Lecture: Taxes and Labor Supply

Hilary Hoynes
hwhoynes@ucdavis.edu
EC230

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

OUTLINE – in this lecture

1) Basic Labor Supply Model, adding taxes

2) Labor Supply Elasticity Estimation: Methodological Issues

3) Estimates of hours/participation elasticities

Emmanuel Saez (2009) “Do Taxpayers Bunch at Kink Points?” American Economic Journal: Economic Policy.

Raj Chetty, R., J. Friedman, T. Olsen and L. Pistaferri “Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records”, Quarterly Journal of Economics 126(2): 749-804, 2011

N. Eissa “Taxation and Labor Supply of Married Women: The Tax Reform Act of 1986 as a Natural Experiment” NBER Working Paper 5023, 1995.

O. Ashenfelter and M. Plant. (1990) "Non-Parametric Estimates of the Labor Supply Effects of Negative Income Tax Programs," Journal of Labor Economics, 8.1 (January), S396-S415.

G. Imbens, D. Rubin, and B. Sacerdote (2001) "The Causal Effect of Income on Labor Supply: Evidence From the Lottery Winner Survey." American Economic Review, Vol. 91, No. 4, September 2001.

MORE ON TAXES AND LABOR SUPPLY IN LATER LECTURES:

Elasticity of Taxable Income

- more general characterization of labor supply: tax avoidance, old Feldstein point that there are other margins then hours of work

Responses to low-income programs (EITC, welfare)

REFERENCES

Surveys in labor economics

- Pencavel (1986), Handbook of Labor Economics, vol 1
- Heckman and Killingsworth (1986) Handbook of Labor Econ vol 1
- Blundell and MaCurdy (1999), Handbook of Labor Economics vol 3
- Keane JEL'2011 (structural)

Surveys in public economics

- Hausman (1985) Handbook of Public Economics vol 1
- Moffitt (2003) Handbook of Public Economics vol 4
- Saez, Slemrod, and Giertz JEL (2012) (reduced form)

STATIC LABOR SUPPLY WITH LINEAR BUDGET CONSTRAINT

Individual faces exogenous deterministic wage (w) and non-labor income (N). Utility is a function of leisure (ℓ) and consumption (c). The choice problem is:

Maximize $U(\ell, y)$ subject to $wh + N = y$

$h = T - \ell$

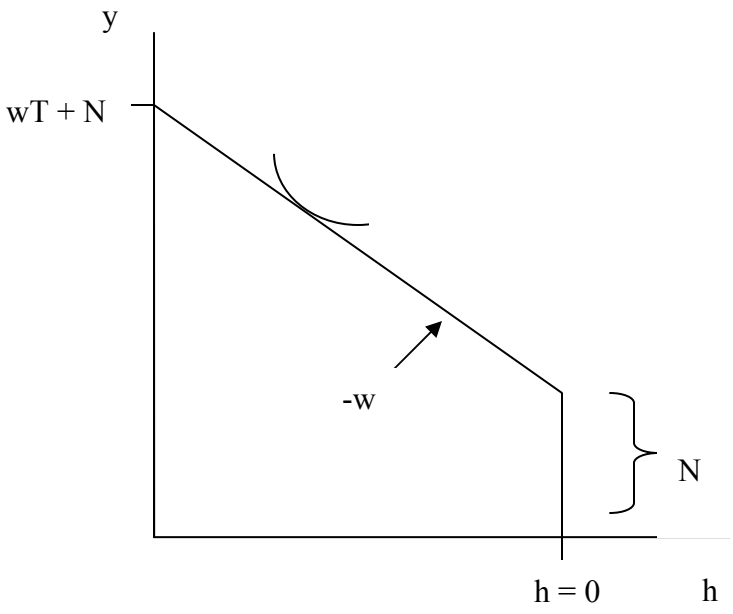
Where:

w	=	hourly wage
h	=	hours worked (ℓ=leisure)
T	=	time endowment
N	=	non-labor income

(Usual) Assumptions:

Increasing in *l* and y (decreasing in h)

Leisure and consumption are normal goods



Deriving budget constraint:

$y = wh + N\,, \quad h = T - \ell$

$y = w(T - \ell) + N$

$y = (wT + N) - (w\ell)$

intercept= wT + N (full income at full hours)
slope = -w (loss in income of one more hour of leisure)

Therefore indifference curves have usual shape. We are typically interested in studying the determinants of <u>hours worked</u> but we model the determinants of leisure and then translate back to hours.

Intensive Margin: FOC

$$wU_y + U_h = 0$$
$$w = -\frac{U_h}{U_y}$$

Yields the optional labor supply function $h = h^*(w, N)$

Extensive Margin

Define w^* = reservation wage

$= -\frac{U_h}{U_y}$, evaluated at h=0

$w > w^*$ then work [h > 0]
$w < w^*$ then no work [h = 0], equivalent to h*<0

Comparative Statics:

Uncompensated elasticity of labor supply $\varepsilon^u = \frac{w}{h}\frac{\partial h}{\partial w}$

substitution effect <0
income effect >0 (if leisure normal)
Can be positive or negative (backward bending labor supply)

Income effect parameter $\eta = w\frac{\partial h}{\partial N}$

If leisure is a normal good, then negative
(Imbens, Rubin, Sacerdote AER 2001)

Compensated elasticity of labor supply $\varepsilon^c = \frac{w}{h}\frac{\partial h^c}{\partial w}$

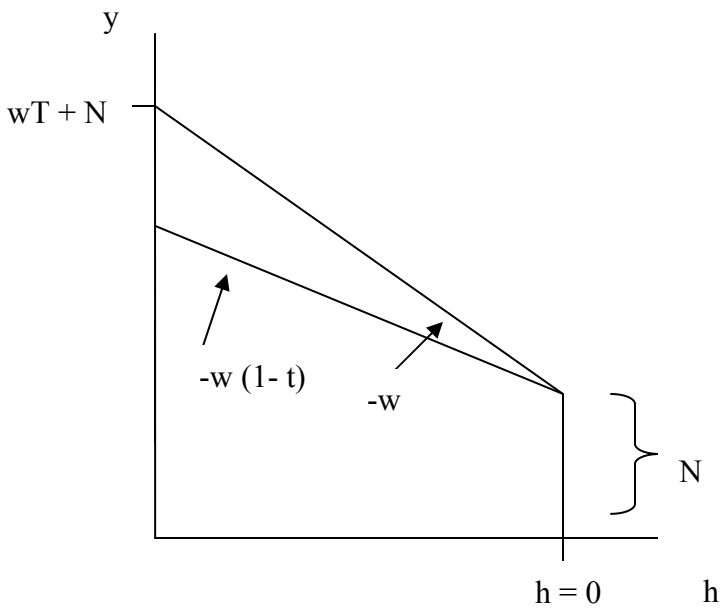
Always positive

Elasticity of participation: $\frac{\partial \Pr(h > 0)}{\partial w} \frac{w}{P} = \frac{\partial \Pr(w > w^*)}{\partial w} \frac{w}{P}$

Positive, increase in wages leads to an increase in participation
(no `income effect' when considering extensive margin)

Adding taxes to labor supply model

Example #1: a uniform, proportional tax denoted as t



Max U (ℓ , y) s.t.

$$w(1-t)h + N = y, \quad h = T - \ell$$

FOC:

$$-\frac{\frac{\partial U}{\partial h}}{\frac{\partial U}{\partial y}} = w(1-t)$$

LS function: $h = h^*(w(1-t),N)$

Observations:

1. net of tax wages belong in the labor supply equation (not gross wage)
2. Policy question: How do taxes affect hours worked?
   Elasticity can tell us; theory does not even tell us the sign!
3. How do taxes affect labor force participation?
   taxes → reduction in net of tax wages; no change in reservation wage
   → probability of work decreases

BASIC CROSS SECTION ESTIMATION

Data on hours or work, wage rates, non-labor income started becoming available in the 1960s:
Current Population Survey (annual, starting in 1960s)
Panel Study of Income Dynamics (panel, 1968-)

Simple OLS regression:

$$h_i = \alpha + \beta w_i + \gamma N_i + \varphi X_i + \varepsilon_i$$

w_i = net-of-tax wage
N_i = non-labor income [including spousal earnings for couples]
X_i = demographic controls [age, experience, education, etc.]
β = measures uncompensated wage effects
φ = income effects

ELASTICITY ESTIMATES FROM BASIC CROSS SECTION

1. Male workers [primary earners when married] (Pencavel,1986 survey):

Small effects
$\varepsilon^u = 0,\ \eta = -0.1,\ \varepsilon^c = 0.1$

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]

1. WAGE CORRELATED WITH TASTES FOR WORK

Cross sectional identification of w, high wage guys have more taste for work independent of wage? Leads upward bias.

Adding taxes (net of tax wage) could lead to downward bias under progressive MTR (high ability means lower net of tax wage and more work)

Controlling for X can help but can never be sure that we have controlled for all the factors correlated with w and tastes for work:

<u>Omitted variable bias</u> →Tax changes provide more compelling identification.

2. MEASUREMENT ERROR IN HOURS

In general w computed as earnings / hours → can create division bias

Let l^* denote true hours, l observed hours (e observed=true)

Compute $w = e / l$ where e is earnings

$$\log l = \log l^* + \mu$$

$$\log w = \log e - \log l = \log e - (\log l^* + \mu) = \log w^* - \mu$$

Spurious negative correlation between observed hours and observed wages. Workers with high misreported hours also have low imputed wages biasing elasticity estimate downward

Solution → tax instruments again

3. NONPARTICIPATION

As we saw above, it is utility maximizing for some individuals not to work.

Practically, wages are unobserved for non-labor force participants

Thus, OLS regression (typically) conditions on <u>workers only</u>

This can bias OLS estimates: low wage earners must have very high unobserved propensity to work to find it worthwhile.

Requires a selection correction pioneered by Heckman in 1970s (e.g. Heckit, Tobit, or ML estimation). Problem is that identification is based on strong functional form assumptions.

Current approach: use panel data to distinguish entry/exit from intensive-margin changes.

4. EXTENSIVE VS INTENSIVE MARGIN

Related issue: want to understand effect of taxes on labor force participation decision.

Interestingly, a small change in net wages could lead individuals may jump from non-participation to part time or full time work (non-convex budget set).

This can be handled using a discrete choice model:

$$P = \phi(w(1-t), N, X)$$

where P is an indicator for whether the individual works [0 or 1] and ϕ is typically specified as logit, probit, or linear prob model.

Again, need tax variation.

5. NON-HOURS RESPONSES

Traditional literature focused purely on hours of work and (later) labor force participation

Problem: income taxes distort many margins beyond hours of work

a) Non-hours margins may be more important quantitatively

b) Hours very hard to measure (most people report 40 hours per week)

Two solutions in modern literature:

a) Focus on total earnings [or taxable income] as a broader measure of labor supply

b) Focus on subgroups of workers for whom hours are better measured, e.g. taxi drivers

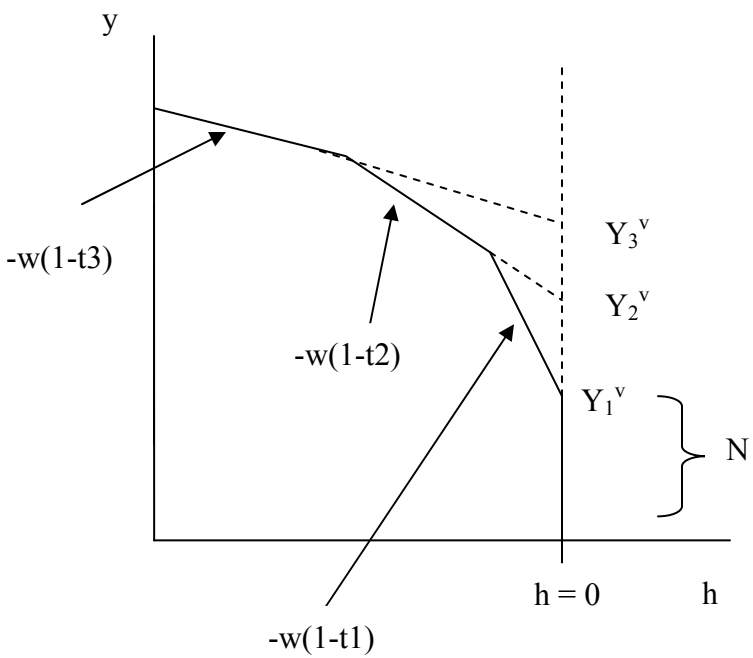
6. NON-LINEAR BUDGET SETS

Tax system is not linear, but instead piecewise linear with varying MTR. This arises from many features in the tax and transfer system: progressive MTR, means-tested transfers, ceiling in payroll tax, social security earnings test, etc.

Same theory applies when considering the linearized tax system.

Consider three, increasing, marginal tax rates. Budget constraint becomes:

$Y = wh + N - T(wh,N)$ where $T(.)$ is the tax function.



$t = t1$ if $E < E1$
$t = t2$ if $E1 < E < E2$
$t = t3$ if $E > E2$

$(t3 > t2 > t1)$

Y is "virtual income" measure

Consider someone on the highest segment

First Order Condition:

$$-\frac{\partial U / \partial h}{\partial U / \partial y} = w(1 - t_3)$$

Which implies the labor supply function:

$$h = h^* (w(1 - t3) , Y^v_3)$$

Or, more generally:

$$h = h^* (w_n , Y^v)$$

where w_n = net of tax wage

$Y^{v`}$ = virtual income

Determining participation: $h^* < 0$ then no work

$h^*(w(1-t1),N)<0$

Given that utility function is concave, and budget set is convex, then we know there is a unique tangency (or corner solution) on one of the segments.

Possibilities: not working ($h^* < 0$)
tangency on 1st, 2nd or 3rd segment
on a kink (expect people to be bunched on convex kinks)

In cross-sectional (reduced form) setting, the main complications are that:

a) w and y are endogenous to choice of hours (tastes for work correlated with tax rate → downward bias in estimated wage elasticity)

b) FOC may not hold if bunched at the kink (mis-specification)

Some help can come with focusing on tax-reform induced changes in the tax rate.

But fundamentally dealing with this requires a different approach.

NON-LINEAR BUDGET SET METHOD

Pioneered by Hausman in late 1970s (Hausman, 1985 PE handbook chapter)
Method uses a structural model of labor supply

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

- Specify preferences, specify errors, etc
- Construct likelihood function given observed labor supply choices
- Find parameters that maximize likelihood

Important insight: need to use virtual incomes to turn problem into standard linear one.

[See example at end of lecture notes]

New literature identifying labor supply elasticities using tax changes has a totally different perspective: taxes are seen as an opportunity to identify labor supply.

RESULTS FROM NONLINEAR BUDGET SETS

- This approach has generally found larger elasticities than earlier literature [Hausman (1981)] Subsequent studies obtain different estimates (MaCurdy, Green, and Paarsh 1990, Blomquist 1995).
- Several studies find negative compensated wage elasticity estimates
- Debate: impose requirement that compensated elasticity is positive or conclude that data rejects model?

Shortcomings of this approach

- Sensitivity to functional form choices, which is a larger issue with structural estimation
- No tax reforms, so does not solve fundamental econometric problem that tastes for work may be correlated with w
- More fundamentally, labor supply model predicts that individuals should bunch at the kink points of the tax schedule. <u>But we observe very little bunching at kinks, so model is rejected by the data</u>

Interest in this approach diminished despite their conceptual advantages over OLS methods.

SAEZ "DO TAXPAYERS BUNCH AT KINK POINTS?" AEJ Policy 2010

-- Basic prediction of kinked budget constraint model is that we should see people bunched at the convex kinks. (And we should see a gap in the distribution at nonconvex kinks.)
-- Some papers have examined particular applications (social security earnings test, welfare recipients around notch, WFTC and hours restriction) but no study has examined this among taxpayers in US.
-- Simple, clever paper using the best data (tax data)

<u>Modeling insights</u>

1. Less curvature in indifference curves (higher substitution elasticity) → more bunching
   $dz^*/z^* = e\,[dt/1-t]$ e=sub elasticity, t=MTR, z=taxable income
2. Therefore if there is little evidence of bunching (and model is valid) → small elasticity of taxable income
3. Later he considers changes to model to explain lack of bunching (uncertainty in income, constrained hours choice)

Data

IRS annual cross-section of taxfilers (1960-2004) N=80,000-200,000/year

He does not use all of the years (high inflation years when tax parameters were not indexed)

Methods:

1. Simple descriptive unconditional exercise
   - Uses histograms and kernel density (local smoother of histogram; within a "band" observations further from the central point are weighted less in average)
2. Uses formulas for relationship between clustering at kinks and elasticities to empirically estimate elasticity.

Results: EITC: Fig 3-5

- Uses data from 1995-2004. Why? Income tax schedule for EITC is stable (in real terms) during this time period.
- Presents figures by number of children (since schedule varies along that dimension)
- Some evidence of bunching around EITC first kink. Results concentrated for those with self employment income (no effect for those with only wage income)

FIG 3B

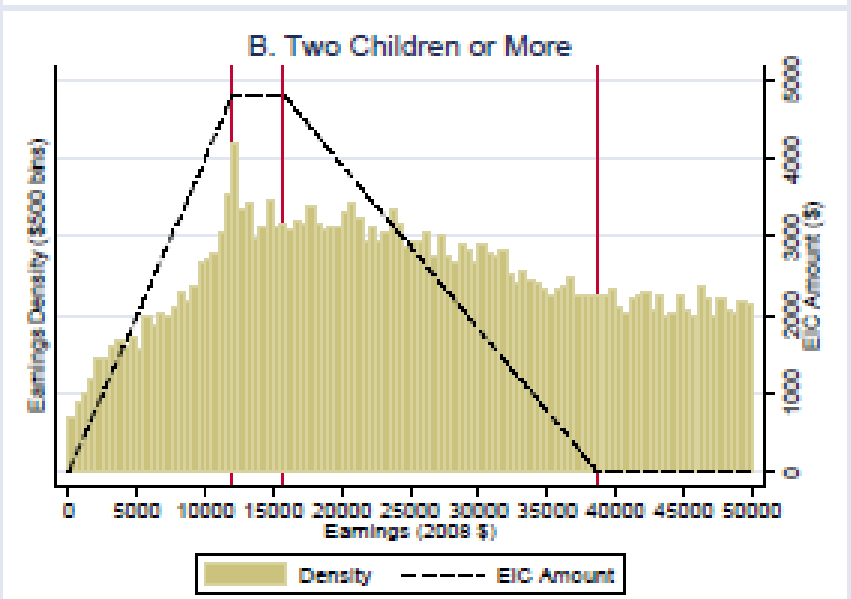
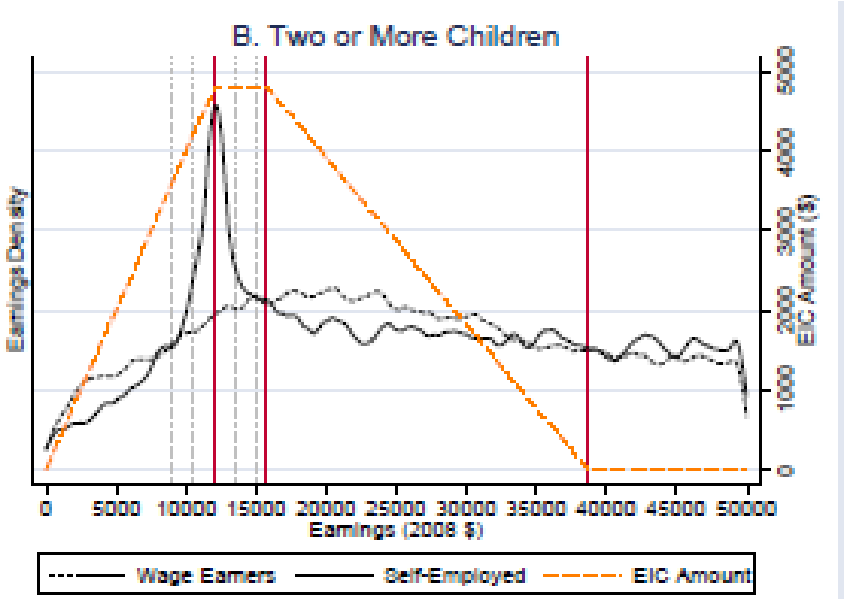


FIG 4B



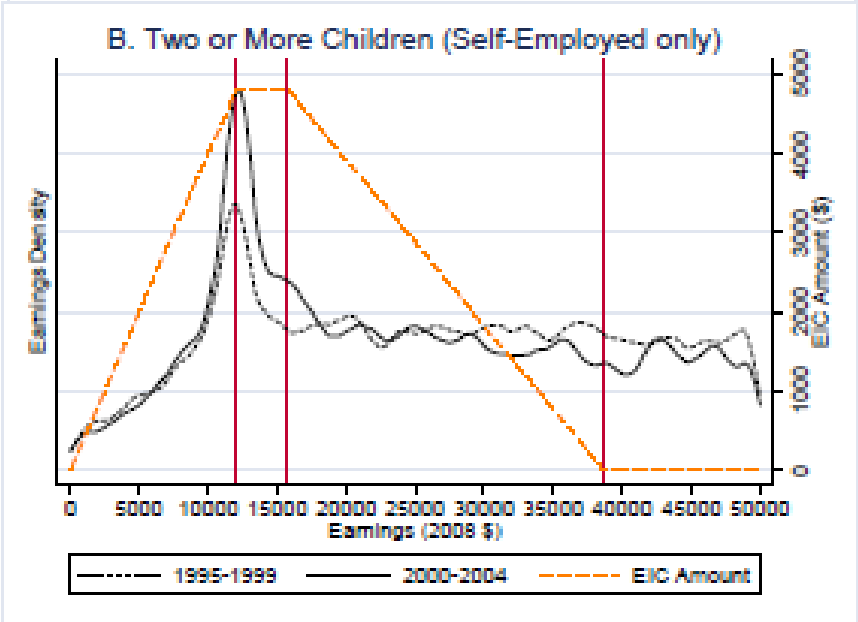


FIG 5B

- Federal Income Tax:
  - More complicated to show since the schedule varies across family types (marital status), number of children, and deductions. Normalize rel to 0.
  - Some evidence of bunching around 1st kink (MTR goes from 0 → 15%)
  - Figures 6/7
  - More evidence for single and HH returns
  - First kink probably the most "visible" to taxpayer. But could the finding be an artifact that those left of 1st kink do not have to file and may not be in data?
  - No evidence at 2nd or later kinks

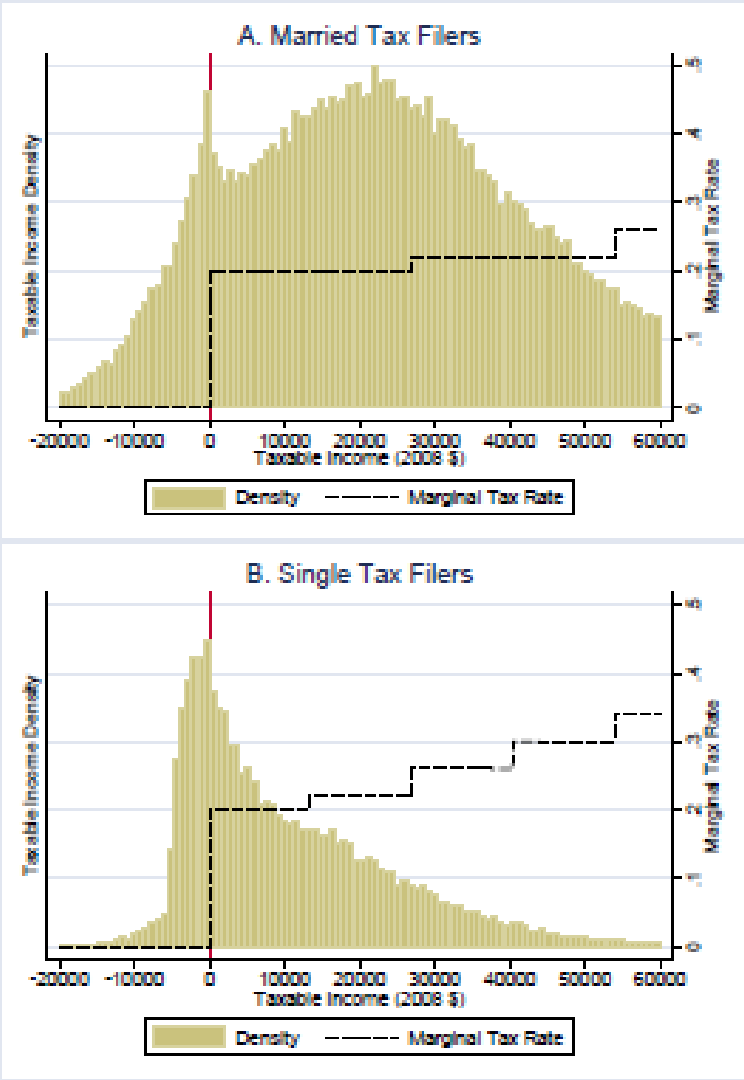


Figure 6. Taxable Income Density, 1960-1969: Bunching around First Kink

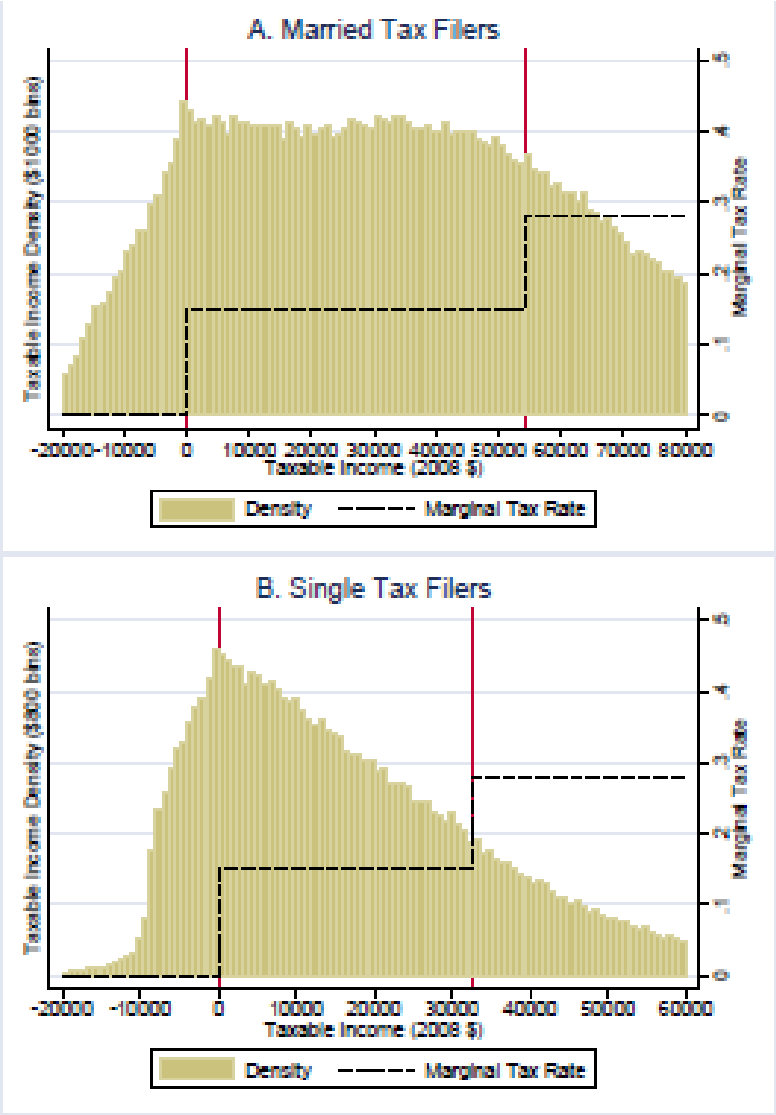


Figure 9. Taxable Income Density, 1988-2002

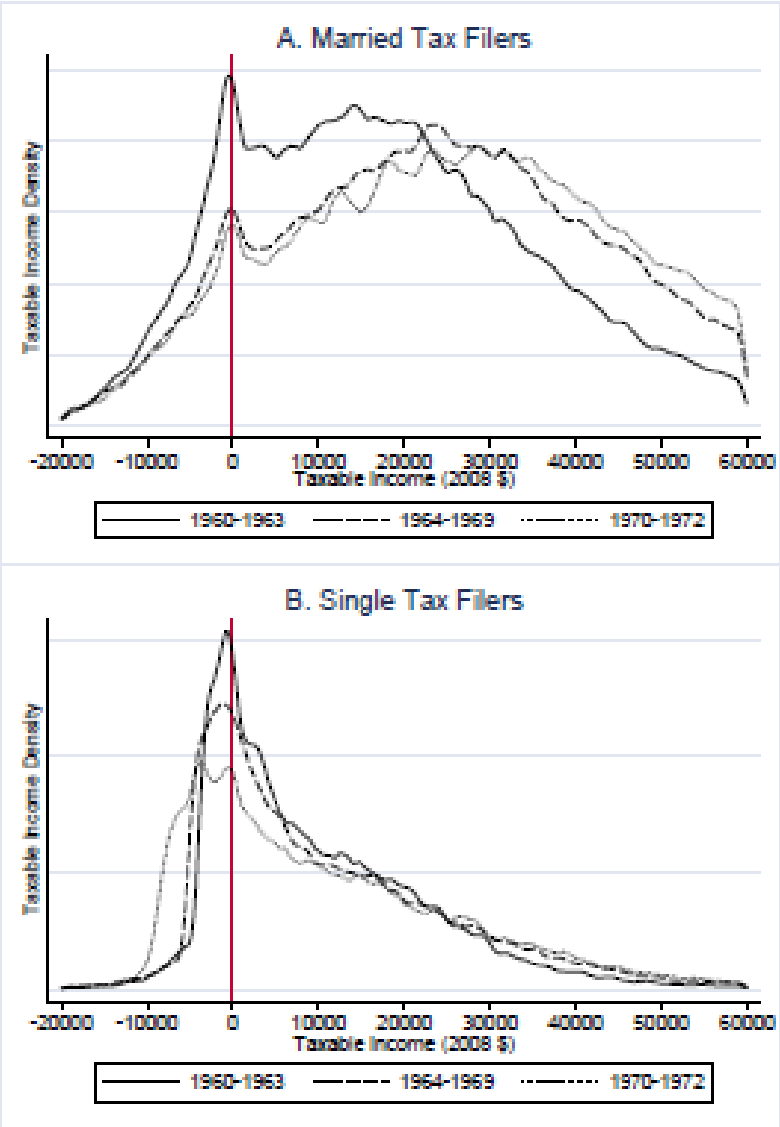


Figure 8. Taxable Income Density, 1960-1972: Dynamics of Bunching

The figure displays the kernel density of taxable income (in 2008 dollars) for married joint tax filers (Panel A) and single tax filers (Panel B) for 3 periods 1960-1963 (solid line), 1964-1969 (dashed line), and 1970-1972 (dotted line). Densities are computed using population weights. As shown in Table 3, the definition of taxable income was stable (in nominal terms) from 1948 to 1963 and was changed in 1964 and in 1970. The graphs show that bunching decreased sharply after the reforms showing that it takes time for tax filers to bunch. Sample sizes for 1960-1963, 1964-1969, 1970-1972 are 76,515, 108,646, 69,365 in Panel A and 34,636, 48,223, 21,775 in Panel B. The bandwidth is $500 in all kernel density estimations.

Implication:

-- Small elasticities for wage earners

-- Simulations using extended model again shows no clustering. So these models are not right or elasticities are small or agents do not know where kinks are.

-- Problematic for research using kinked budget constraint methods

OTHER EVIDENCE ON LACK OF BUNCHING

Friedberg 2000: Social Security Earnings Test

1) Uses CPS data on labor supply of retirees receiving Social Security benefits
2) Studies bunching based on responses to Social Security earnings test
   - Earnings test: phaseout of SS benefits with earnings above an exempt amount around $14K/year
   - Today: Phaseout rate varies by age group: 50% (below 66), 33% (age 66), 0 (above 66)
3) Friedberg exploits a 1983 reform (change in SS earnings test)

Estimates elasticities using Hausman method, finds relatively large compensated and uncompensated elasticities.

Finds some evidence of bunching on kink (but data is not as good as Saez). May be that this is more salient.

Ironically, lost social security benefits are considered delayed retirement with an actuarial adjustment of future benefits → (a) No kink if person has average life expectancy, (b) kink if person has less than average life expectancy

WHY NOT MORE BUNCHING AT THE KINKS?

1) True intensive elasticity of response may be small (may partly be due to inability to adjust hours)

2) Randomness in income generation process: Saez, 2002 shows that year-to-year income variation too small to erase bunching if elasticity is large

3) Information and salience

a) Liebman and Zeckhauser: "Schmeduling" (behavioral model where individuals confuse MTR with average tax rate)

b) Chetty and Saez (2009): information significantly affects bunching in EITC field experiment

4) Adjustment costs and institutional constraints (Chetty et al 2009)

CHETTY, FRIEDMAN, OLSEN, AND PISTAFERRI (2011)

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)

2) Could explain why macro studies find larger elasticities

3) Question: How much are elasticity estimates affected by frictions?

Chetty et al. 2011: Model

1) Firms post jobs at a set hours value (wage-hours offers). Firm can not change hours after matching with worker (hours constraint)

2) Workers draw from this distribution and must pay search cost to reoptimize (so only search if gains exceed costs). Nest “standard” model by setting search costs to 0.

3) Therefore not all workers locate at optimal choice

4) Bunching at kink and observed responses to tax reforms attenuated

Chetty et al. 2011: Testable Predictions

Model generates three predictions:

1) [Size] Larger tax changes generate larger observed elasticities

Large tax changes are more likely to induce workers to search (potential gains more likely to exceed search costs) for a different job

2) [Scope] Tax changes that apply to a larger group of workers generate larger observed elasticities

Firms tailor jobs to preferences of common workers
Individual bunching (individual locate at kink through search) vs aggregate bunching (aggregation of worker's preferences by unions or firms).

3) [Search Costs] Workers with lower search costs exhibit larger elasticities from individual bunching

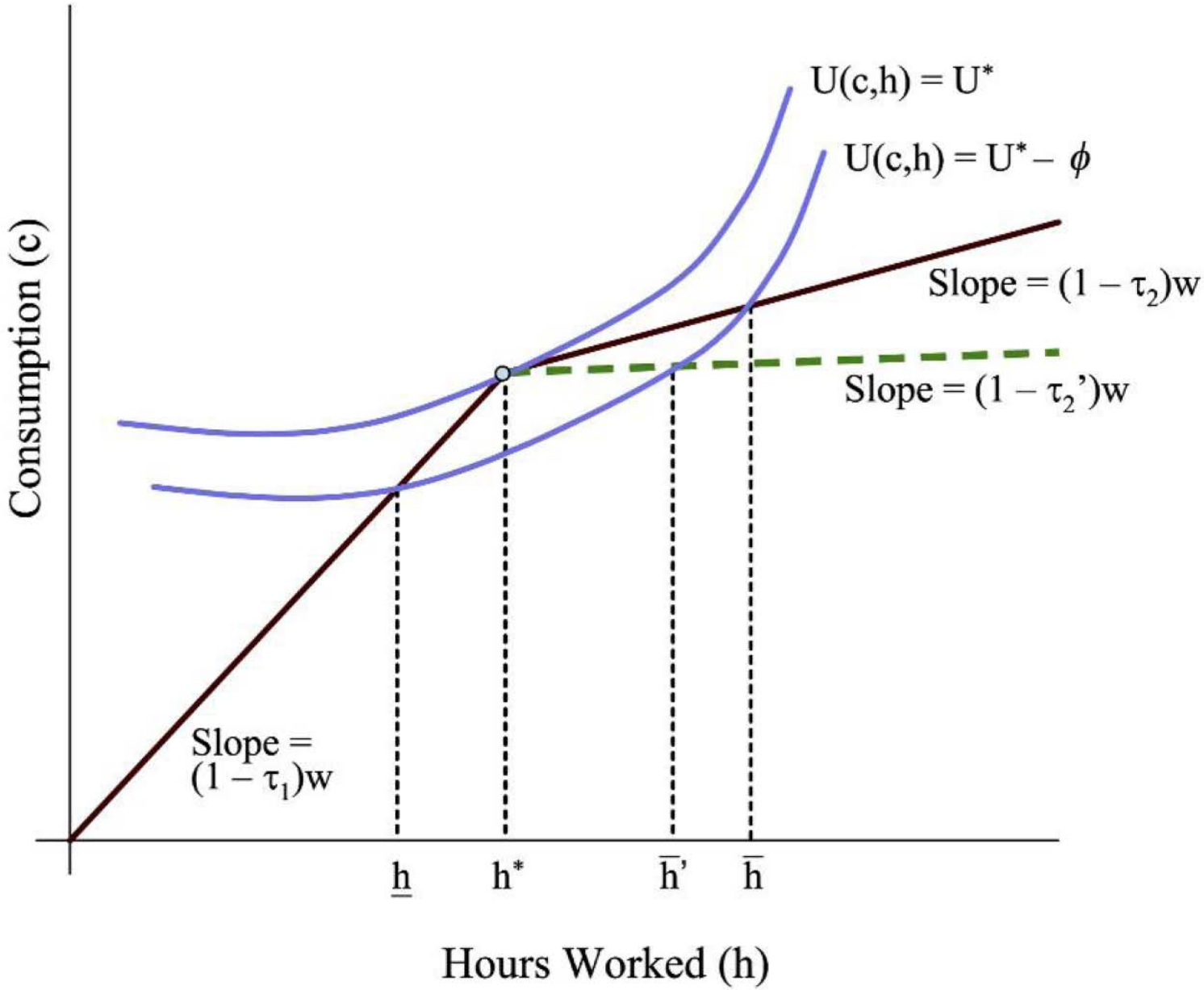


FIGURE I
Bunching at Kinks with Search Costs

Chetty et al. 2011: Data

Matched employer-employee panel data with admin tax records for full population of Denmark 1994-2001

1) Income vars: wage earnings, capital and stock income, pension contributions

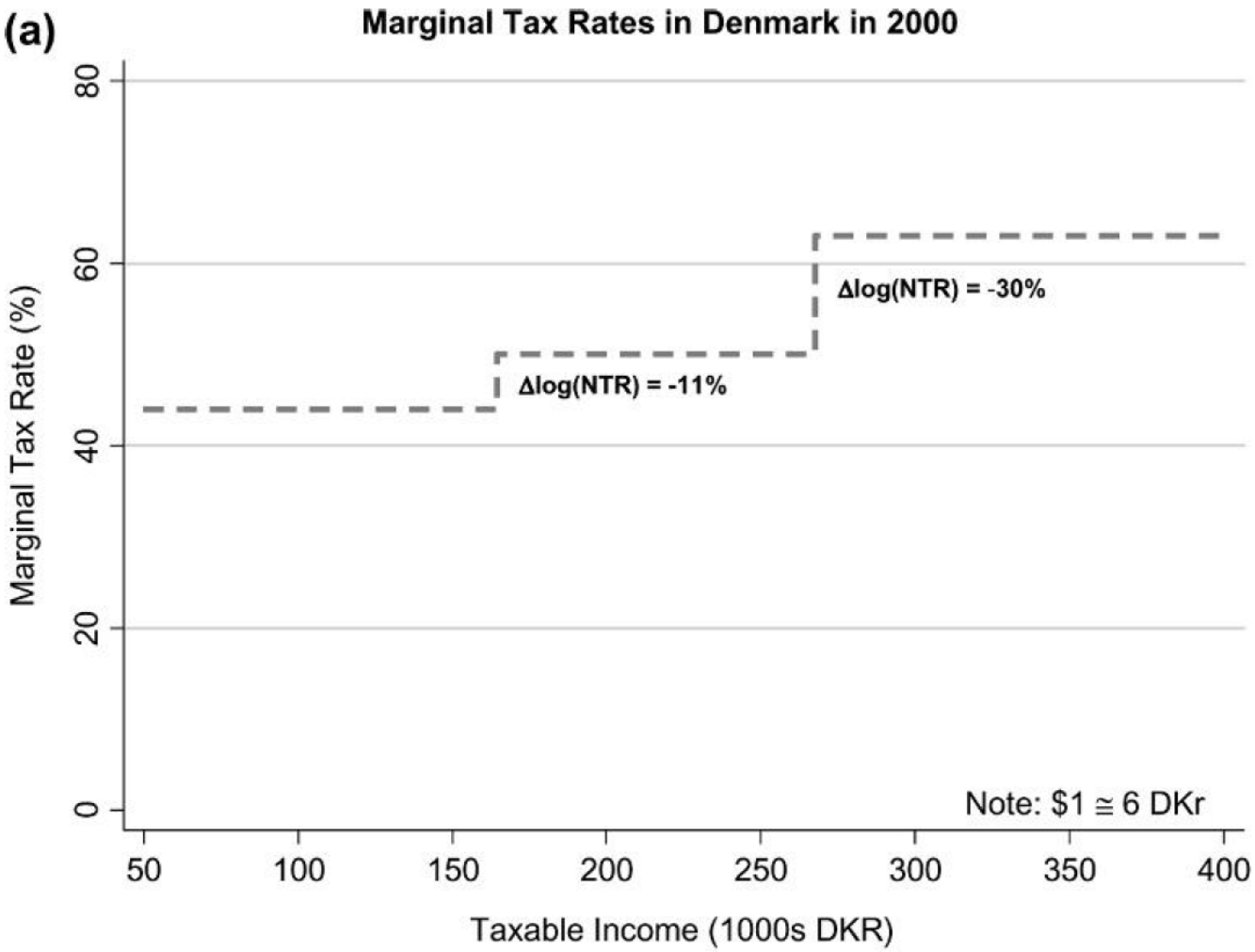
2) Employer vars: tenure, occupation, employer ID

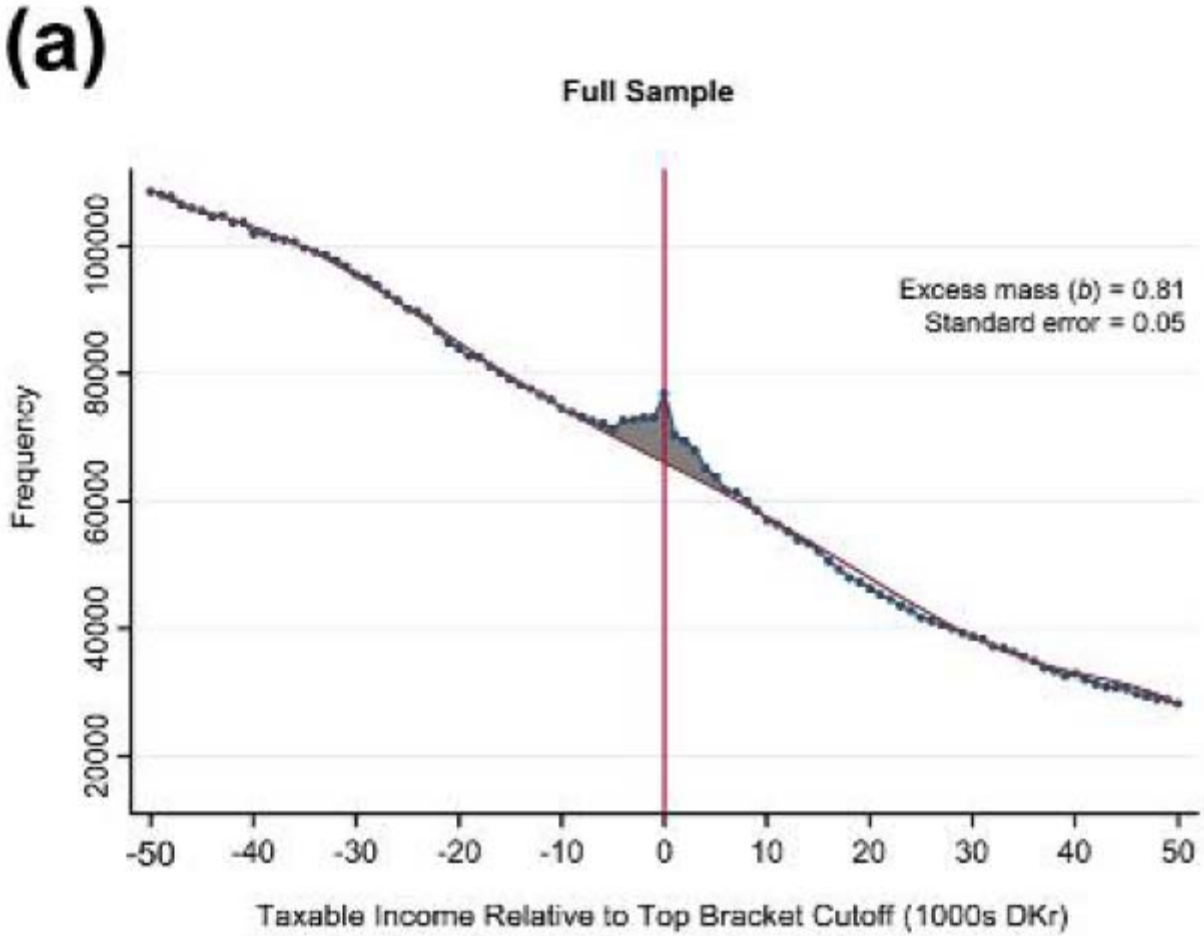
3) Demographics: education, spouse ID, kids, municipality

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

Approximately 2.42 million people per year

Institutional Setting
-- heavily unionized labor force
-- 3 bracket MTR system, 25% pay top MTR
-- tax variation: within tax-year (across brackets) and across year. Larger variation exists within tax-year.





Empirical distribution (1000 bins) of taxable income for those within 50,000 of top bracket.

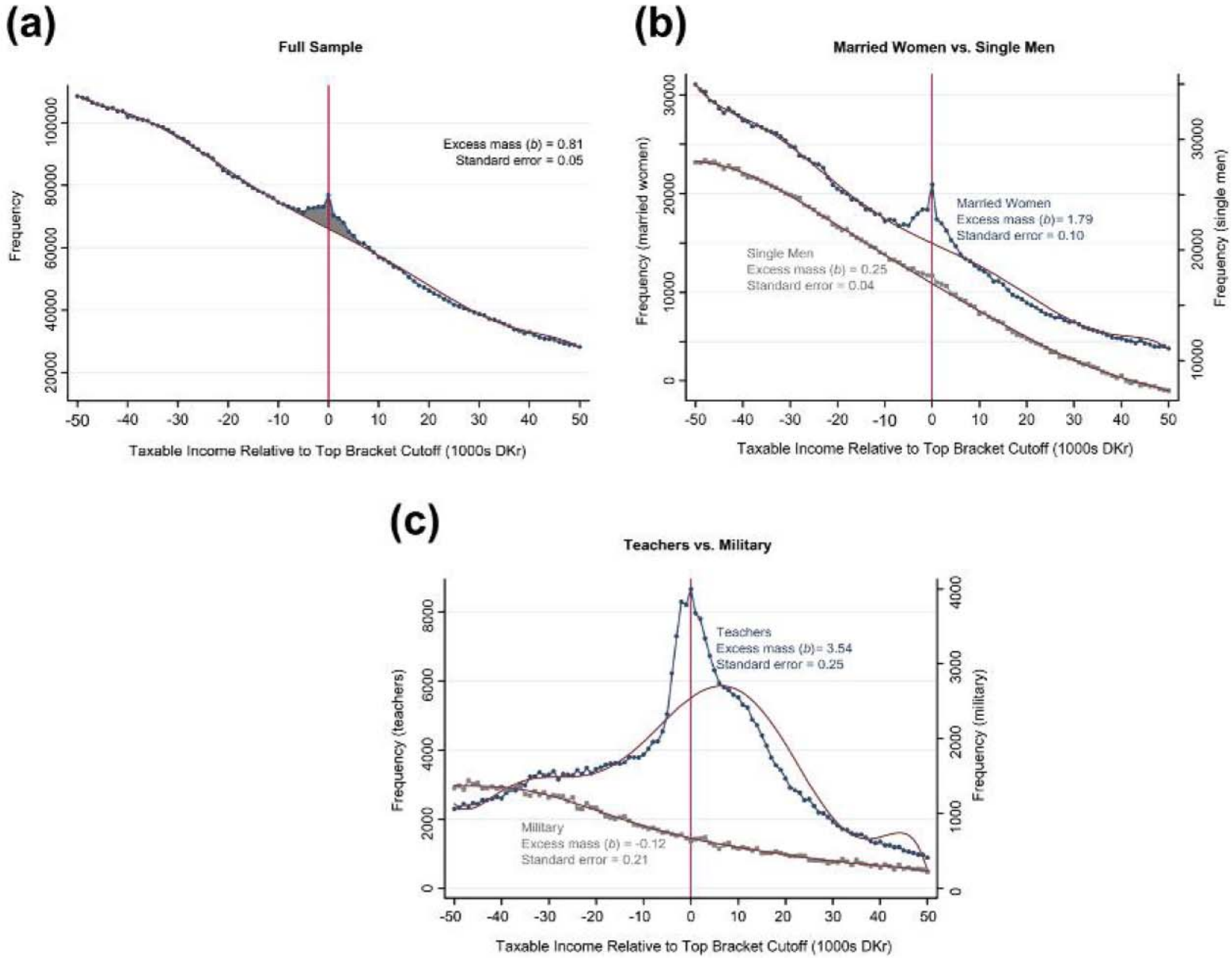


FIGURE III
Income Distributions around the Top Tax Cutoff for Wage Earners

Figure IV: shows that the excess mass moves as the kink increases (automatic indexing of brackets)

Figure VI: Consistent with prediction that larger bunching with tax changes are larger (only present for top bracket)

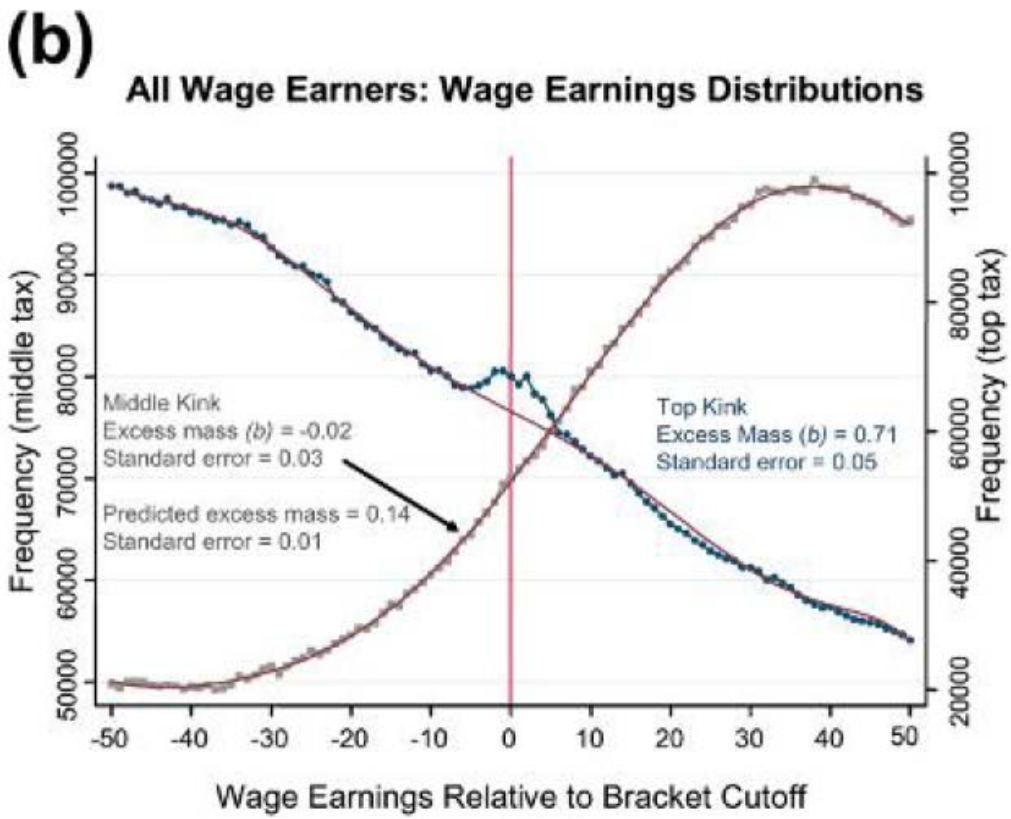


FIGURE VII
Teachers' Wage Earnings Distributions

Shows that bunching holds not using own incentives (e.g. bottom graph) but group's incentives.

Chetty et al. 2011: Results

1) Search costs attenuate observed behavioral responses substantially

2) Firm responses and coordination critical for understanding behavior: individual and group elasticities may differ significantly

3) Nonlinear budget set models may fit data better if these factors are incorporated

4) Standard method of estimating elasticities using "small tax reforms" on same data yields close-to-zero elasticity estimate

5) Placebo test: Much more bunching (and at all kinks) for self employed (who presumably do not face hours constraints or adjustment costs).

6) Illustrates aggregate bunching, that is the bunching for worker's who do not have own incentives to locate at the kink.

Chetty et al 2010 Thoughts

External validity: Denmark less complex labor market, highly unionized. Easier to facilitiate supply-side optimizing.

Cross-sectional tax variation is not always larger than “small tax reforms” (see Eissa TRA86)

REDUCED FORM LABOR SUPPLY LITERATURE: RESULTS

Literature exploits variation in taxes/transfers to estimate hours and participation elasticities

1) Return to simple model where we ignore non-linear budget set issues

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

NEGATIVE INCOME TAX (NIT) EXPERIMENTS

1) Best way to resolve identification problems: exogenously increase the marginal tax rate 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 τ (50%-80%) [based on family earnings]

5) Analysis by Rees (1974), Munnell (1986) book, Ashenfelter and Plant JOLE'90, and others

6) Several groups (varying G, τ) with randomization within each; approx. N = 75 households in each group.

EXPERIMENTAL PRIMER

Why we like them:
Evaluate changes without model, functional form assumptions.
Evaluate new policies (no “natural” variation)

How we evaluate them:
Difference in mean T – C
Why no pre vs post?

What do we need to check:
1) T is exogenous: balance in Xs and pre-RA y between T and C
2) no impact of T on sample over time (attrition)

Table 1
Parameters of the 11 Negative Income Tax Programs

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

NOTE.—Terms are explained in text.

Economics of NIT experiments
-- combination of G, τ lead to reduction in h, lfp
-- decrease in G → increase in h, lfp, B
-- decrease in τ → ambiguous impact on h, B, increase in lfp

NIT Experiments: Ashenfelter and Plant 1990

1) Present non-parametric evidence of labor supply effects
* nonparametric = simple T, C differences
* challenge: nonrandom assignment
* solution: compare T, C within income strata

2) Compare implied benefit payments to treated vs. control households
* could have looked at h, lfp. They instead looked at impacts on benefits (DWL)

* $\hat{\Delta}_j = \sum_i w_{ij}(D_{ij} - D_{ij}^0)$, average over income strata i

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

4) This is the relevant parameter for expenditure calculations and potentially for welfare analysis (revenue method of calculating DWL)

5) Shortcoming: approach does not decompose estimates into income and substitution effects and intensive vs. extensive margin [they focused on “program evaluation”]

6) Hard to identify the key elasticity relevant for policy purposes and predict labor supply effect of other programs

Table 3
Experimental Payment minus Predicted Control Payment for 3-Year Dual-headed Experimental Families, Attrition Families Excluded (Standard Errors in Parentheses)

G ($)	τ	Declining Tax Rate	Preexperimental Payment ($)	Payments for Year of Experiment ($) 1	2	3	Postexperimental Payment ($)
3,800	.5	No	193.78	248.46	368.95*	389.24*	138.56
			(143.45)	(149.58)	(170.75)	(182.99)	(188.20)
3,800	.7	No	124.96	185.18	317.28	218.37	−47.85
			(223.77)	(237.91)	(252.99)	(325.57)	(314.66)
3,800	.7	Yes	−33.37	68.94	158.44	324.84	29.28
			(178.05)	(176.07)	(213.59)	(230.50)	(222.42)
3,800	.8	Yes	75.40	336.06	221.54	160.83	91.52
			(229.44)	(237.18)	(245.92)	(264.53)	(261.84)
4,800	.5	No	52.02	85.17	294.55	337.23	70.22
			(192.31)	(184.85)	(201.73)	(221.73)	(219.58)
4,800	.7	No	220.76	288.33	496.85*	543.25*	178.32
			(160.04)	(169.04)	(197.88)	(204.50)	(194.03)
4,800	.7	Yes	136.99	281.98*	423.30*	348.03*	23.96
			(127.36)	(137.19)	(157.51)	(162.38)	(140.58)
4,800	.8	Yes	−16.87	305.09	417.90	317.39	121.47
			(175.54)	(209.24)	(234.32)	(274.11)	(239.59)
5,600	.5	No	−163.12	200.75	664.41*	717.15*	124.93
			(252.05)	(258.13)	(283.28)	(280.65)	(287.04)
5,600	.7	No	−59.97	23.34	386.12	744.94*	267.69
			(164.95)	(156.41)	(200.59)	(263.80)	(259.45)
5,600	.8	Yes	−27.64	−51.03	117.85	273.44	121.53
			(121.47)	(126.67)	(138.52)	(157.96)	(169.26)

Because so many T groups it is hard to see how changing G and t affect outcomes.

Table 4
Experimental Payment minus Predicted Control Payment for 5-Year Dual-headed Experimental Families, Attrition Families Excluded (Standard Errors in Parentheses)

G ($)	τ	Declining Tax Rate	Preexperimental Payment ($)	Payment for Year of Experiment ($) 1	2	3	4	5	Postexperimental Payment ($)
3,800	.5	No	102.24	345.68	526.02	110.30	390.07	169.82	229.70
			(185.55)	(221.42)	(241.53)	(265.28)	(307.01)	(286.76)	(309.06)
3,800	.7	No	81.16	23.30	−99.33	98.20	−16.42	−122.01	−406.46
			(309.85)	(316.06)	(330.14)	(383.52)	(388.07)	(352.95)	(314.40)
3,800	.7	Yes	6.99	490.00	176.14	23.22	324.70	−59.79	−598.09*
			(234.01)	(288.13)	(272.87)	(300.28)	(386.93)	(331.68)	(102.72)
3,800	.8	Yes	−130.30	349.73	189.80	329.94	1207.82*	1108.49*	307.38
			(271.23)	(286.56)	(280.63)	(365.58)	(463.10)	(487.83)	(453.29)
4,800	.5	No	−23.66	30.15	160.40	399.28	419.73	434.30	251.09
			(183.73)	(208.90)	(199.26)	(236.33)	(247.25)	(254.52)	(242.45)
4,800	.7	No	−129.98	25.71	−4.47	569.10	493.42	219.74	−38.46
			(185.46)	(208.14)	(211.44)	(314.73)	(357.32)	(340.60)	(228.01)
4,800	.7	Yes	75.66	224.96	387.66	340.71	−130.10	34.61	189.49
			(234.21)	(280.43)	(367.56)	(404.05)	(308.90)	(445.67)	(491.52)
4,800	.8	Yes	467.89	325.17	599.43*	398.62	537.21	506.95	346.28
			(252.40)	(276.31)	(274.39)	(280.50)	(365.56)	(351.98)	(337.43)
5,600	.5	No	−224.97	560.51	723.08*	782.53*	592.40	313.82	−53.07
			(286.39)	(298.21)	(306.90)	(327.39)	(366.88)	(387.31)	(325.66)
5,600	.7	No	−158.74	500.18	1194.68*	890.38*	825.39	435.01	588.91
			(239.17)	(311.24)	(416.25)	(391.61)	(467.76)	(609.49)	(510.52)
5,600	.8	Yes	−6.48	193.54	617.29*	906.13*	888.72	877.71	75.21
			(175.15)	(199.51)	(255.89)	(315.98)	(337.38)	(398.38)	(216.12)

Attrition is a real problem here. Key is that they collected earnings data through survey and there is no incentive to stick with it if you expect no benefit (e.g. do not receive NIT)

NIT Experiments: Overall 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

PROBLEMS WITH NIT DESIGN

This early attempt at experimentation in the US was not ultimately successful. Experiments were poorly designed.

-- nonrandom selection into experiment (selected on income) [Lesson: keep it simple random T, C]

-- self reported earnings with incentives for T to underreport so that they got NIT payment [Lesson: need to match to administrative records: UI, SS, firm tax records.]

-- selective sample attrition: after initial year, data collected based on voluntary income reports by families → those in less generous groups/far above break-even point had much less incentive to report → attrition rates higher in these groups → no longer a random sample of treatment + controls [Ashenfelter-Plant JOLE'90]

Nonrandomness undoes the simple T/C comparison that is so powerful in randomized studies. So much statistical modeling was used here.

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

It is unusual to have experimentally manipulated income to identify income effects (on labor supply). This paper provides evidence from lottery winnings.

-- Survey of lottery winners and nonwinners (=winners of small prize) matched to Social Security administrative data to estimate income effects.

* matching to administrative data is a plus (pre trends)

-- Lottery generates random assignment conditional on playing

* variation in prize amount is random

-- Find significant but relatively small income effects: $\eta = w\frac{\delta l}{\delta y}$ = -0.05 to -0.10

-- Identification threat: differential response-rate among groups

* 49% for nonwinners, 42% for winners

TABLE 2—SUMMARY STATISTICS BASIC SAMPLE: PRE-LOTTERY CHARACTERISTICS AND POST-LOTTERY OUTCOMES

Variable	All (N = 496) Mean	(SD)	Nonwinners (N = 259) Mean	Winners (N = 237) Mean	[t-stat]	Big winners (N = 43) Mean	[t-stat]
Yearly prize	26.4	(50.8)	0	55.2	[14.4]	160.0	[20.4]
Year won	1986.2	(1.2)	1986.4	1986.1	[−3.0]	1985.9	[−1.1]
Tickets bought	3.3	(2.9)	2.2	4.6	[10.2]	5.0	[0.9]
Age	50.2	(13.7)	53.2	46.9	[−5.2]	50.3	[1.8]
Age > 55	0.35	(0.48)	0.43	0.27	[−3.9]	0.40	[2.1]
Age > 65	0.15	(0.36)	0.19	0.10	[−2.9]	0.21	[2.6]
Male	0.63	(0.48)	0.67	0.58	[−2.1]	0.84	[3.9]
Years of schooling	13.7	(2.2)	14.4	13.0	[−7.8]	12.8	[−0.6]
College	0.65	(0.48)	0.78	0.51	[−6.6]	0.53	[0.4]
Working then	0.78	(0.41)	0.77	0.80	[0.9]	0.86	[1.1]
Earnings year −6	13.8	(13.4)	15.6	12.0	[−3.0]	14.6	[1.6]
Earnings year −5	14.1	(13.8)	16.0	12.1	[−3.1]	15.2	[1.9]
Earnings year −4	14.2	(14.1)	16.2	12.0	[−3.3]	16.1	[2.5]
Earnings year −3	14.8	(14.8)	16.6	12.8	[−2.9]	17.1	[2.5]
Earnings year −2	15.6	(15.3)	17.6	13.5	[−3.0]	16.8	[1.9]
Earnings year −1	16.3	(15.7)	18.0	14.5	[−2.5]	17.3	[1.5]
Earnings year 0	16.1	(15.8)	18.2	13.7	[−3.3]	13.8	[0.1]
Earnings year 1	15.4	(16.2)	18.5	12.0	[−4.5]	9.5	[−1.4]
Earnings year 2	14.7	(16.3)	17.7	11.4	[−4.3]	8.4	[−1.6]
Earnings year 3	14.2	(16.3)	17.1	10.9	[−4.3]	8.7	[−1.2]
Earnings year 4	13.8	(16.3)	16.9	10.4	[−4.5]	7.5	[−1.6]
Earnings year 5	13.6	(16.3)	16.7	10.3	[−4.4]	7.8	[−1.4]
Earnings year 6	13.2	(16.4)	15.8	10.5	[−3.6]	6.8	[−2.0]
Positive earnings year −6	0.69	(0.46)	0.69	0.70	[0.3]	0.70	[−0.0]
Positive earnings year −5	0.71	(0.45)	0.68	0.74	[1.5]	0.65	[−1.5]
Positive earnings year −4	0.71	(0.45)	0.69	0.73	[1.1]	0.72	[−0.2]
Positive earnings year −3	0.70	(0.46)	0.68	0.73	[1.4]	0.74	[0.2]
Positive earnings year −2	0.71	(0.46)	0.68	0.74	[1.6]	0.70	[−0.7]
Positive earnings year −1	0.71	(0.45)	0.69	0.74	[1.2]	0.70	[−0.7]
Positive earnings year 0	0.71	(0.45)	0.69	0.73	[1.1]	0.70	[−0.6]
Positive earnings year 1	0.68	(0.47)	0.68	0.68	[−0.0]	0.49	[−3.0]
Positive earnings year 2	0.63	(0.48)	0.64	0.62	[−0.5]	0.42	[−3.0]
Positive earnings year 3	0.60	(0.49)	0.62	0.58	[−0.8]	0.40	[−2.8]
Positive earnings year 4	0.58	(0.49)	0.61	0.55	[−1.3]	0.33	[−3.4]
Positive earnings year 5	0.59	(0.49)	0.59	0.58	[−0.4]	0.35	[−3.4]
Positive earnings year 6	0.56	(0.50)	0.57	0.55	[−0.4]	0.33	[−3.4]
Car value	18.2	(17.8)	16.7	20.0	[2.0]	29.6	[3.5]
Net car value	15.5	(14.9)	15.3	15.7	[0.3]	25.7	[4.0]
Housing value	166.3	(111.6)	174.9	156.9	[−1.8]	218.1	[4.4]
Net housing value	122.1	(95.5)	144.6	97.6	[−5.4]	112.3	[1.4]
Retirement accounts	64.7	(102.8)	92.6	34.4	[−6.1]	34.6	[0.0]
Other financial assets	84.3	(151.9)	91.8	76.1	[−1.1]	127.1	[2.0]
Total financial assets	133.4	(192.5)	164.5	99.4	[−3.8]	150.9	[2.0]

Mixes up evaluation of validity of design: Comparing Xs and Y pre-treatment [key is testing for differences between columns 3 & 4]

with

Unconditional treatment effects

Problem: unbalanced T and C groups.

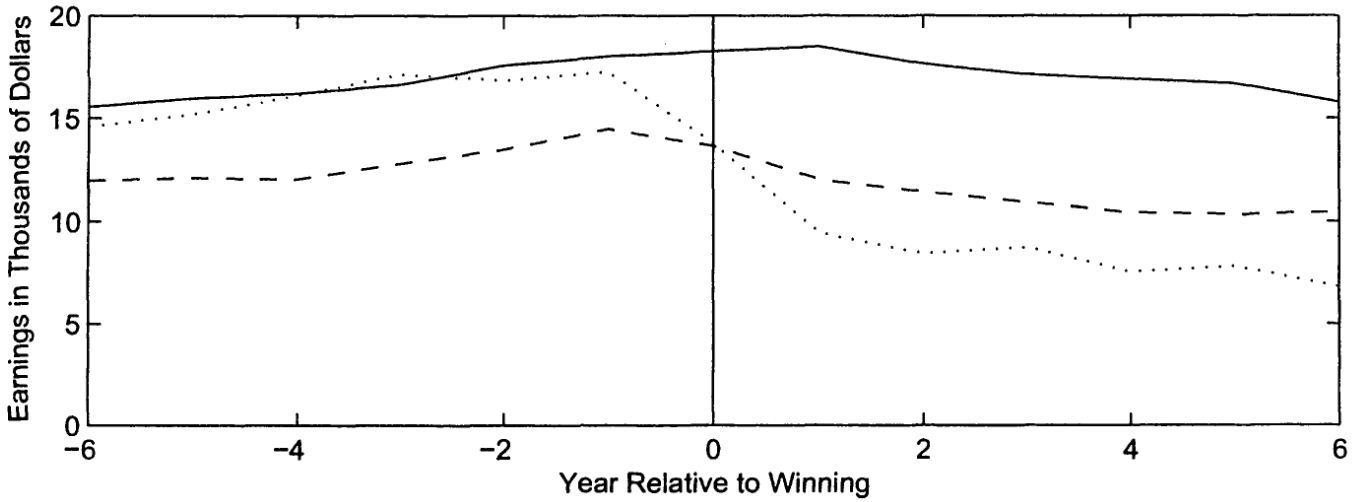


FIGURE 1. AVERAGE EARNINGS FOR NONWINNERS, WINNERS, AND BIG WINNERS

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

Good pre-trends
Larger impact for larger prize winners

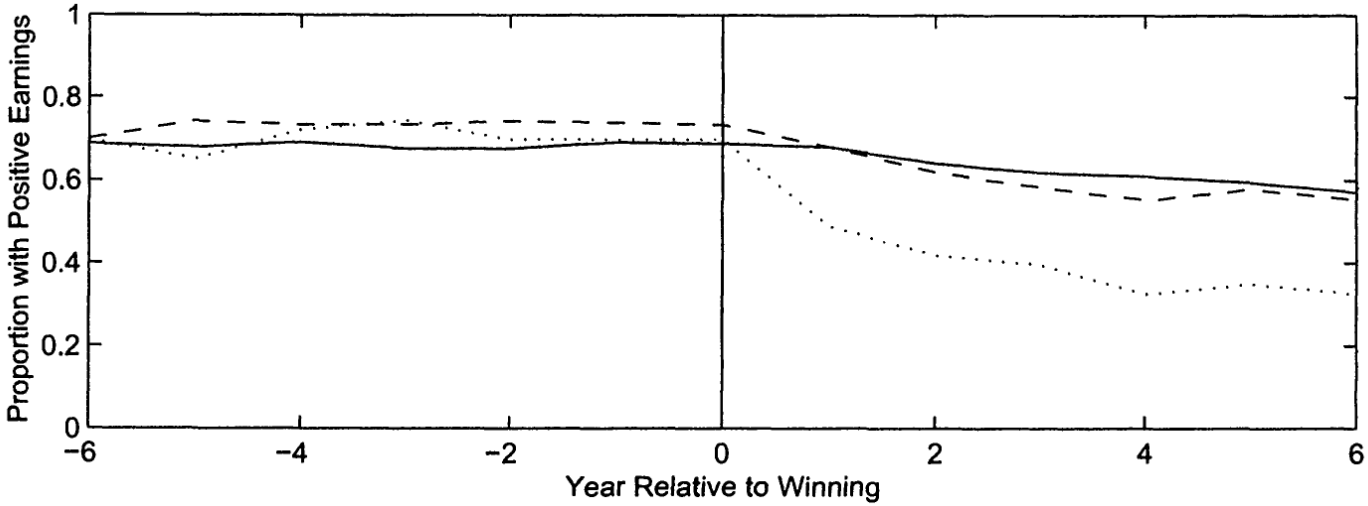


FIGURE 2. PROPORTION WITH POSITIVE EARNINGS FOR NONWINNERS, WINNERS, AND BIG WINNERS

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

Same.

Parametric model results (regress on continuous prize measure) yields estimate for income effect of -0.05 – 0.10

Threats and thoughts:

External validity: lottery winners not random
Attrition

INSTRUMENTAL VARIABLE METHODS

1) Another strategy to overcome endogeneity is instrumenting for wage rate

2) Mroz (1987): often-cited survey/meta-analysis of earlier studies

-- Uses PSID to test widely-used IV's for married women's wage

$$l = \alpha + \beta w + \gamma X + \varepsilon$$
$$w = \theta Z + \mu$$

-- Uses Hausman specification/overidentification test to show that many instruments violate $\mathrm{E}Z\varepsilon = 0$. Goal is to see which are credible.

Hausman Test

1) Suppose you can divide instrument set into those that are credibly exogenous (Z) and those that are questionable (Z^*)

2) Null hypothesis: both are exogenous

3) Alternative hypothesis: Z^* is endogenous

4) Compute IV estimate of β with small and large instrument set and test for equality of the coefficients

5) Note that is often a very low power test (accept validity if instruments are weak)

MROZ RESULTS

Background variables he maintains as credible [unemp rate, parent's ed, wife's age and ed]

Tests show that the following variables fail the Hausman test: labor market experience, age hourly earnings, and previous reported wages

Shows that earlier estimates in the literature are very nonrobust.

This study contributed to emerging view that policy variation (taxes) was necessary to identify parameters.

Blundell et al, *Econometrica*, Use demographics, tax reform in IV setting

TAX REFORMS AND LABOR SUPPLY

Modern studies use tax reforms as a "natural experiment" to evaluate the effect of taxes on labor supply (and other outcomes). Can get around problem of endogenous net of tax wages and wages more generally

- Advantage of tax reform: policies can affect some groups and not others, creating natural treatment and control groups.
- We have seen lots of changes in tax laws to provide experiments to examine.

<u>TRA86 Tax Reform Act of 1986</u>

- Tons of papers on this. Why? (See Auerbach and Slemrod JEL)
- Most significant policy change in postwar period
- Goals of TRA86: Horiz Equity, Efficiency (eliminate tax preferences), Simplicity
- Result: Broaden base + reduce rates MTR
- 1986: 14 brackets, 11% - 50%
- 1990: 5 brackets 0, 15, 28, 33, 28
- Increase standard deduction and personal exemptions
- We will see later papers using this variation to look at impact of taxes on low end (EITC) and high end

EISSA (1995): TRA86 AND MARRIED WOMEN'S LABOR SUPPLY

Never published (not sure why), but great teaching paper and also very influential paper

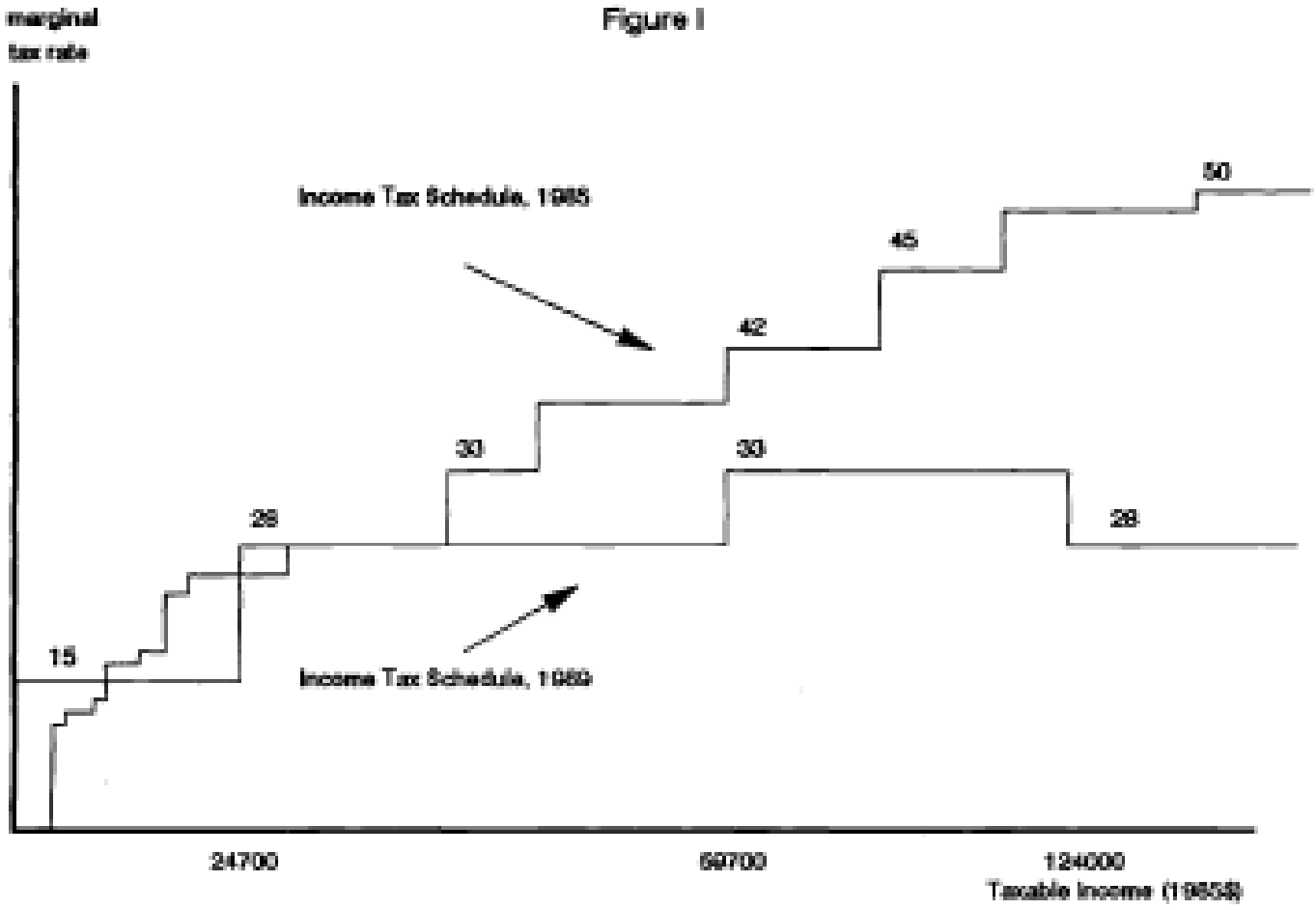
-- Established convincingly that married women are sensitive to taxes, have higher elasticity of labor supply

-- Added to our knowledge that participation margin is more sensitive than hours margin

-- Good example of early difference in difference methodology

-- Eissa focuses on high income women because they had the highest reductions in MTR (see figure from paper)

Figure I


marginal
tax rate
Income Tax Schedule, 1985
Income Tax Schedule, 1989
50
45
42
33
33
28
28
15
24700
59700
124000
Taxable Income (1985$)

Economics: Secondary Earner Labor supply model

Most common approach is to model labor supply of husband and wife sequentially

(1) Husband (or primary earner) maximizes utility ignoring wife (just like single agent model)

$$\text{Max}(l_h, Y) \quad \text{s.t.} \quad w_h h_h + N = Y \quad \rightarrow \quad h_h\,*$$

(2) Wife (or secondary earner) maximizes utility conditioning on husband's optimal labor supply decision

Therefore, she takes $N + w_h h_h^*$ as given

$\text{Max}(l_w, C)$ s.t. $w_w h_w + (w_h h_h^* + N) = C$

Graph this

Comparative statics of secondary earner model

- Earnings of husband increase ↑ (through increase in h or w) → nonlabor income of wife ↑ → income effect → hours and employment of wife ↓.
- Taxes? Decrease in taxes leads to:
  - ↑ net nonlabor income → hours and employment of wife fall
  - ↑ w_w → hours (?), employment ↑
  - KEY: with progressive taxes, she gets the change in MTR which is exogenous to her own labor supply, but comes through her husband. Her first hour MTR is his last hour MTR.

Diff-in-Diff (DD) Methodology:

Step 1: Simple Difference

Outcome: LFP (labor force participation)

Two groups: Treatment group (T) which faces a change [women married to high income husbands] and control group (C) which does not [women married to middle income husband]

Simple Difference estimate: $\Delta = LFP^T - LFP^C$ captures treatment effect if absent the treatment, LFP equal across 2 groups

Note: Rarely holds in nonexperimental setting (always holds when T and C status is randomly assigned)

What to test for: Compare LFP before treatment happened (period 0)
$\Delta_0 = LFP_0^T - LFP_0^C$

Step 2: Diff-in-Difference (DD)

If $\Delta_0 \neq 0$, we can estimate the DD: $\Delta\Delta = (LFP_1^T - LFP_1^C) - (LFP_0^T - LFP_0^C)$
(0 = after reform, 1 = 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.

Regression estimation of Unconditional DD:

$$LFP_{it} = \alpha + \beta_0 AFTER + \beta_1 TREAT + \gamma AFTER * TREAT + \varepsilon$$

$$\hat{\gamma} = (LFP_1^T - LFP_1^C) - (LFP_0^T - LFP_0^C)$$

Diff-in-Diff (DD) Methodology

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

Can 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

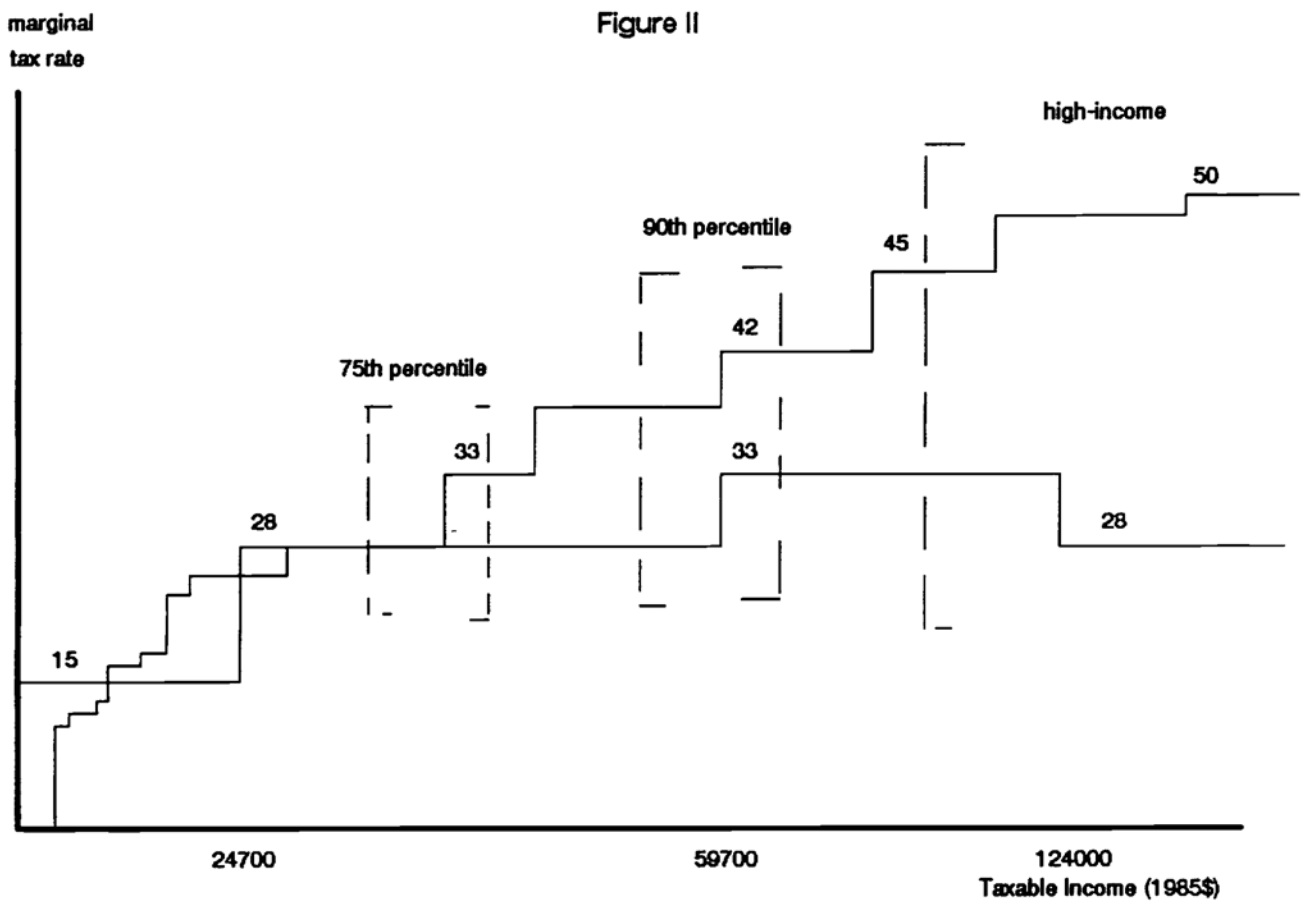
(b) if $DD^{placebo} = 0$, then DD=DDD but DDD has higher s.e.

BACK TO EISSA

Treatment women in >=99th percentile of $\boxed{N + w_h h_h}$ distribution

Control 75th percentile or 90th percentile

Tradeoff: 90th better control but gets some treatment



Data CPS

1984 – 1986 before (83 – 85)

1990 – 1992 after (89 – 91) TRA86 phased in by 88

Predictions?

-- Employment of women in 99th p will rise relative to women in 90th p

→ Her MTR ↓ → net wage ↑ → LFP ↑

-- But we have to believe that her net of tax nonlabor income did not change much. Why?

→ husband's MTR ↓ (→ ↑ earnings) (or no change if elas small)

→ But TRA86 broadened base

→ overall effect on her after tax non-labor income is small

* To the extent which his ↓ net earnings are not captured, then, this estimate is an underestimate of total effect.

Results

Unconditional difference in difference

	Ave Y	ΔMTR(Tab IIa)	ΔLFP (Tab III)	D in D
99th p	> 90K	-13.9 pp	+9.0 pp	
90th p	67K	-6.9 pp	+4.5 pp	+4.5 pp. (2.8) 13%
75th p	47K	-4.1 pp	+5.3 pp	+3.7 pp (2.8) 12%

	Ave Y	ΔMTR(Tab IIa)	Δhours (Tab III	D in D
99th p	> 90K	-13.9 pp	+163	
90th p	67K	-6.9 pp	+96	+67 (64.8) 6%
75th p	47K	-4.1 pp	+55	+108 (65.1) 9%

Conditional D-D

$$\Pr(Work) = \alpha_0 + \alpha_1 Z_{it} + \alpha_2\, high_i + \alpha_3\, Post86_t + \alpha_4 (High_i * Post86)_{it}$$

Z_{it} = age, educ, # kids, young kids, race, CC, year & state fixed effects

<u>Expectations</u>
$\alpha_2 < 0$ baseline inc. effect
$\alpha_3 > 0$ secular trend
$\alpha_4 > 0$ Main test of TRA86

<u>Results</u>
-- Large response for participation, less for hours
-- Consistent w/ lit showing greater responsiveness on participation margin than hours margin (Mroz, Hausman)

Table VIa
Predicted Participation Rate
Probit Estimates

Group	Before TRA86	After TRA86	Change	Differences-in-Differences Estimate
Control: 75th Percentile				
High	.460	.497	.037 (8.1%)	
75th Percentile	.741	.739	-.002 (-0.3%)	.039 (8.4%)
Control: 90th Percentile				
High	.480	.542	.062 (13.1%)	
90th Percentile	.640	.654	.014 (2.2%)	.048 (10.9%)

Predicted Participation is calculated as

$$P(lfp) = \Phi(\bar{z}\hat{a})$$

where z corresponds to the average characteristics in the sample.

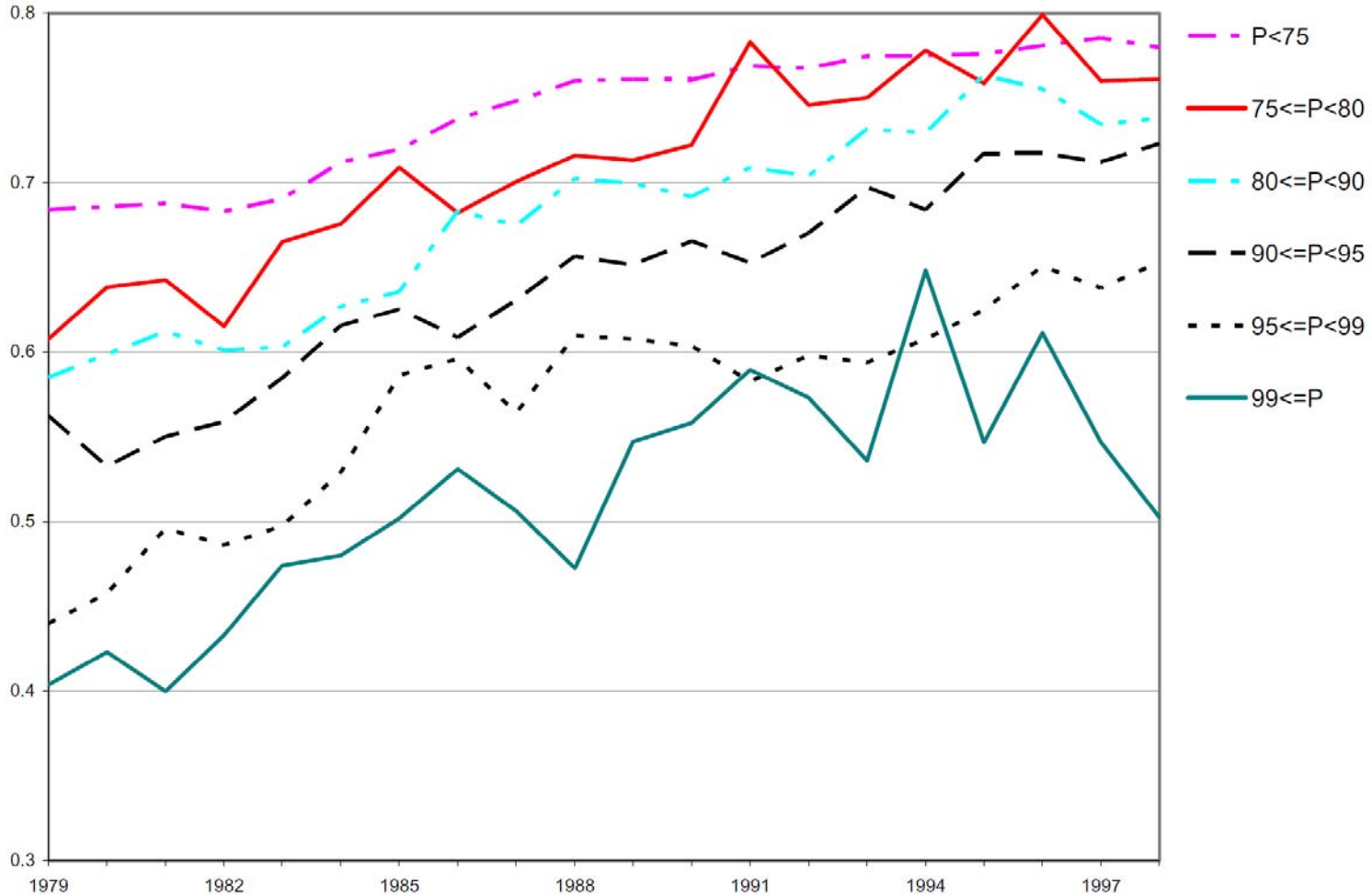
Table VIb
Elasticity of Participation
High-Income Group

Control Group	Elasticity
75th Percentile	0.4
90th Percentile	0.6

Caveats of Eissa Results

- Does "common trends" assumption hold?
  - Possible story: Assortative mating on unobservables. Trend toward "power couples." Used to be that prof men had nonworking spouses; now more common to have working prof spouse. Yet in middle class more stable situation with working middle class spouse.
  - Demand or supply shock to 99th p (e.g. work in different sectors) → different trends for T and C reflecting inequality literature
- LFP is very different between T and C (never a good thing)
- Things to examine in DD model that were not known then:
  - Placebo treatment (use data for pre periods, redo DD using placebo treatment, say comparing year 0 and year -1)
  - Useful (necessary) to plot outcome variables in T and C year on year for whole period; examine whether the trends are similar in pre period. Look for change.
  - Liebman and Saez (2006) plot full time-series CPS plot and show that Eissa's results are not robust using admin data (SSA matched to SIPP) [unfortunately, IRS public tax data does not break down earnings within couples] See next page.

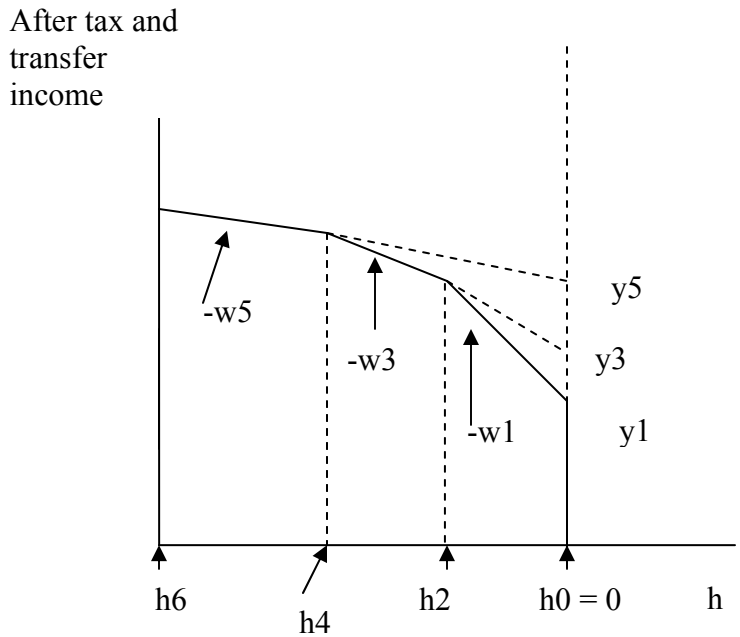
Figure 10
Fraction of Married Women with Positive Annual Earnings by Income Group in March CPS



Notes: Groups are based on other household income (husband's earnings plus asset income) as described in Eissa (1995). Group 1 <=75th percentile. Group 75 is >75th percentile and <= 80th percentile. Group 80 is >80th and <=90th. Group 90 is >90th and <=95th. Group 95 is >95th and <=99th. Group 99 is >99th.

Econometrics of Kinked Budget Constraints: Convex Budget Set

(Hausman's model)



Preferences:

$h^* = h(w, y, \varepsilon)$ (functional form for labor supply equation)

h = observed hours

ε = taste shifter

Comments:
-Virtual income y3, y5 are a function of observed non-labor income and tax system.
-Need one preference assumption (either labor supply equation, IUF or DUF)
-Assume gross wage is exogenous
-Assume gross nonlabor income is exogenous

Ex: Functional form Hausman used for labor supply equation was

$$hi^* = \alpha wi + \beta yi + z\gamma + \varepsilon$$

which implied the following for the IUF:

$$v(wi, yi) = \exp(\beta wi)[yi + \frac{\alpha}{\beta} wi - \frac{\alpha}{\beta\text{^}2} + \frac{z\gamma}{\beta}]$$

4 Steps in constructing the likelihood function:

1. What do you observe?
2. Identify possible states
3. Determine economic decision rule that justifies each choice
4. Derive probabilities associated with each choice

(Step 1) What do you observe?
Hours (0, or continuous hours worked)
Hourly wage rate, for workers
Nonlabor income
Covariates

(Step 2) States of the World

0	$h = 0$
1	$0 < h < h2$
2	$h = h2$
3	$h2 < h < h4$
4	$h = h4$
5	$h4 < h < h6$
6	$h= h6$

Define the labor supply function for each segment:

$h(wi,yi,\varepsilon)$ linear labor supply curve for net wage wi, and net nonlabor income yi
i=1,2,3

Ex:	w1=w	(no taxes)
	w3=w(1-t1)	1st marginal tax rate
	w5=w(1-t2)	2nd marginal tax rate
	y1=N	observed nonlabor income
	y3	virtual income y3
	y5	virtual income y5

(Step 3) Economic Decision Rules

State 0

$h = 0$

$h(w1, y1, \varepsilon) \leq 0$

Desired hours given w1, y1 are <0

State 1

$0 < h < h2$

$h = h(w1, y1, \varepsilon)$

Desired hours given w1,y1 are between 0 and h2

State 2

$h = h2$

$h(w1, y1, \varepsilon) \geq h2$ AND $h(w3, y3, \varepsilon) \leq h2$

Note that being on kink has higher probability than any given point on segment.

State 3

$h2 < h < h4$

$h = h(w3, y3, \varepsilon)$

State 4

$h = h4$

$h(w3, y3, \varepsilon) \geq h4$ AND $h(w5, y5, \varepsilon) \leq h4$

State 5

$h4 < h < h6$

$h = h(w5, y5, \varepsilon)$

State 6

$h = h6$

$h(w5, y5, \varepsilon) \geq h6$

Then translate desired hours rule into rule about unobservable (ε)
Derive probability that choice was made

<u>(Step 4) Create Likelihood Function</u>

$$L(h) = \prod_{i=0,2,4,6} [\Pr(\delta i = 0)]^{(\delta i)} \prod_{i=1,3,5} [f(h \mid wi, yi)]^{(\delta i)}$$

where δi = 1 if state i is observed, and = 0 otherwise.