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 (same as previous lecture):

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

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

Individual solves

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

FOC: $w \partial u / \partial c + \partial u / \partial l = 0$ defines uncompensated (Marshallian) labor supply function $l^u(w, R)$

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

Income effect parameter: $\eta = w \partial l / \partial R \leq 0$: $\$ 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: $\varepsilon^c = (w/l) \partial l^c / \partial w > 0$

Slutsky equation: $\partial l / \partial w = \partial l^c / \partial w + l \partial l / \partial R \Rightarrow \varepsilon^u = \varepsilon^c + \eta$
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 (Ordinary Least Square) regression:

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

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

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

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

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

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

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)
ISSUE WITH OLS REGRESSION:

$w_i$ correlated with tastes for work $\epsilon_i$

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

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 regression

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 $\Rightarrow$ Tax changes provide more compelling identification
Negative Income Tax (NIT) Experiments

1) Best way to resolve identification problems: exogenously increase the tax rate / non-labor income 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>
<thead>
<tr>
<th>Program Number</th>
<th>$G$ ($)</th>
<th>$\tau$</th>
<th>Declining Tax Rate</th>
<th>Break-even Income ($)</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
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) Response of married women is concentrated along the extensive margin
5) Earnings of treated married 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: \( \eta = w \partial l / \partial R \) 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

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

Source: Imbens et al. (2001), p. 783
Responses to Low-Income Transfer Programs

1) Particular interest in treatment of low incomes in a progressive tax/transfer 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
Overall Costs of Anti Poverty Programs

1) US government (fed+state and local) spent $1000bn in 2016 on income-tested programs

   a) About 5% of GDP but 15% of $6 Trillion govt budget (fed+state+local).

   b) About 50% is health care (Medicaid)

2) Only $250 billion in cash (1.3% of GDP, or 25% of transfer spending)
1996 US Welfare Reform

1) Largest change in welfare policy

2) Reform modified AFDC 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

4) EITC also expanded during this period: general shift from welfare to “workfare”
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
Figure TANF 1.
AFDC/TANF Families Receiving Income Assistance
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

Earned Income Tax Credit (EITC) program

1) EITC started small in the 1970s but was expanded in 1986-88, 1994-96, 2008-09: today, largest means-tested cash transfer program [$70bn in 2016, 30m families recipients]

2) Eligibility: families with kids and low earnings.

3) Refundable Tax credit: administered through income tax 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)
EITC Amount as a Function of Earnings

Earnings ($)
0 5000 10000 15000 20000 25000 30000 35000 40000
Subsidy: 34%
Phase-out tax: 16%
Single, 2+ kids
Married, 2+ kids
Single, 1 kid
Married, 1 kid
No kids
EITC Amount ($)
0 1000 2000 3000 4000 5000
Source: Federal Govt
Figure 1: Earned Income Tax Credit by Number of Children and Filing Status, 2013

Theoretical Behavioral Responses to the EITC

**Extensive margin:** positive effect on Labor Force Participation. Meyer-Rosenbaum (2001) find that 60% of LFP increase of single mothers in 1990s due to EITC expansion.

**Intensive margin:** earnings conditional on working;

1) Phase in: (a) Substitution effect: work more due to 40% increase in net wage, (b) Income effect: work less ⇒ Net effect: ambiguous; probably work more

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

3) Phase out: (a) Substitution effect: work less, (b) Income effect: also work less ⇒ Net effect: work less
EITC and Intensive Labor Supply Response: Bunching at Kinks

1) Basic labor supply theory predicts that we should observe bunching of individuals at the EITC kink points:

Some individuals find it worthwhile to work more when subsidy rate is 40% but not when subsidy rate falls to 0% ⇒ Utility maximizing labor supply is to be exactly at the kink

2) Amount of bunching is proportional to compensated elasticity: if labor supply is inelastic, then kinks in the budget set are irrelevant and do not create bunching

Saez AEJ’10 finds bunching around 1st kink point of EITC but only for the self-employed ⇒ likely due to cheating to maximize tax refund (and not labor supply)
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

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

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
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...
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

Panel A. One child

Source: Saez (2010), p. 191

Panel B. Two children or more

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 zero self-employment income (and hence whose
EIC amount ($)

Panel A. One child

Panel B. Two or more children

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 tax filers with one dependent child and panel B for tax filers 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
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 earnings density).

Panel A. One child

Panel B. Two or more children

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 tax filers with one dependent child and panel B for tax filers 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
EITC Empirical Studies

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 increases or decreases with earnings

Such confusion might be good for the government as the EITC induces work along participation margin without discouraging work along intensive margin
Chetty, Friedman, Saez AER’13 EITC information

Use US population wide tax return data since 1996

1) Substantial heterogeneity fraction of EITC recipients bunching (using self-employment) across geographical areas ⇒ Information about EITC varies across areas

2) Places with high self-employment EITC bunching display wage earnings distribution more concentrated around plateau ⇒ Evidence of wage earnings response to EITC along intensive margin

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

Percent of Wage Earners

W-2 Wage Earnings

EITC Amount ($)

$0 $10K $20K $30K $25K $35K $25K $5K

0% 0.5% 1% 1.5% 2% 2.5% 3% 3.5%

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

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

Source: Chetty, Friedman, and Saez NBER'12
Welfare reform and consumption: Meyer and Sullivan 2004

1) Examine the consumption patterns of single mothers and their families from 1984–2000 using CEX data

2) Question: did single mothers’ consumption fall because they lost welfare benefits and were forced to work?
Fig. 2. Total consumption: single mothers, 1984–2000.

Meyer and Sullivan: Results

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

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 recession?

Households getting SNAP (food stamps) surged from 12M in ’07 to 20M in ’10

But households getting TANF increased only slightly from 1.7M in ’07 to 1.85M in ’10
Long-term effects of Redistribution: Evidence from the Israeli Kibbutz

Abramitzky ’13 book based on series of academic papers

Kibbutz are egalitarian and socialist voluntary communities in Israel, thrived for almost a century within a 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
Long-term effects of Redistribution: Evidence from the Israeli Kibbutz

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

They use a DD strategy: pre-post reform and comparing reform Kibbutz to non-reform Kibbutz. They find that

1) Students are 3 percentage points more likely to graduate

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

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’14 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 severe, others 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]
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


Abramitzky, Ran and Victor Lavy, 2014 “How Responsive is Investment in Schooling to Changes in Redistributive Policies and in Returns?”, Econometrica, 82(4), 1241-1272 (web)


