Taxation, labor supply & income distribution

Andreas Peichl

ZEW & U Mannheim

Canazei, 2014-01-14

Motivation

- How do taxes (tax reforms) affect labor supply and income distribution?
- Optimal design of (redistributive) public policies depends on behavioral responses:
  - Diamond and Saez (2011): optimal top tax rate: $\tau = \frac{1-g}{1-g+ae}$ (g = marginal social welfare weight, a = parameter of Pareto distribution, e = LS elasticity)
  - With $e = 0.25$: $\tau = 73\,\%$; with $e = 0.6$: $\tau = 53\%$ (for US)
- This talk: Overview on how to estimate labor supply preferences / elasticities and some applications – focus on structural models
- More applications for welfare and policy analysis: see André & Ugo ...

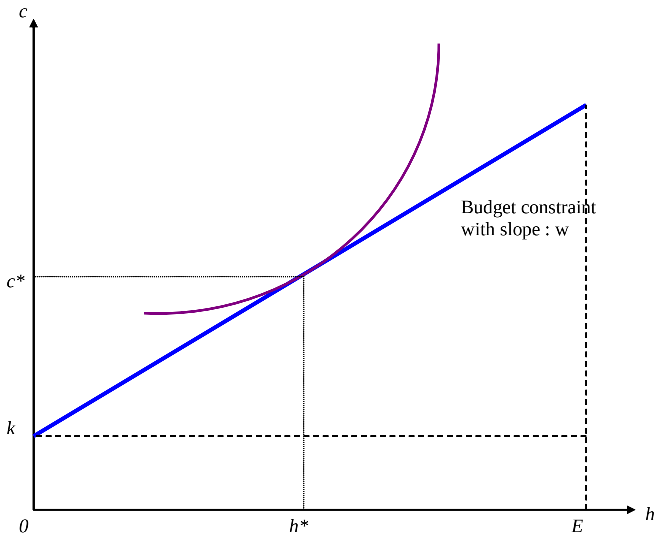
Outline

1. Theoretical background
2. Identification
3. Structural LS models
4. Applications of structural models
5. Limitations & comparison with (natural-) experiments
6. Appendix: examples NOT using structural models

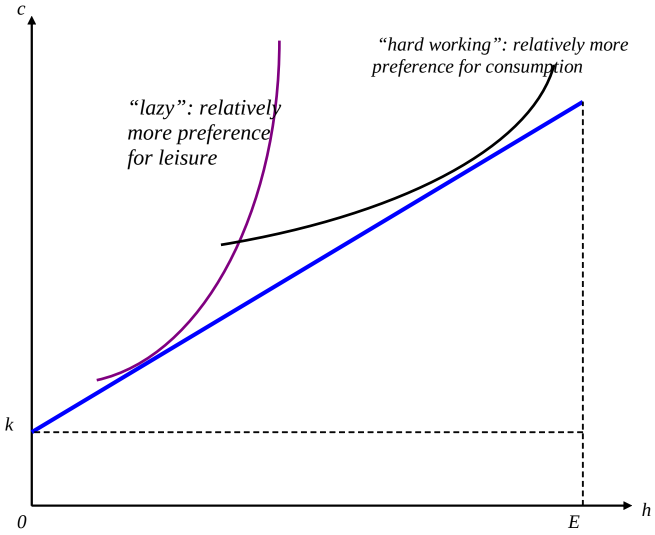
Outline

1. Theoretical background

Basic LS model



Basic LS model



Outline

2. Identification

Identification

- Understanding **causal relationship** behind empirical results
- Y and X are correlated. 3 reasons which are not mutually exclusive:
  - **Cause**: Changes in X drive changes in Y
  - **Reverse Cause**: Changes in Y drive changes in X
  - **Correlated variable**: Changes in Z drive X and Y
- 4 broad approaches for identification (see appendix for examples):
  - *Field Experiments* – you generate the variation
    - US NIT experiments in 1960s/70s (Ashenfelter and Plant 1990)
  - *Natural Experiments* – you know what generated the variation
    - Eissa (1995): top income MTR cut in 1986
  - *IVs* – you have a variable that can provide you exogenosu variation
    - Blau/Kahn (2007) using "grouping instrument" (Blundell et al. 1998)
  - *Parametric Identification* – you rely on econometric assumptions

Outline

3. Structural labor supply models

Structural LS

- Disadvantage: identification less credible (usually due to parametric assumptions only; but validation against quasi-experimental methods)
    - Nonlinearities and discontinuities from tax-benefit rules: individuals with same gross wage receive different net wages.
    - Regional variation in tax-benefit rules
- Advantages of structural model:
    - capture behavior of the whole population (under certain assumptions)
    - Estimation of LS preferences can be used to simulate effects of alternative policies
- We focus on two margins of LS:
    - extensive margin: participation (social benefits may discourage employment)
    - intensive margin: hours (high marginal taxes may reduce effort supply)

Discrete Choice Models

- van Soest (1995), Aaberge et al. (1995, 1999), ...
- Agents i can choose between limited number J of discrete alternatives
  - corresponding to work duration h_j, $j = 1, ...J$ (or job char.): e.g. $h_1 = 0$ (inactive), $h_2 = 20$ (part-), $h_3 = 40$ (full-), $h_4 = 50$ (overtime)
- For individual i, the utility from choosing alternative j is

$$V_{ij} = U_{ij} + \varepsilon_{ij}$$

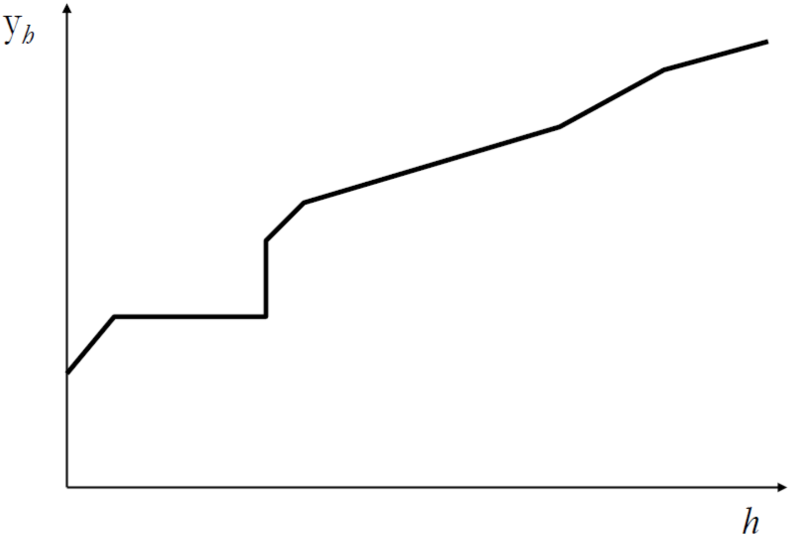
$U_{ij} = \mathbf{x}'_{ij}\mathbf{b}_j$: deterministic function; ε_{ij} unobservable random term.

- Multinomial Logit (McFadden 74): Prob of choosing alternative j:

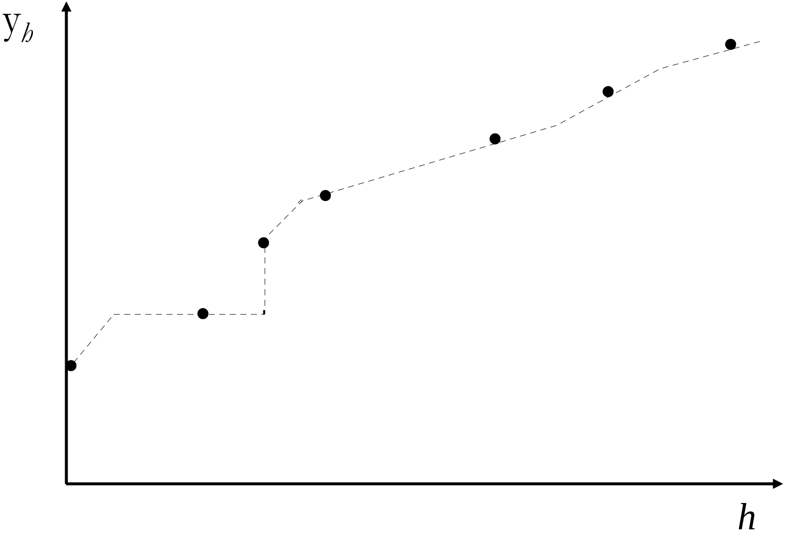
$$P_{ij} = \frac{\exp(U_{ij})}{\sum_J \exp(U_{ik})} = \frac{\exp(\mathbf{x}'_{ij}\mathbf{b}_j)}{\sum_J \exp(\mathbf{x}'_{ik}\mathbf{b}_k)}$$

- Random Utility Model: Direct relationship btw maximizing (log) likelihood and utility-max behavior

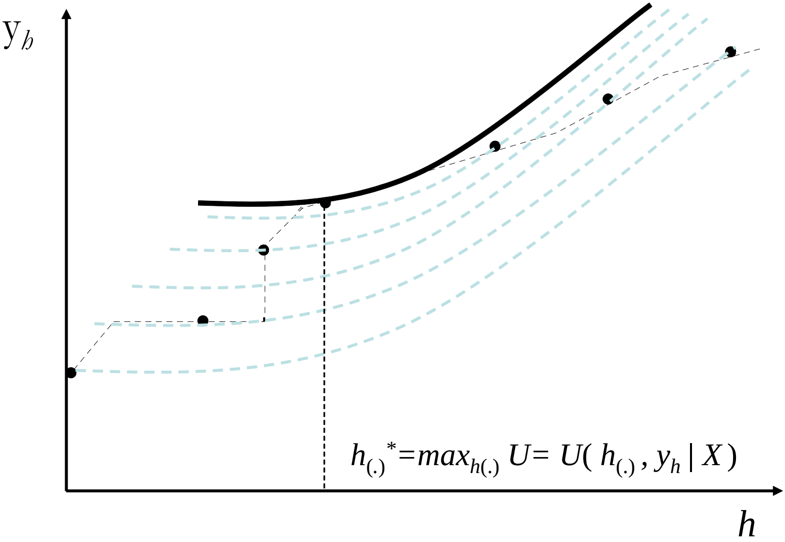
A discrete choice set



A discrete choice set



A discrete choice set

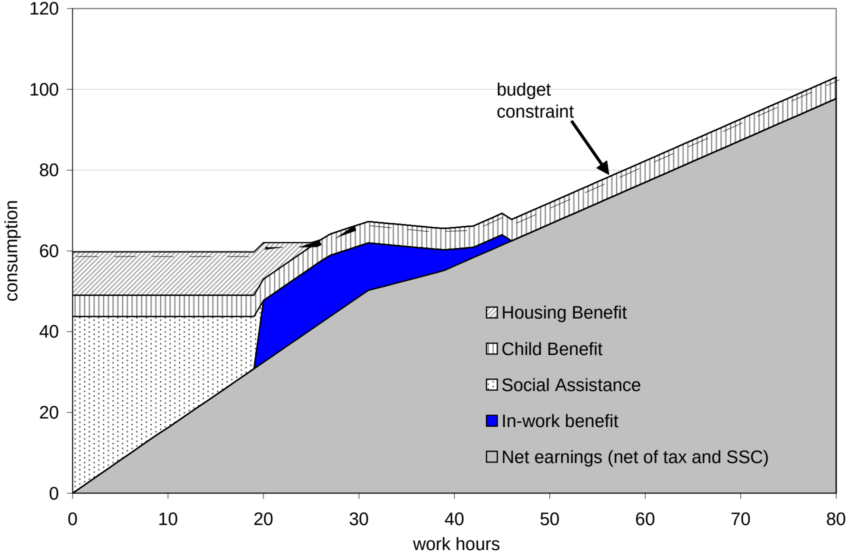


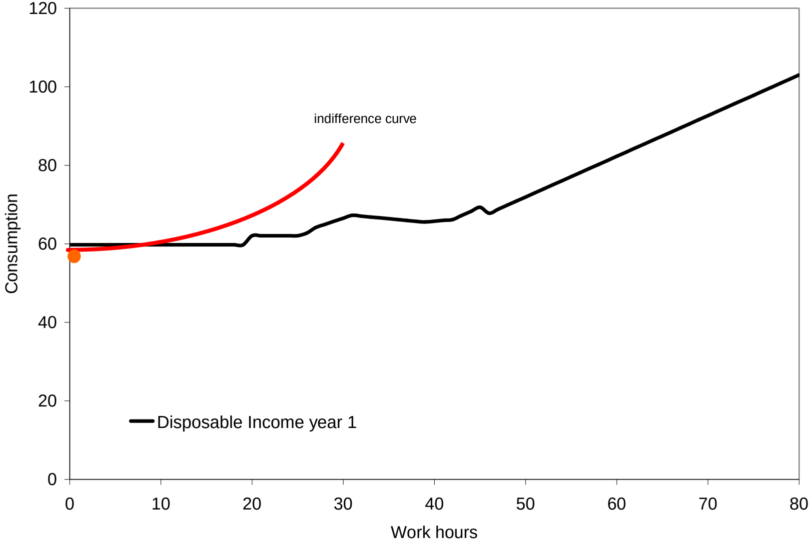
Real-world tax-benefit policies

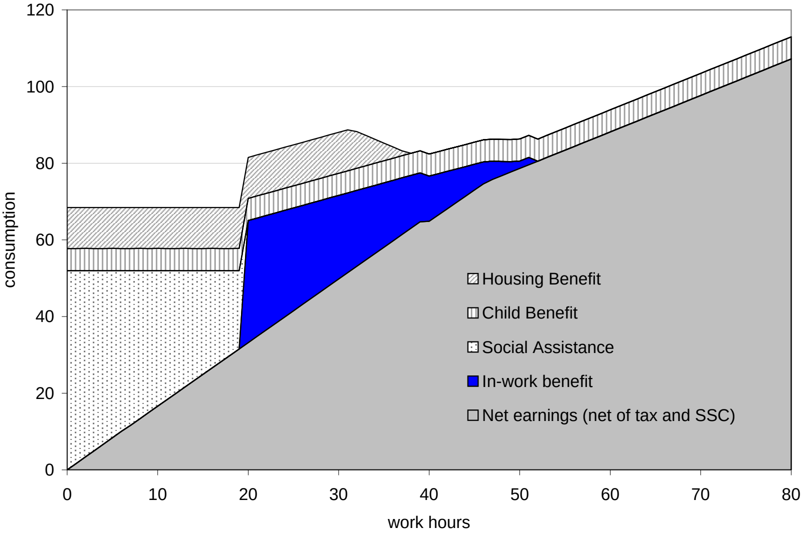
- In the present static framework, consumption c_{ij} is equivalent to disposable income and calculated as a function

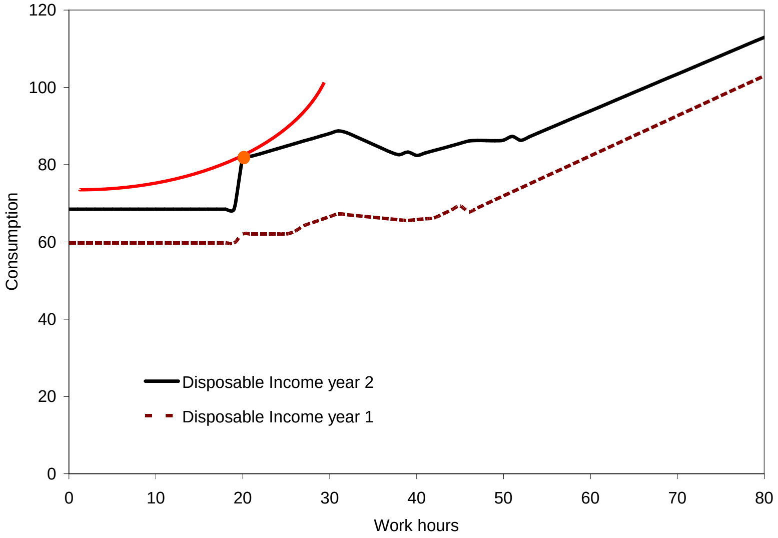
$$c_{ij} = D(w_i h_j, k_i, z_i)$$

  of gross labor income $w_i h_j$ (for any alternative j), exogenous income k_i and household characteristics z_i
- Function D relies on a fairly complex set of tax-benefit rules approximated by tax-benefit *microsimulation (e.g. EUROMOD)*


budget
constraint
Housing Benefit
Child Benefit
Social Assistance
In-work benefit
Net earnings (net of tax and SSC)
consumption
work hours
120
100
80
60
40
20
0
0
10
20
30
40
50
60
70
80


120
100
80
60
40
20
0
Consumption
indifference curve
Disposable Income year 1
0
10
20
30
40
50
60
70
80
Work hours


120
100
80
60
40
20
0
consumption
0
10
20
30
40
50
60
70
80
work hours
Housing Benefit
Child Benefit
Social Assistance
In-work benefit
Net earnings (net of tax and SSC)



120
100
80
60
40
20
0
Consumption
0
10
20
30
40
50
60
70
80
Work hours
Disposable Income year 2
Disposable Income year 1

Outline

4. Applications

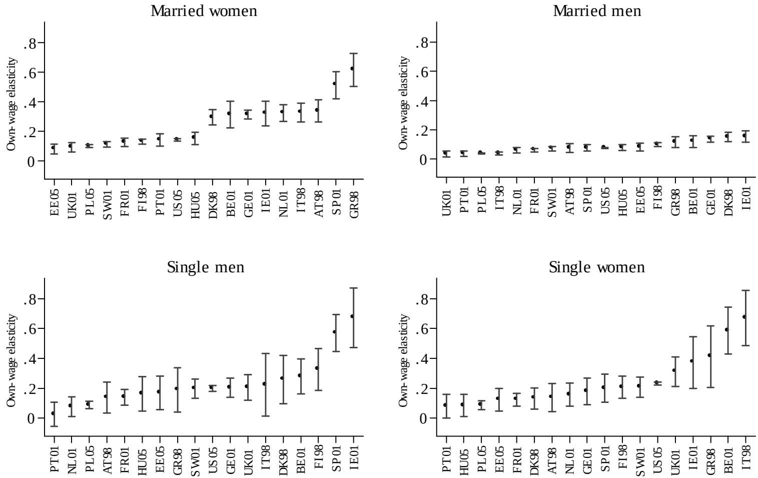
Applications

1. LS elasticities – Bargain, Orsini, Peichl (2014)
2. LS (& distributional) effects of tax reforms – Bargain et al. (2013b)[1]
   - many more applications – e.g. Aaberge et al. (1995/99, ...), Blundell et al. + surveys
3. Design of optimal tax system (Aaberge / Colombino, 2012/13) and "inverse" optimal tax problem (Bargain et al., 2014 a,b)
4. Effect of tax policy on inequality – Bargain et al. (2013b)
5. Comparison of welfare and income distribution – Decoster, Haan (2010), Bargain et al. (2013a)

[1] "et al." = Dolls / Neumann / Pestel / Peichl / Siegloch + Decoster / Fuest

1. Labor Supply Elasticities in the EU/US

- BOP (14) use EUROMOD (& TAXSIM) to estimate LS elas
- Common specification, various robustness checks



1. Labor Supply Elasticities

- Bargain & Peichl (2013): survey of DC LS elas & meta analysis
  - preference changes over time and differences in estimation methods
  - Other surveys: Pencavel (1986), Heckman (1993), Blundell & MaCurdy (1999), Meghir & Phillips (2010), Keane (2011)
  - Macro literature: Chetty et al. (2011), Chetty (2012), Keane & Rogerson (2012), Jäntti et al. (2013)
- Löffler et al. (2014): analysis of DC modelling assumptions
  - assumption of wage exogeneity crucial for outcomes!
  - Relaxing it: LS ela increases from 0.25 ($\tau = 73\,\%$) to 0.6 ($\tau = 53\%$)

Table 6: Joint estimation results single females

	(1)	(2)	(3)	(4)	(5)
Part time			0.0491**	-0.0873***	-0.142***
			(0.0249)	(0.0332)	(0.0356)
Over time			-0.121***	-0.0465	-0.0628**
			(0.0293)	(0.0288)	(0.0268)
N	7455	7455	7455	7455	7455
Pseudo R^2	0.172	0.468	0.471	0.489	0.492
Log-likelihood	-1715.5	-2214.2	-2202.8	-2128.3	-2114.7
AIC	3475.1	4496.3	4477.7	4334.7	4315.4
BIC	3627.2	4731.5	4726.7	4604.4	4612.8
Correlation					
$\rho_{FC,\ln W}$				-0.999***	-0.928***
				(0.003)	(0.042)
$\rho_{C,\ln W}$					-0.123
					(0.092)
$\rho_{L,\ln W}$					0.312***
					(0.107)
Elasticities					
Intensive	0.223	0.249	0.293	0.484	0.577
Extensive	0.369	0.313	0.365	1.047	0.974

2. Labor supply effects of tax reforms

- Bargain et al. (2013b) use EM to simulate EU tax-benefit system:

	Baseline		*Reform scenarios*			
	Hours worked		Change labour supply (FTE) in %			
	Mean/ week	FTE in Mio.	*EUavg* *Sc. 1*	*EUavg* *Sc. 2*	*EUavg_p* *Sc. 1*	*EUavg_p* *Sc. 2*
EU	29.9	71.1	-0.1	-1.0	-0.6	-2.6
AT	32.0	2.1	-0.1	-0.9	-0.6	-2.2
BE	32.7	2.6	2.5	5.6	1.8	3.7
FI	33.2	1.7	2.0	4.6	1.6	3.6
FR	30.8	17.3	0.5	1.3	0.3	0.6
GE	30.0	23.5	0.4	0.0	-0.3	-2.3
GR	25.3	1.3	-3.1	-10.2	-3.7	-12.1
IR	28.1	0.7	-1.3	-4.7	-1.8	-6.6
IT	26.7	8.4	-1.4	-4.9	-1.9	-6.6
NL	31.3	5.2	0.2	-0.2	-0.3	-1.7
PT	34.5	2.0	-0.3	-1.2	-0.5	-2.0
SP	27.7	6.4	-2.4	-8.0	-2.9	-9.9
	net taxes base		*relative change in net taxes*			

	EUavg – Sc. 1			
	%+	%dY	Gap+	Gap-
EU	55	0.0	19.9	-23.6
AT	35	-2.2	17.2	-28.3
BE	50	-0.3	19.2	-22.3
FI	51	0.0	19.1	-19.6
FR	31	-2.8	14.5	-26.5
GE	66	1.0	20.0	-24.5
GR	80	8.5	30.8	-12.7
IR	28	-2.7	21.2	-33.8
IT	63	0.9	19.6	-21.8
NL	40	-0.6	16.8	-16.5
PT	68	4.5	29.4	-17.9
SP	60	0.9	20.1	-20.2

2. Labor supply effects of tax reforms

- More examples: see André & Ugo,... + Rolf's presentations in the last Winter School(s)
- Ugo's talk: Important (how) to account for (demand) constraints / job opportunities (see also Aaberge et al., 1995/99, ...; Creedy & Duncan, 2005, Haan & Steiner, 2006, Peichl & Siegloch, 2012) and frictions (Chetty, 2012)

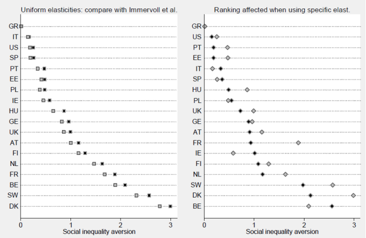
3. Design of optimal tax system

- Aaberge & Colombino (2012): Simulate EOp-optimal tax:

Table 6. Optimal three-parameter tax systems under various EOp social objective criteria $\left(\tilde{W}_k\right)$

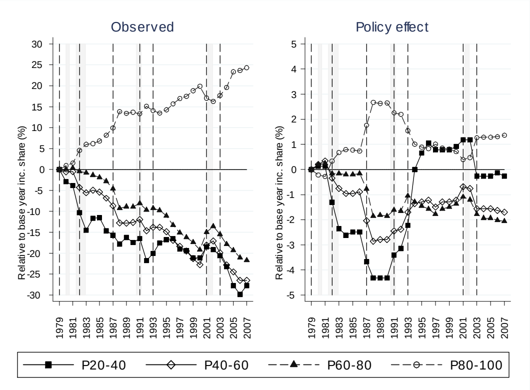
k	1	2	3	∞
t_1	.856	.251	0	0
t_2	.776	.531	.168	0
c	12,500	3,500	-3,500	-5,790

- "inversion": inequality aversion of the planner given LS & tax system?



4. Effect of tax policy on inequality

- Decomposition methods (accounting for labor supply), e.g. Bargain & Callan (2010), see also Dardanoni & Lambert (2002), Lambert & Thoresen (2009)
- Bargain et al. (2013b): effect of tax policy reforms on inequality in US over time (applying τ in $t+1$ to y in t and vice versa):



5. Comparison of welfare and income distribution

- Decoster & Haan (2010, for DE) and Bargain et al. (2013a, for EU&US) use DC LS model to retrieve individual preferences
- compare welfare criteria respecting preferences (Fleurbaey) to income

Table 3 Average percentile position of households in the global welfare ranking - by country and metrics

	Income	*Ref. preferences* Any metric	*Full heterogeneity in preferences* Rent	RW *p25*	RW *p50*	RW *p75*	Wage	Δpp Rent-Wage
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AT	43.6	47.4	41.3	49.1	54.4	58.1	61.0	19.7
BE	49.2	48.6	49.9	47.9	45.4	43.3	42.1	- 7.9
DK	47.2	42.5	48.0	39.9	35.2	32.2	31.3	- 16.7
FI	29.7	23.9	34.3	18.6	15.5	13.7	13.9	- 20.4
FR	34.4	34.5	34.1	35.5	36.1	37.1	37.3	3.2
GE	36.3	38.9	35.9	40.4	43.8	46.7	50.4	14.5
IE	53.1	56.2	46.5	53.8	60.6	66.5	73.9	27.4
NL	47.6	51.3	47.4	53.0	57.1	60.4	64.6	17.2
PT	19.1	17.8	21.8	15.4	13.9	12.8	12.3	- 9.5
SW	38.1	33.4	41.9	29.1	25.8	24.0	23.8	- 18.2
UK	45.0	45.7	44.2	46.2	47.1	47.7	48.4	4.2
US	63.3	62.2	63.4	61.7	60.1	58.5	56.7	- 6.7

Note: For each metric, we compute the percentile position of each household in the global ranking and average them across all households from the respective country. Reference wages for the "rent + reference wage" metrics (RW) are *p25*-, *p50*- and *p75*-wages of the

Outline

5. Limitations & comparison with (natural-) experiments

Limitations of Structural Models

- Can we trust ex-ante predictions from structural LS models?
  - not all institutions accounted for (child-care, in-kind transfers)
  - possibly strong assumptions on household decision making process
  - models often static, unitary, supply-side
- Non-hours responses: income taxes distort many margins beyond hours of work / participation
  - focus on elasticity of total earnings (wh) or taxable income as a broader measure of labor supply (Saez et al., 2012, JEL)

Comparison with Natural Experiments

Natural experiments:

- frequently used to assess labor supply (or tax base) elasticity
  - esp. Difference-in-difference (DiD) or Regression Discontinuity (RD)
- Advantage: no need to impose structure / capture variety of behavior
- Issues with DiD estimators:
  - difficult to find convincing control vs treatment group
  - unobserved characteristics may affect selection into treatment and outcomes
- Concerns about limited policy relevance:
  - limited external validity of DiD (specific in time and space) and RD (around discontinuity)
  - only certain subset of population
- Also: often no direct link to welfare measure

Comparison with Field Experiments

- Recent literature advocate **experimental evaluation designs**
- Directly solve natural-experiment problems if true randomization
- Huge literature on welfare programmes in developing countries
  - e.g. conditional cash transfer programs like Progresa (Attanasio et al, 2003, Alzúa et al., 2010)
- But still problem of external validity:
  - overgeneralization: impact of treatment on group studied may not be informative about impact on other groups
  - randomization bias: individuals who are willing to participate in trial not representative of those who would participate in subsequent program
- Hawthorne effect (subjects' behavior in trial may be different from what it would normally be, ex: teachers in the STAR projects)

Validating Structural Models using Experiments

Validation / comparison of structural models and experiments

- Rarely done – need more systematic feedback into structural design
- Need to mix structural model and (natural) experiments
  - structural model captures behavior and policy effect
  - how well does it perform on subgroups (experiment)?
  - behavior captured by structural model while "other" changes captured by control group
- Examples: Hansen and Liu (2012), Geyer et al. (2012), Pronzato (2012), Thoresen and Vatto (2012), Bargain and Doorley (2013)
  - Findings: structural model usualyl performs well (but some execptions)

Outline

6. Conclusions

Conclusions

- LS responses important for policy and welfare analysis
- Structural models useful tools
- but need to think (more) carefully about identification (e.g. use quasi-experimental variation to estimate parameters)
  - "*It could be especially problematic twenty years from now, when President Chelsea Clinton looks for an economist to appoint to head the Federal Reserve, and the only thing she can find in the American Economic Association are experts on game shows and sumo wrestling.*" (Mankiw, 2007)
- Alternative approach: sufficient statistics (Chetty, 2009)

The End

Thank you!

Comments? Questions?

peichl@zew.de

Appendix

1. Empirical examples

1. Experiments
   - Examples: NIT experiment conducted in 1960s/70s in several US cities
   - Findings (Ashenfelter and Plant 1990): Significant labor supply response but small overall (Implied earnings elasticity for males (females) around 0.1 (0.5))
   - But: various problems with experimental design and very expensive to conduct
2. Tax reforms as natural experiments
   - Example: Eissa (1995): TRA 1986 cut top income MTR from 50% to 28% from 1986 to 1988
   - DiD strategy: compare women in top 1% households (treatment) with women in 90th percentile and 75th percentile (controls)
   - Labor supply elasticity of around 1
   - However, several problems with DiD design: Liebman and Saez (2006) show that Eissa's results are not robust using admin data

3. Instruments
    - Examples: Blau and Kahn (2007) using "grouping instrument" (Blundell, Duncan, Meghir, 1998)
4. Bunching
    - Saez (2010) observes that only non-parametric source of identification for elasticity in a cross-section is amount of bunching at kinks

2. Experiments

Negative Income Tax

- Best way to resolve identification problems: exogenously increase the marginal tax rate
- NIT experiment conducted in 1960s/70s in Denver, Seattle, and other cities
- First major social experiment in U.S.
- Provided lump-sum welfare grants G combined with a steep phaseout rate τ (50%-70%)
- Analysis by Rees (1974), Ashenfelter and Plant (1990), and others
- Several groups, with randomization within each; approx. N = 75 households in each group

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

Source: Ashenfelter and Plant (1990)

NIT Experiments: Findings (Ashenfelter and Plant 1990)

- Significant labor supply response but small overall
- Implied earnings elasticity for males around 0.1
- Implied earnings elasticity for women around 0.5
  - Response of women is concentrated along the extensive margin

Problems with Experimental Design

Estimates from NIT not considered very credible today for two reasons:

1. Self reported earnings
   - Treatments had financial incentives to under-report earnings.
   - Reported earnings not well correlated with actual payments
     $\rightarrow$Lesson: need to match with administrative records
2. Selective attrition
   - After initial year, data was collected based on voluntary income reports by families to qualify for the grant
   - Those in less generous groups/far above breakeven point had much less incentive to report
   - Consequently attrition rates were much higher in these groups

     $\rightarrow$No longer a random sample of treatment + controls

3. Tax reforms as natural experiments

Tax Reform Variation

- Modern studies use tax changes as natural experiments
- Representative example: Eissa (1995)
- Uses the Tax Reform Act of 1986 to identify the effect of MTRs on labor force participation and hours of married women
- TRA 1986 cut top income MTR from 50% to 28% from 1986 to 1988
  - But did not significantly change tax rates for the middle class
- Substantially increased incentives to work of wives of high income husbands relative wives of middle income husbands
- DD strategy: compare women in top 1% households (treatment) with women in 90th percentile and 75th percentile (controls)
- Data: CPS, 1983-85 and 1989-91

Table IIa
Marginal Tax Rate

Group	Before TRA86	After TRA86	Change	Relative Change
High	.521 (.002)	.382 (.001)	-.139 (.002)	
75th Percentile	.365 (.001)	.324 (.001)	-.041 (.001)	**-.098** **(.002)**
90th Percentile	.430 (.001)	.360 (.001)	-.07 (.001)	**-.069** **(.002)**

The marginal tax rate is calculated using family wage and salary, self-employment, interest, dividend, farm and social-security income. I assume all couples file jointly, and that all itemize their deductions. Itemized deductions and capital gains are imputed using Statistics of Income data. These figures include the secondary earner deduction, as well as social security taxes. Standard errors are in parentheses. Before TRA86 is tax years 1983-1985; After TRA86 is tax years 1989-1991.

Source: Eissa 1995

Table III
Differences-in-Differences Estimates
CPS Married Women Before and After TRA86

A: Labor Force Participation

Group	Before TRA86	After TRA86	Change	Difference-in-Difference
High	0.464 (.018) [756]	0.554 (.018) [718]	0.090 (.025) {19.5%}	
75th Percentile	0.687 (.010) [3799]	0.740 (.010) [3613]	0.053 (.010) {7.2%}	**0.037 (.028) {12.3%}**
90th Percentile	0.611 (.010) [3765]	0.656 (.010) [3584]	0.045 (.010) {6.5%}	**0.045 (.028) {13%}**

Source: Eissa 1995

B: Hours Conditional on Employment

Group	Before TRA86	After TRA86	Change	Difference-in-Difference
High	1283.0 (46.3) [351]	1446.3 (41.1) [398]	163.3 (61.5) {12.7%}	
75th Percentile	1504.1 (14.3) [2610]	1558.9 (13.9) [2676]	54.8 (20.0) {3.6%}	**108.6 (65.1) {9.4%}**
90th Percentile	1434.1 (16.4) [2303]	1530.1 (15.9) [2348]	96.0 (22.8) {6.8%}	**67.3 (64.8) {6.2%}**

Each cell contains the mean for that group, along with standard errors in (), number of observations in [], and % increase in {}. Means are unweighted.

Source: Eissa 1995

Eissa 1995: Results

- Participation elasticity around 0.4 but large standard errors
- Hours elasticity of 0.6
- Total elasticity (unconditional hours) is $0.4 + 0.6 = 1$

Eissa 1995: Caveats

- Does the common trends assumption hold?
- Potential story biasing the result:
  - Trend toward "power couples" and thus DD might not be due to taxes
  - In the 1980s, professionals had non-working spouses
  - In the 1990s, professionals married to professionals
  - While for middle class, always married to working middle class wives
- Problem: starting from very different levels for T and C groups
- Liebman and Saez (2006) show that Eissa's results are not robust using admin data (SSA matched to SIPP)

Bianchi, Gudmundsson, and Zoega 2001

- Use 1987 “no tax year” in Iceland as a natural experiment
- In 1987-88, Iceland switched to a withholding-based tax system
- Workers paid taxes on 1986 income in 1987; paid taxes on 1988 income in 1988; 1987 earnings never taxed
- Data: individual tax returns matched with data on weeks worked from insurance database
- Random sample of 9,274 individuals who filed income tax-returns in 1986-88

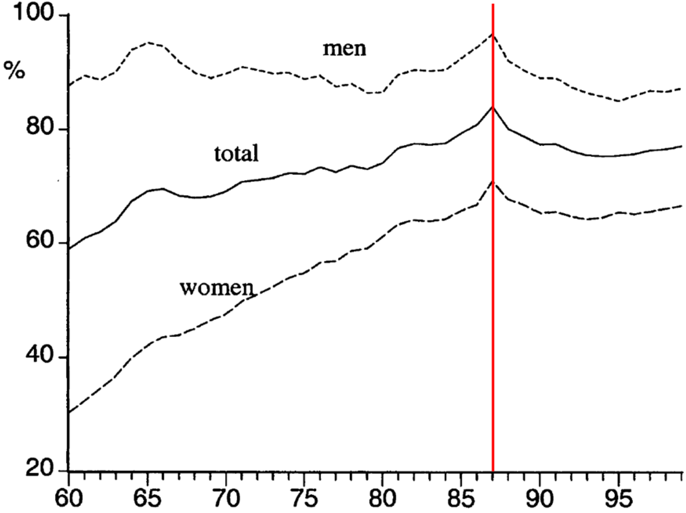


FIGURE 1. THE EMPLOYMENT RATE IN ICELAND, 1960–1996

Source: Bianchi, Gudmundsson, and Zoega 2001

Bianchi, Gudmundsson, and Zoega 2001

- Large, salient change: $\Delta \log(1 - MTR) \approx 49\%$, much bigger than most studies
- Note that elasticities reported in paper are w.r.t. *average* tax rates:

$$\begin{aligned}
\varepsilon_{L,T/E} &= \frac{\sum (L_{87} - L_A)/L_A}{\sum T_{86}/E_{86}} \\
\varepsilon_{E,T/E} &= \frac{\sum (E_{87} - E_A)/E_A}{\sum T_{86}/E_{86}}
\end{aligned}$$

- Estimates imply earnings elasticity w.r.t. marginal tax rate of roughly 0.37 (Chetty 2012)

4. Grouping instrument

Changing Elasticities: Blau and Kahn 2007

- Identify elasticities from 1980-2000 using "grouping instrument" (Blundell, Duncan, Meghir, 1998, Econometrica)

    1. Define cells (year/age/gender/education) and compute mean wages
    2. Instrument for actual wage with mean wage

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

- Result: total hours elasticity (including int + ext margin) shrank from 0.4 in 1980 to 0.2 today

- Interpretation: elasticities shrink as women become more attached to the labor force

5. Bunching

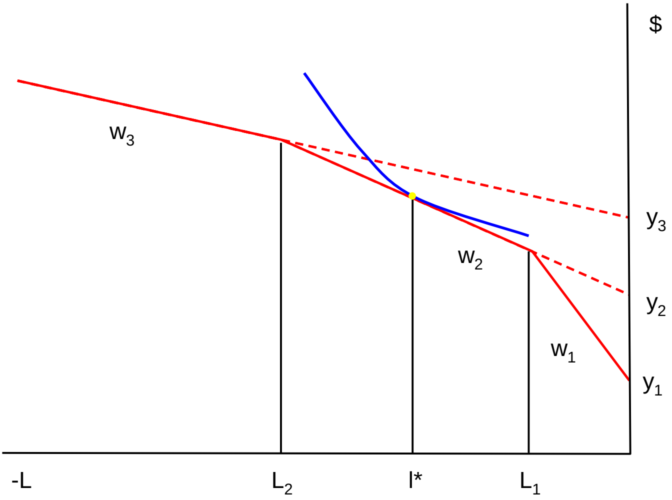
Traditional labor supply

- Traditional approach to estimating elasticities with non-linear budget sets pioneered by Hausman (1981)
- Assume an uncompensated labor supply equation:

$$l_i = \alpha + \beta w_i(1 - \tau_i) + \gamma y_i + \upsilon_i$$

- Error term υ_i is normally distributed with variance σ^2
- Observed variables: w_i, τ_i, y_i, and l_i
- Technique: (1) construct likelihood function given observed labor supply choices on NLBS, (2) find parameters (α, β, γ) that maximize likelihood
- Important insight: need to use "virtual incomes" in lieu of actual unearned income with NLBS

Non-Linear Budget Set Estimation: Virtual Incomes



Source: Hausman 1985

Hausman (1981) Application

- Hausman applies method to 1975 PSID cross-section
  - Finds significant compensated elasticities and large income effects
  - Elasticities larger for women than for men
- Shortcomings of this implementation
  1. Sensitivity to functional form choices
  2. No tax reforms, so does not solve fundamental econometric problem that tastes for work may be correlated with w

NLBS and Bunching at Kinks

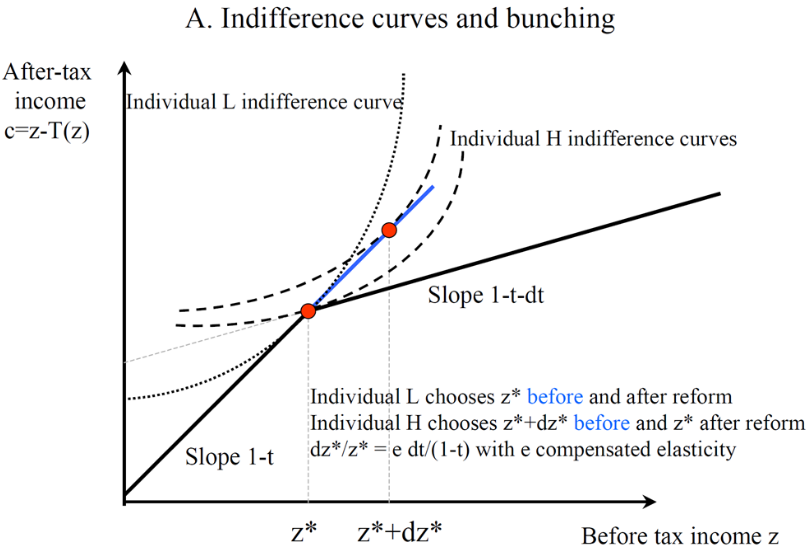
- Subsequent studies obtain different estimates (MaCurdy, Green, and Paarsh 1990, Blomquist 1995)
- Several studies find **negative** compensated wage elasticity estimates
- Debate: impose requirement that compensated elasticity is positive or conclude that data rejects model?
- Fundamental source of problem: labor supply model predicts that individuals should bunch at the kink points of the tax schedule
  - But we observe very little bunching at kinks, so model is rejected by the data
  - Interest in NLBS models diminished despite their conceptual advantages over OLS methods

Saez 2010: Bunching at Kinks

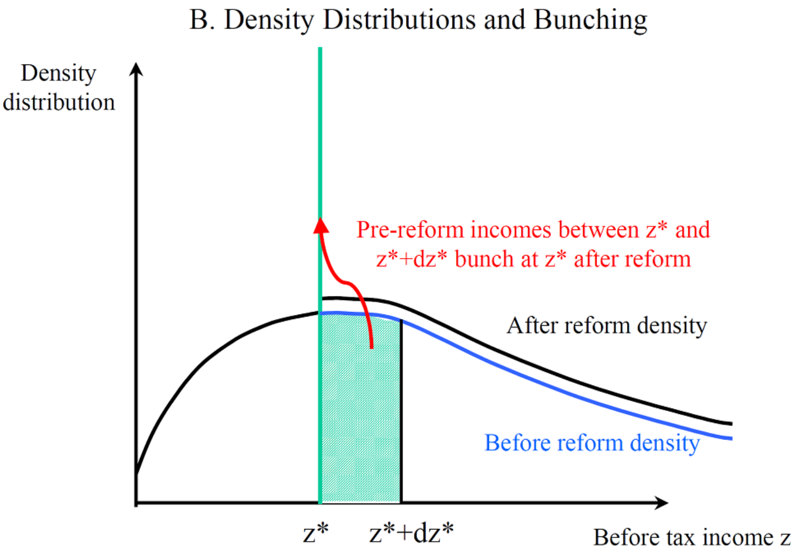
- Saez observes that only non-parametric source of identification for elasticity in a cross-section is amount of bunching at kinks
  - Intuition: discontinuous reduction in wage rate at kink yields source of non-parametric identification
  - All other cross-sectional tax variation is contaminated by smooth heterogeneity in tastes
- Derives an estimator for the compensated taxable income elasticity using amount of bunching at kinks

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

- Currently a popular approach (esp. when adapted to account for frictions) because it yields highly credible estimates



Source: Saez (2009)

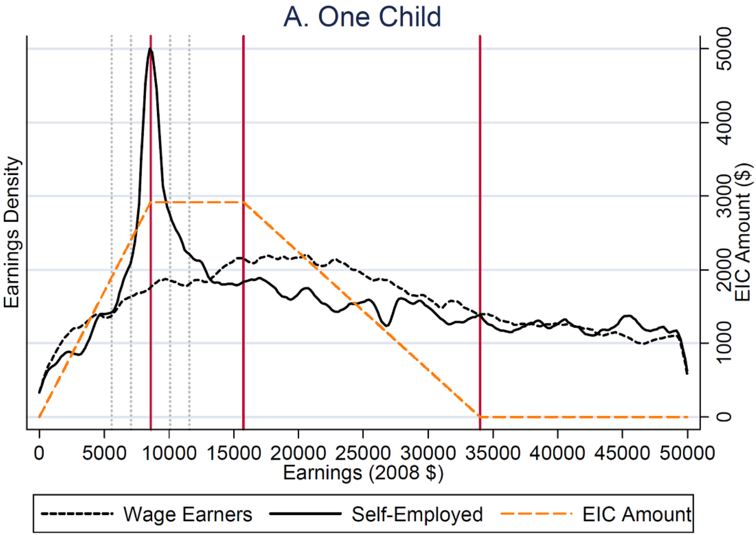


Source: Saez (2009)

Saez 2010: Bunching at Kinks

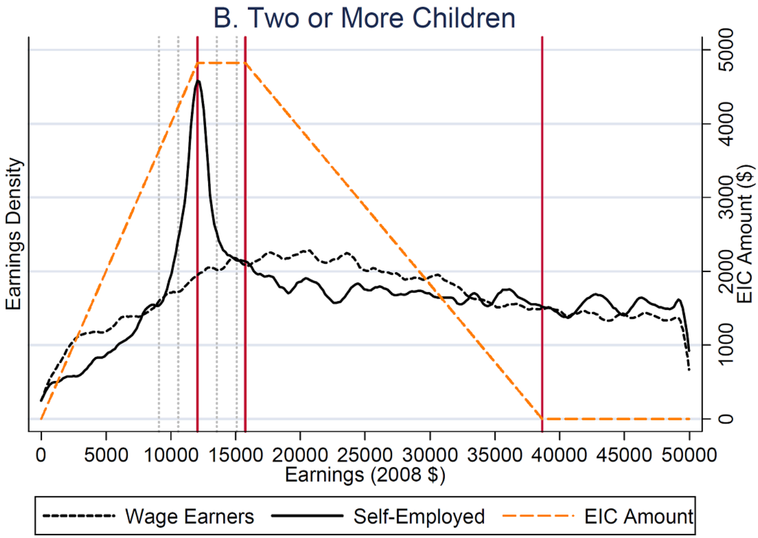
- Saez implements this method using individual tax return micro data (IRS public use files) from 1960 to 2004
- Advantage of dataset over PSID: very little measurement error
- Finds sharp bunching around first kink point of the EITC for self-employed
  - Later shown to be largely due to reporting effects
- However, no bunching observed at any kink for wage earners

Earnings Density and the EITC: Wage Earners vs. Self-Employed



Source: Saez (2009)

Earnings Density and the EITC: Wage Earners vs. Self-Employed



Source: Saez (2009)

6. Limitations & comparison with (natural-) experiments

Limitations of Structural Models

- Structural models made for *ex-ante* evaluation of policy changes to study hypothetical/future reforms for policy advice
- But how much do we trust the models?
  - not all institutions are accounted for (child-care, in-kind transfers)
  - possibly strong assumptions on household decision making process
  - models often static, unitary, supply-side
  - simulated responses rarely validated against data or other approaches (natural experiments)

Limitations of Structural Models

- **Non-hours responses**: income taxes distort many margins beyond hours of work / participation
  - hours very hard to measure (most report 40 hours/week)
  - non-hours margins may be more important quantitatively
- Two solutions in modern literature:
  - focus on subgroups of workers for whom hours are better measured, e.g. taxi drivers
  - focus on elasticity of total earnings (wh) or **taxable income** as a broader measure of labor supply (Saez et al., 2012, JEL)

Comparison with Natural Experiments

Natural experiments:

- frequently used to assess labor supply (or tax base) elasticity
  - esp. Difference-in-difference (DiD) or Regression Discontinuity (RD)
- Identification requires exogenous variation in the data
- Advantage: no need to impose structure / capture variety of 'true' behaviors

Comparison with Natural Experiments

- Issues with DiD estimators:
  - difficult to find convincing control vs treatment group
  - unobserved characteristics may affect both selection into treatment group and measured effects
- Concerns about limited policy relevance:
  - limited external validity of DiD (specific in time and space) and RD (around discontinuity)
  - only certain subset of population
  - Additional data to test the underlying assumptions of DiD design: ‘Placebo’ or ‘falsification’ tests
- Also: often no direct link to welfare measure; no GE effects

Comparison with Field Experiments

- Recent literature advocate **experimental evaluation designs**
- Directly solve natural-experiment problems if true randomization
  - Shortcoming of previous randomized experiments (esp. US Negative Income Tax 1960/1970s, Ashenfelter and Plant, 1990; Moffitt, 2003)
- e.g. *Self-Sufficiency Project (SSP):* large randomized trial of wage subsidies for SA recipients in Canada
  - induced recipients to find and hold a fulltime job
  - influential in welfare-reform in several countries (in-work support in UK, FR, GE, NL,...)

Comparison with Field Experiments

- Huge literature on welfare programmes in developing countries
  - e.g. conditional cash transfer programs like Progresa (Attanasio et al, 2003, Alzúa et al., 2010)
- But still problem of external validity:
  - overgeneralization: impact of treatment on group studied may not be informative about impact on other groups
  - randomization bias: individuals who are willing to participate in trial not representative of those who would articipate in subsequent program
- Hawthorne effect (subjects' behavior in trial may be different from what it would normally be, ex: teachers in the STAR projects)

Validating Structural Models using Experiments

Validation / comparison of structural models and experiments

- Rarely done – need more systematic feedback into structural design
- Need to mix structural model and (natural) experiments
    - structural model captures behavior and policy effect
    - how well does it perform on subgroups (experiment)?
    - behavior captured by structural model while "other" changes captured by control group

Validating Structural Models

- Example: Hansen and Liu (2012)
- Use RD results of Lemieux-Milligan (2008):
  - before 1989, much lower SA for childless recipients in Quebec when under age 30
  - sharp discontinuity reduces employment after 30 (next slide)
  - RD results: employment drops by 4.9 ppt (7.9%) for target group (high school dropouts)

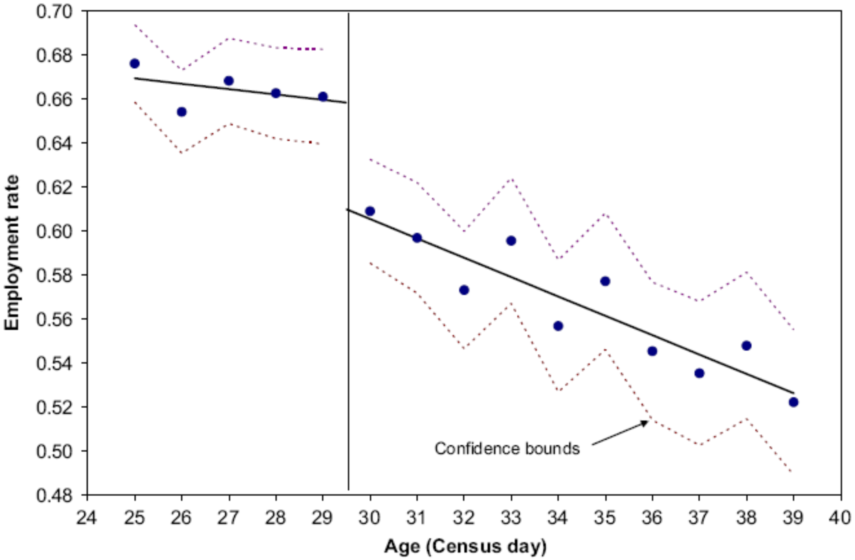


Fig. 3. Employment rate in Census week, Quebec 1986.

Validating Structural Models

- Example: Hansen and Liu (2012) based on RD results of Lemieux-Milligan (2008):
  - employment drops by 4.9 ppt (7.9%) for high school dropouts
- Compare RD to prediction of structural model:
  - For high school dropouts: model predicts reduction in employment due to SA of 4.3 ppt (6.7%)
  - close to RD results of 4.9 ppt (7.9%)
- results relatively robust to specification of LS model
  - if the case, encouraging that structural model performs well esp. given large change in policy environment