated as transfers so can still have labor supply effect of taxes if compensated elasticity (or income effects) large.

Alternative plausible story: utility depends on relative consumption ⇒ Earnings only $10,000 would seem good in 1900 but low today.
Figure 2. Average Weekly Hours Worked per Person, by Age Group

Source: Authors' estimates, based on information from Kendrick (1961, 1973), the census, and the CPS.

Ramey and Francis AEJ'09
Long-run cross-country panel: Prescott 2005

Uses data on hours worked by country in 1970 and 1995 for 7 OECD countries [total hours/people age 15-64]

Technique to identify elasticity: calibration of GE model

Rough intuition: posit a labor supply model, e.g.

$$u(c, l) = c - \frac{1^{1+1/\varepsilon}}{1 + 1/\varepsilon}$$

Finds that elasticity of $\varepsilon = 1.2$ best matches time series and cross-sectional patterns

Note that this is analogous to a regression without controls for other variables

Results verified in subsequent calibrations by Ohanina-Raffo-Rogerson JME’08 and others using more data
Table 2

Actual and Predicted Labor Supply

In Selected Countries in 1993–96 and 1970–74

<table>
<thead>
<tr>
<th>Period</th>
<th>Country</th>
<th>Labor Supply*</th>
<th>Differences (Predicted Less Actual)</th>
<th>Prediction Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Actual</td>
<td>Predicted</td>
<td></td>
</tr>
<tr>
<td>1993–96</td>
<td>Germany</td>
<td>19.3</td>
<td>19.5</td>
<td>.2</td>
</tr>
<tr>
<td></td>
<td>France</td>
<td>17.5</td>
<td>19.5</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Italy</td>
<td>16.5</td>
<td>18.8</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>22.9</td>
<td>21.3</td>
<td>−1.6</td>
</tr>
<tr>
<td></td>
<td>United Kingdom</td>
<td>22.8</td>
<td>22.8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Japan</td>
<td>27.0</td>
<td>29.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>25.9</td>
<td>24.6</td>
<td>−1.3</td>
</tr>
<tr>
<td>1970–74</td>
<td>Germany</td>
<td>24.6</td>
<td>24.6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>France</td>
<td>24.4</td>
<td>25.4</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>Italy</td>
<td>19.2</td>
<td>28.3</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>Canada</td>
<td>22.2</td>
<td>25.6</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>United Kingdom</td>
<td>25.9</td>
<td>24.0</td>
<td>−1.9</td>
</tr>
<tr>
<td></td>
<td>Japan</td>
<td>29.8</td>
<td>35.8</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>United States</td>
<td>23.5</td>
<td>26.4</td>
<td>2.9</td>
</tr>
</tbody>
</table>

*Labor supply is measured in hours worked per person aged 15–64 per week.
Sources: See Appendix.

Reconciling Micro and Macro Estimates

Recent interest in reconciling micro and macro elasticity estimates (see Chetty-Guren-Manoli-Weber ’11)

Three potential explanations

a) Statistical Bias: culture differs in countries with higher tax rates [Alesina, Glaeser, Sacerdote 2005, Steinhauer 2013 for Swiss communities by language]

b) Macro-elasticity captures long-term response which could be larger than short-term response due to frictions (Chetty ’12).

c) Other programs: retirement, education affect labor supply at beginning and end of working life (Blundell-Bozio-Laroque ’11)
Strong evidence that variation in aggregate hours of work across countries happens among the young and the old: (a) schooling–work margin (b) presence of young children (for women), (c) early retirement

Serious cross-country time series analysis would require to put together a better tax wedge by age groups which includes all those additional govt programs [welfare, retirement, child care]

This has been done quite successfully in the case of retirement by series of books by Gruber and Wise, Retirement around the world

⇒ Need to develop a more comprehensive international / time series database of tax wedges by age and family types
There are certain key margins where tax rates impinge on earnings decisions. For many male workers this is at the beginning and at the end of their working lives. These are the schooling-work margins and the early retirement margins. Indeed much of the difference in male employment across OECD countries occurs at these points in the life-cycle.

Male employment by age – US, FR and UK 2005

Source Blundell (2009), Mirrlees Review
Male Hours by age – US, FR and UK 2005

Source Blundell (2009), Mirrlees Review
Male employment by age – US, FR and UK 1975

Source: Blundell (2009), Mirrlees Review
Female Employment by age – US, FR and UK 2005

[Graph showing female employment by age for US, FR, and UK from 16 to 78 years old.]

Source: Blundell (2009), Mirrlees Review
Female Hours by age – US, FR and UK 2005

Source: Blundell (2009), Mirrlees Review
For women earnings are influenced by taxes and benefits not only at these margins but also when there are young children in the family.

For women with younger children it is not usually just an employment decision that is important it is also whether to work part-time or full-time.

Often the employment margin is referred to as the extensive margin of work and the part-time or hours of work decisions more generally as the intensive margin.

Female Employment by age – US, FR and UK 1975

Source Blundell (2009), Mirrlees Review
Long-term effects: Evidence from the Israeli Kibbutz

Abramitzky '15 book based on series of academic papers

Kibbutz are egalitarian and socialist communities in Israel, thrived for almost a century within a more capitalist society

1) Social sanctions on shirkers effective in small communities with limited privacy

2) Deal with brain drain exit using communal property as a bond

3) Deal with adverse selection in entry with screening and trial period

4) Perfect sharing in Kibbutz has negative effects on high school students performance but effect is small in magnitude (concentrated among kids with low education parents)
Long-term effects: Evidence from the Israeli Kibbutz

Abramitzky-Lavy ECMA’14 show that high school students study harder once their kibbutz shifts away from equal sharing.

Uses a DD strategy: pre-post reform and comparing reform Kibbutz to non-reform Kibbutz. Finds that

1) Students are 3% points more likely to graduate

2) Students are 6% points more likely to achieve a matriculation certificate that meets university entrance requirements

3) Students get an average of 3.6 more points in their exams

Effect is driven by students whose parents have low schooling; larger for males; stronger in kibbutz that reformed to greater degree.
Culture of Welfare across Generations

Conservative concern that welfare promotes a culture of dependency: kids growing up in welfare supported families are more likely to use welfare

Correlation in welfare use across generations is obviously not necessarily causal

Dahl, Kostol, Mogstad QJE’2014 analyze causal effect of parental use of Disability Insurance (DI) on children use (as adults) of DI in Norway

Identification uses random assignment of judges to denied DI applicants who appeal [some judges are severe, some lenient]

Find evidence of causality: parents on DI increases odds of kids on DI over next 5 years by 6 percentage points

Mechanism seems to be learning about DI availability rather than reduced stigma from using DI [because no effect on other welfare programs use]
Judge allowance rate in the other cases a judge has handled. We note the judge leniency measure is calculated from all cases the judge has ever handled, not just the cases in our estimation sample. On average, each judge has handled a total of 380 cases. The mean of the leniency variable is .15 with a standard deviation of .06. The histogram reveals a wide spread in judge leniency, with approximately 22% of cases allowed by a judge at the 90th percentile compared to approximately 9% at the 10th percentile.

Figure 3: Effect of Judge Leniency on Parents (First Stage) and Children (Reduced Form).

Notes: Baseline sample, consisting of parents who appeal an initially denied DI claim during the period 1989-2005 (see Section 3 for further details). There are 14,893 individual observations and 79 different judges. Panel (A): Solid line is a local linear regression of parental DI allowance on judge leniency. Panel (B): Solid line is a local linear regression of child DI receipt on their parent’s judge leniency measure. All regressions include fully interacted year and department dummies. The histogram of judge leniency is shown in the background of both figures (top and bottom 0.5% excluded from the graph).

Source: Dahl, Kostol, Mogstad (2013)
REFERENCES


10-069 call for papers on "Future Research in the Social, Behavioral, and Economic Sciences" September 2010. (web)


