Public Economics Lectures
Part 1: Introduction

Raj Chetty and Gregory A. Bruich

Harvard University
Fall 2009

What is Public Economics?

- Public economics focuses on answering two types of questions
  1. How do government policies affect the economy?
  2. How should policies be designed to maximize welfare?
- Three motivations for studying these questions:
  1. Practical Relevance
  2. Academic Interest
  3. Methodology

Motivation 1: Practical Relevance

- Interest in improving economic welfare $\rightarrow$ interest in public economics
- Almost every economic intervention occurs through government policy (i.e. involves public economics) via two channels:
  - Price intervention: taxes, welfare, social insurance, public goods
  - Regulation: min wages, FDA regulations (25% of products consumed), zoning laws, labor laws, min education laws, environment, legal code
- Government directly employs one sixth of U.S. workforce

Motivation 1: Practical Relevance

- Stakes are extremely large because of broad scope of policies
  - Ex. Tax reforms immediately affect millions
- Contentious debate on the appropriate role of government in society
  - Controversial: liberals vs. conservatives
  - McCain: "Obama proposes higher taxes. Therefore 5M fewer jobs"
  - Obama: "Alternative energy investment will create 5M U.S. jobs"
- Which view is right? Injecting science into these political debates has tremendous practical value

Motivation 2: Academic Interest

- Public economics is typically the end point for many other subfields of economics
- Macro, development, labor, and corporate finance questions often ultimately motivated by a public economics issue
  - Ex 1: Macro studies on costs of business cycles and intertemporal models of household behavior
  - Ex 2: Labor studies on employment effects of the minimum wage
- Natural to combine public finance with another field
- Understanding public finance can help sharpen your research focus and ensures you are working on relevant issues

Motivation 3: Methodology

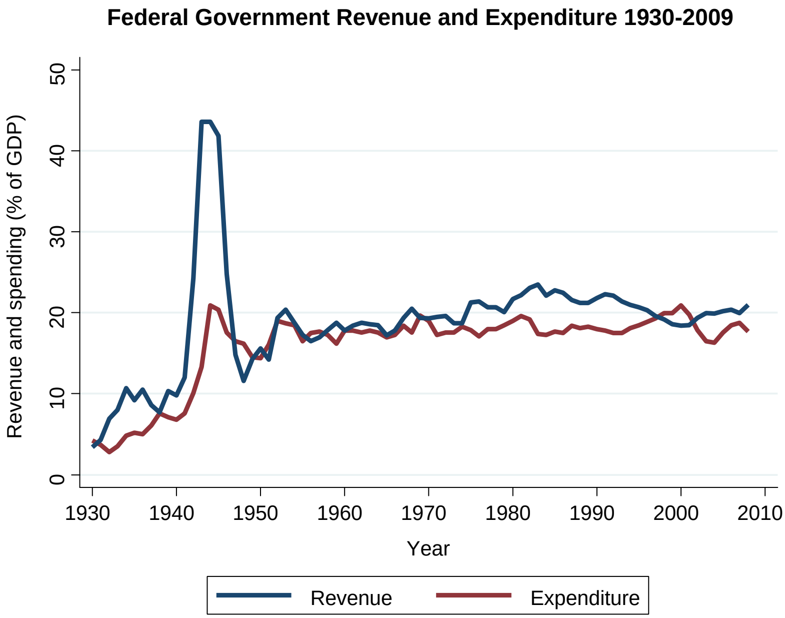
- Modern public economics tightly integrates theory with empirical evidence to derive quantitative predictions about policy
  - What is the optimal income tax rate?
  - What is the optimal unemployment benefit level?
- Combining applied theory and evidence is a useful skill set that is at the frontier of many fields of economics

Methodological Themes

1. Micro-based empirics but both micro and macro theory
2. Two styles of work: structural and reduced-form
3. Neoclassical, but growing interest in implications of behavioral econ for public policy
4. Focus primarily on developed countries because of data availability, but growing interest in developing countries
5. Long run focus in theory: focus on ideal design of systems for long run welfare but short-run focus in empirics
6. Two approaches to research: bringing in new ideas from other fields vs. innovating within public economics

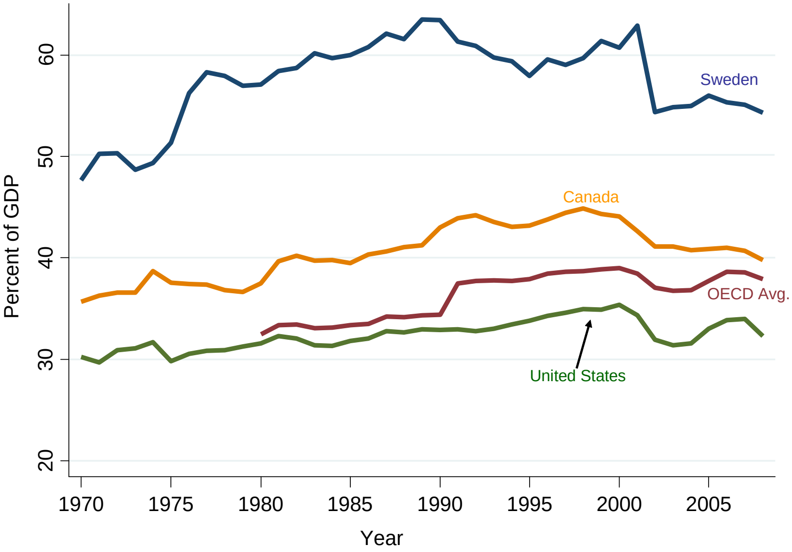
Background Facts: Size and Growth of Government

- Government expenditures = 1/3 GDP in the U.S.
- Higher than 50% of GDP in some European countries
- Decentralization is a key feature of U.S. govt
    - One third of spending (10% of GDP) is done at state-local level (e.g. schools)
    - Two thirds (20% of GDP) is federal



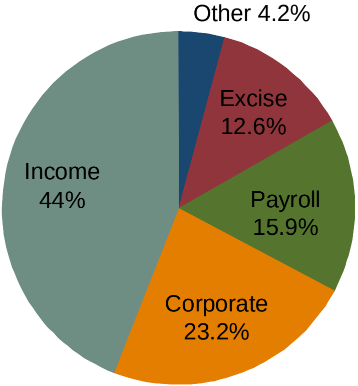
Source: Office of Management and Budget, Historical Tables, FY 2011

Total Government Spending by Country, 1970-2007

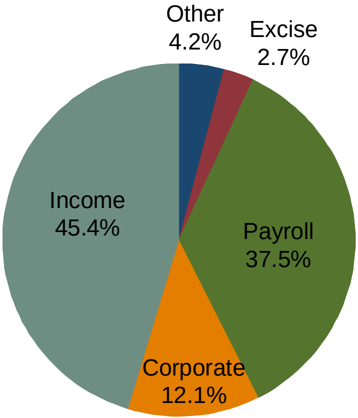


Source: OECD Economic Outlook (2009)

Federal Revenues (% of total revenue)



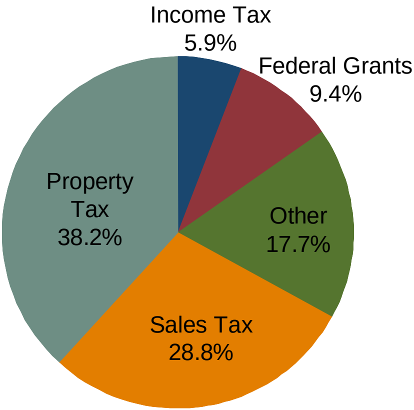
1960



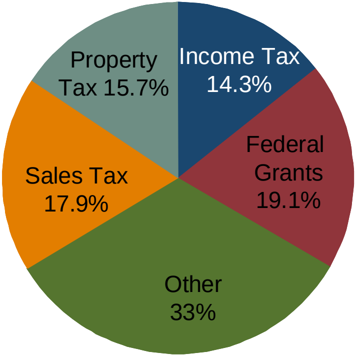
2008

Source: Office of Management and Budget, historical tables, government receipts by source

State/Local Revenues (% of total revenue)



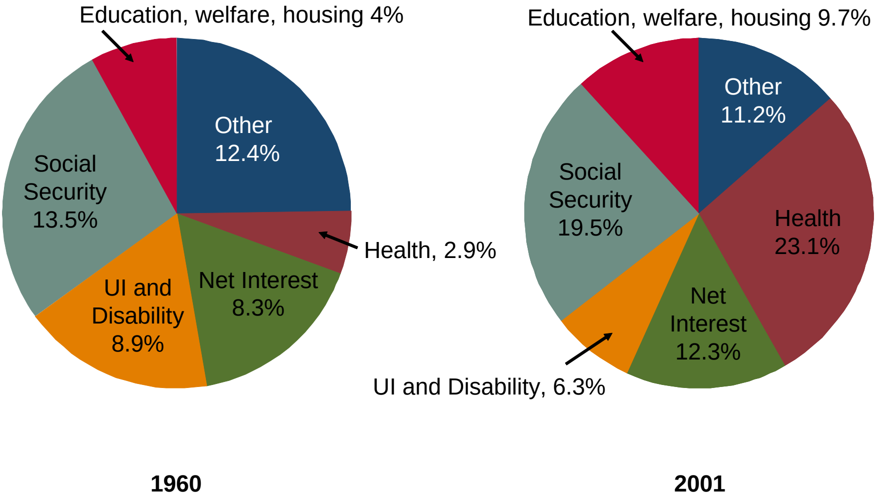
1960



2007

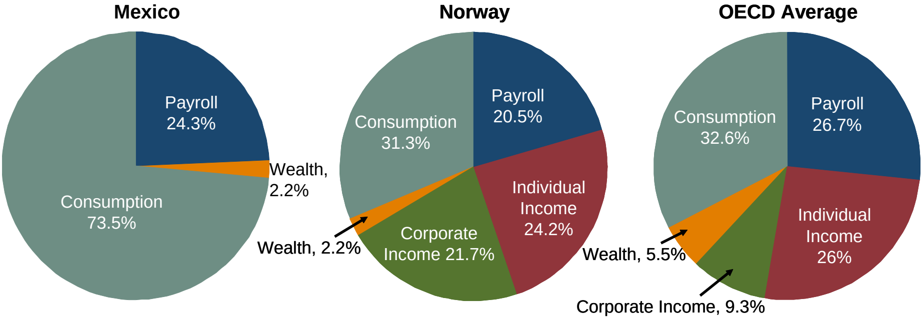
Source: U.S. Census Bureau, 2007 Summary of State & Local Government

Federal Spending (% of total spending)



Source: Office of Management and Budget, historical tables, government outlays by function

International Tax Revenue by Type of Tax (2001, % of Total)

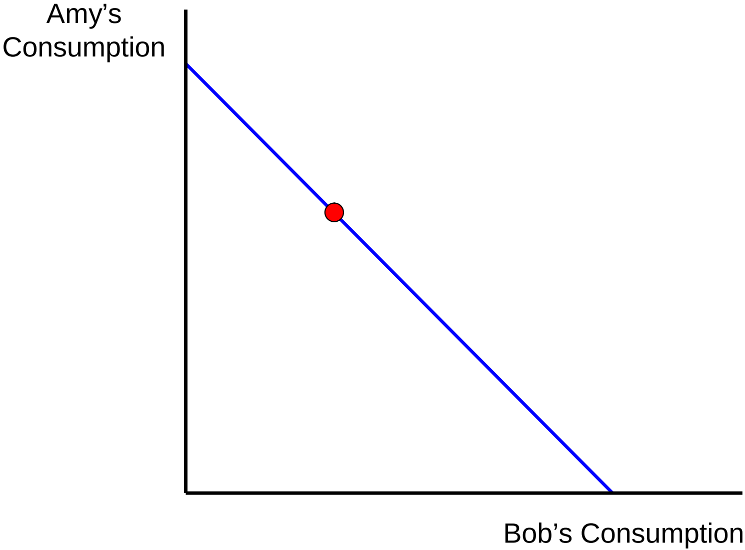


Source: OECD 2002

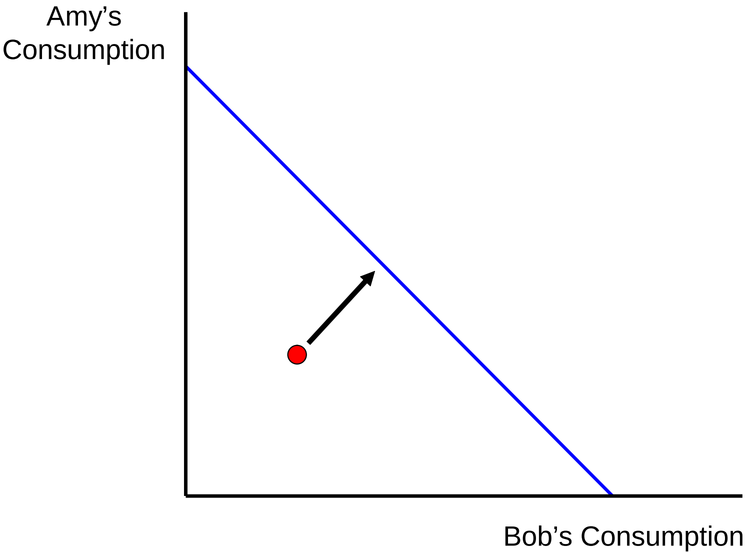
Government Intervention in the Economy

- Organzing framework: "When is government intervention necessary in a market economy?"
  - Market first, govt. second approach
  - Why? Private market outcome is efficient under broad set of conditions (1st Welfare Thm)
- Course can be split into two parts:
  1. How can govt. improve efficiency when private market is inefficient?
  2. What can govt. do if private market outcome is undesirable due to redistributional concerns?

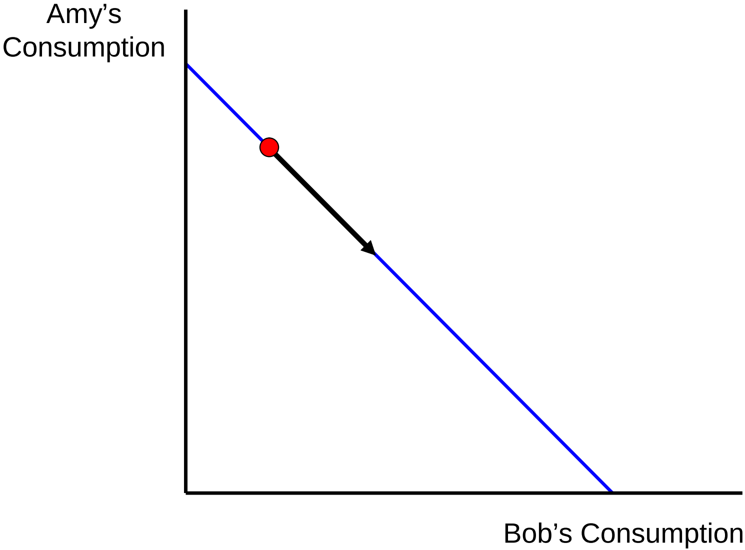
Efficient Private Market Allocation of Goods



First Role for Government: Improve Efficiency



Second Role for Government: Improve Distribution



First Welfare Theorem

- Private market provides a Pareto efficient outcome under three conditions

  1. No externalities
  2. Perfect information
  3. Perfect competition

- Theorem tells us when the government should intervene

Failure 1: Externalities

- Markets may be incomplete due to lack of prices (e.g. pollution)
  - Achieving efficient Coasian solution requires an organization to coordinate individuals – that is, a government
- This is why govt. funds public goods (highways, education, defense)
- Questions: What public goods to provide and how to correct externalities?

Failure 2: Asymmetric Information and Incomplete Markets

- When some agents have more information than others, markets fail
- Ex. 1: Adverse selection in health insurance
  - Healthy people drop out of private market $\longrightarrow$ unraveling
  - Mandated coverage could make everyone better off
- Ex. 2: capital markets (credit constraints) and subsidies for education
- Ex. 3: Markets for intergenerational goods
  - Future generations' interests may not be fully reflected in market outcomes

Failure 3: Imperfect Competition

- When markets are not competitive, there is role for govt. regulation
  - Ex: natural monopolies such as electricity and telephones
- This topic is traditionally left to courses on industrial organization and is not covered in this course
- But taking the methodological approach of public economics to problems traditionally analyzed in IO is a very promising area

Individual Failures

- Recent addition to the list of potential failures that motivate government intervention: people are not fully rational
- Government intervention (e.g. by forcing saving via social security) may be desirable
- This is an "individual" failure rather than a traditional market failure
- Conceptual challenge: how to avoid paternalism critique
  - Why does govt. know better what's desirable for you (e.g. wearing a seatbelt, not smoking, saving more)
- Difficult but central issues to policy design

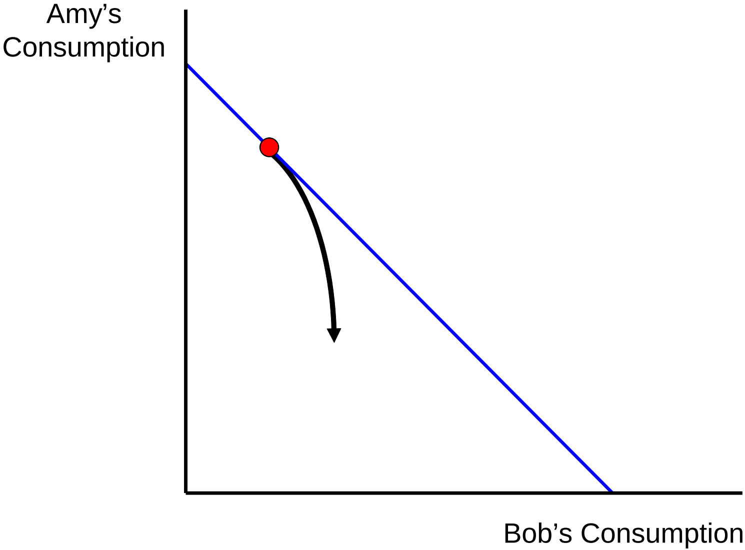
Redistributional Concerns

- Even when the private market outcome is efficient, may not have good distributional properties
- Efficient markets generally seem to deliver very large rewards to small set of people (top incomes)
- Government can intervene to redistribute income through tax and transfer system

Why Limit Government Intervention?

- One solution to these issues would be for the government to oversee all production and allocation in the economy (socialism).
- Serious problems with this solution
  1. Information: how does government aggregate preferences and technology to choose optimal production and allocation?
  2. Government policies inherently distort incentives (behavioral responses in private sector)
  3. Politicians not necessarily a benevolent planner in reality; face incentive constraints themselves
- Creates sharp tradeoffs between costs and benefits of government intervention
  - Providing more public goods requires higher taxes and distorts consumption decisions
  - Redistribution distorts incentives to work

Equity-Efficiency Tradeoff



Three Types of Questions in Public Economics

1. Positive analysis: What are the observed effects of government programs and interventions?

2. Normative analysis: What should the government do if we can choose optimal policy?

3. Public choice/Political Economy

    - Develops theories to explain why the government behaves the way it does and identify optimal policy given political economy concerns

    - Criticism of normative analysis: fails to take political constraints into account

Course Outline

1. Tax Incidence and Efficiency
2. Optimal Taxation
3. Income Taxation and Labor Supply
4. Social Insurance
5. Public Goods and Externalities

Public Economics Lectures
Part 2: Incidence of Taxation

Raj Chetty and Gregory A. Bruich

Harvard University
Fall 2009

Outline

References on Tax Incidence

- Kotlikoff and Summers (1987) handbook chapter
- Atkinson and Stiglitz text chapters 6 and 7
- Chetty, Looney, and Kroft (2009)

Definition

- Tax incidence is the study of the effects of tax policies on prices and the distribution of utilities
- What happens to market prices when a tax is introduced or changed?
  - Increase tax on cigarettes by $1 per pack
  - Introduction of Earned Income Tax Credit (EITC)
  - Food stamps program
- Effect on price $\rightarrow$ distributional effects on smokers, profits of producers, shareholders, farmers, ...

Economic vs. Statutory Incidence

- Equivalent when prices are constant but **not** in general
- Consider the following argument:
  - Government should tax capital income b/c it is concentrated at the high end of the income distribution
- Neglects general equilibrium price effects
  - Tax might be shifted onto workers
  - If capital taxes $\longrightarrow$ less savings and capital flight, then capital stock may decline, driving return to capital up and wages down
  - Some argue that capital taxes are paid by workers and therefore *increase* income inequality (Hassett and Mathur 2009)

Overview of Literature

- Tax incidence is an example of positive analysis
  - Typically the first step in policy evaluation
  - An input into thinking about policies that maximize social welfare
- Theory is informative about signs and comparative statics but is inconclusive about magnitudes
  - Incidence of cigarette tax: elasticity of demand w.r.t. price is crucial
  - Labor vs. capital taxation: mobility of labor, capital are critical

Overview of Literature

- Ideally, we would characterize the effect of a tax change on utility levels of all agents in the economy
- Useful simplification in practice: aggregate economic agents into a few groups
- Incidence analyzed at a number of levels:

  1. Producer vs. consumer (tax on cigarettes)
  2. Source of income (labor vs. capital)
  3. Income level (rich vs. poor)
  4. Region or country (local property taxes)
  5. Across generations (social Security reform)

Partial Equilibrium Incidence: Key Assumptions

1. **Two good economy**
   - Only one relative price → partial and general equilibrium are same
   - Can be viewed as an approx. of incidence in a multi-good model if
     - the market being taxed is "small"
     - there are no close substitutes/complements in the utility fn
2. **Tax revenue is not spent on the taxed good**
   - Tax revenue is used to buy untaxed good or thrown away
3. **Perfect competition among producers**
   - Relaxed in some studies of monopolistic or oligopolistic markets

Partial Equilibrium Model: Setup

- Two goods: x and y
- Government levies an **excise** tax on good x
  - **Excise/Specific tax**: levied on a quantity (e.g. gallon, pack, ton)
  - **Ad-valorem tax**: fraction of prices (e.g. sales tax)
- Let p denote the pretax price of x and $q = p + t$ denote the tax inclusive price of x
- Good y, the numeraire, is untaxed

Partial Equilibrium Model: Demand

- Consumer has wealth Z and has utility $u(x, y)$
- Let $\varepsilon_D = \frac{\partial D}{\partial p} \frac{q}{D(p)}$ denote the price elasticity of demand
  - **Elasticity**: % change in quantity when price changes by 1%
  - Widely used concept because elasticities are unit free

Partial Equilibrium Model: Supply

- Price-taking firms
- Use $c(S)$ units of the numeraire y to produce S units of x
- Cost of production is increasing and convex:

$$c'(S) > 0 \text{ and } c''(S) \geq 0$$

- Profit at pretax price p and level of supply S is $pS - c(S)$
- With perfect optimization, the supply function for good x is implicitly defined by the marginal condition $p = c'(S(p))$
- Let $\varepsilon_S = \frac{\partial S}{\partial p} \frac{p}{S(p)}$ denote the price elasticity of supply

Partial Equilibrium Model: Equilibrium

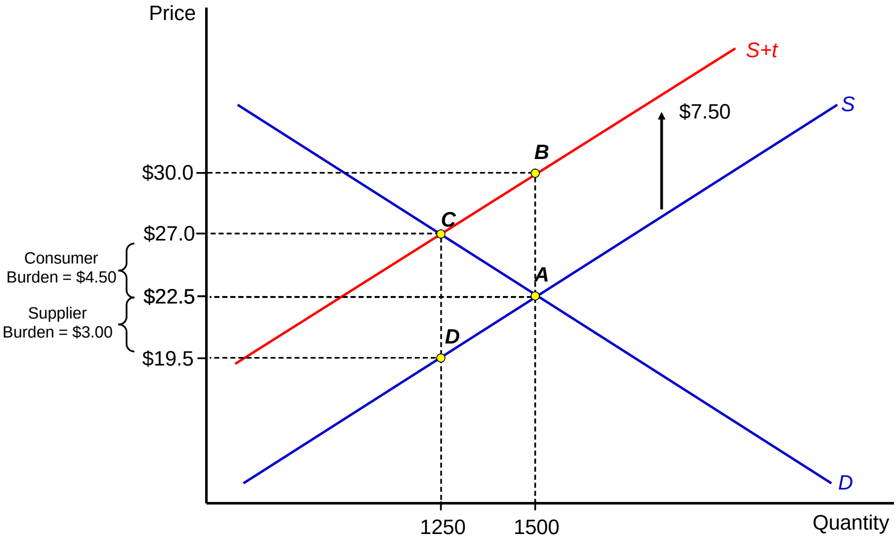
- Equilibrium condition

$$Q = S(p) = D(p + t)$$

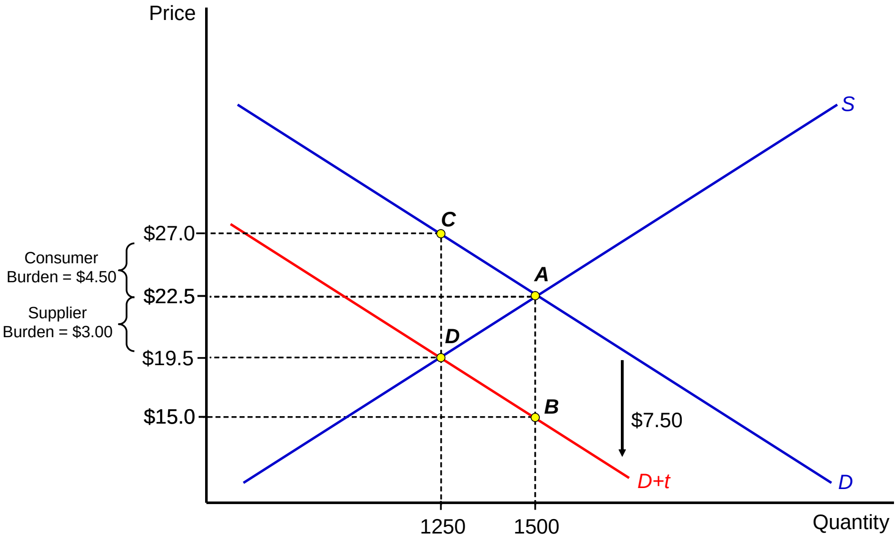
  defines an equation $p(t)$

- Goal: characterize $\frac{dp}{dt}$, the effect of a tax increase on price
- First consider some graphical examples to build intuition, then analytically derive formula

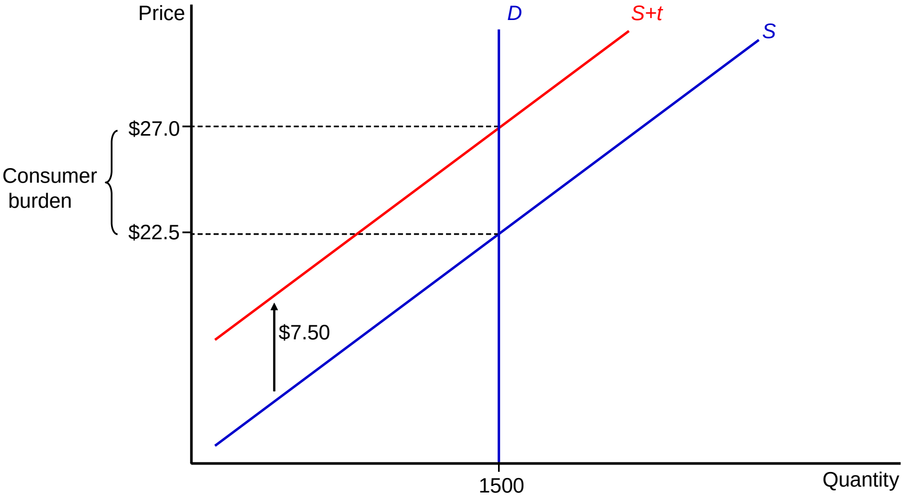
Tax Levied on Producers



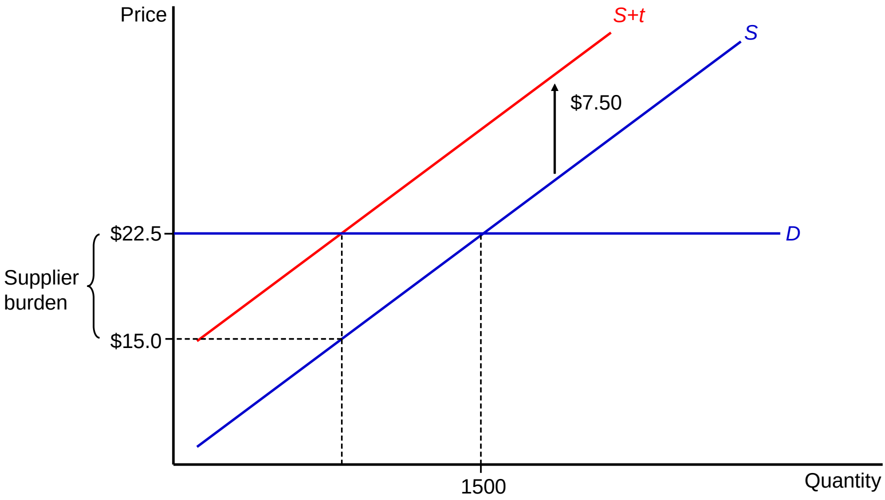
Tax Levied on Consumers



Perfectly Inelastic Demand



Perfectly Elastic Demand



Formula for Tax Incidence

- Implicitly differentiate equilibrium condition

$$D(p+t) = S(p)$$

to obtain:

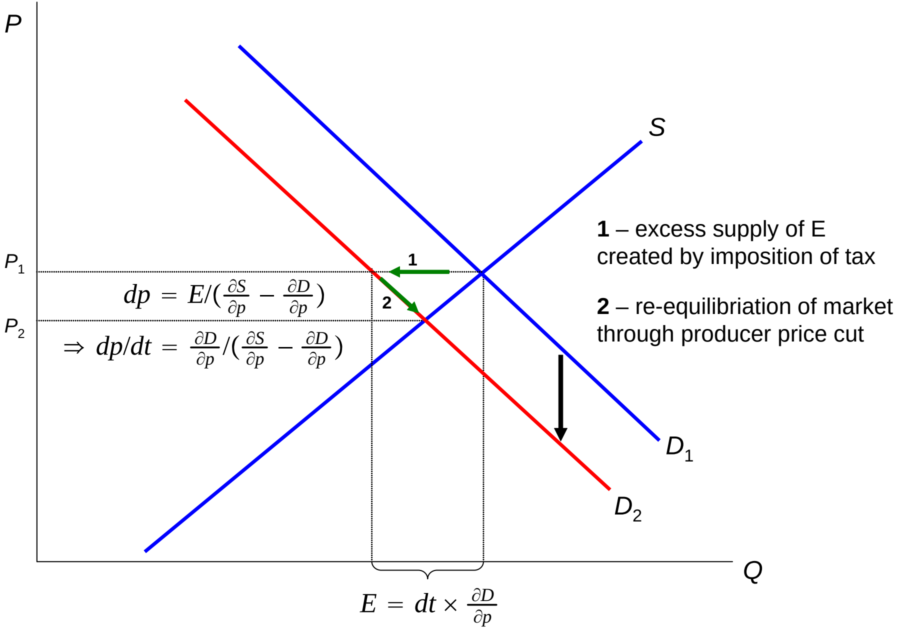
$$\frac{dp}{dt} = \frac{\partial D}{\partial p} \frac{1}{\left(\frac{\partial S}{\partial p} - \frac{\partial D}{\partial p}\right)}$$

$$\Rightarrow \frac{dp}{dt} = \frac{\varepsilon_D}{\varepsilon_S - \varepsilon_D}$$

- Incidence on consumers:

$$\frac{dq}{dt} = 1 + \frac{dp}{dt} = \frac{\varepsilon_S}{\varepsilon_S - \varepsilon_D}$$

Formula for Tax Incidence



Tax Incidence with Salience Effects

- Central assumption of neoclassical model: taxes are equivalent to prices ($\frac{dx}{dt} = \frac{dx}{dp}$)
- In practice, are people fully aware of marginal tax rates?
- Chetty, Looney, and Kroft (2009) test this assumption and generalize theory to allow for salience effects
- **Part 1:** Test whether "salience" (visibility of tax-inclusive price) affects behavioral responses to commodity taxation
  - Does effect of a tax on demand depend on whether it is included in **posted** price?
- **Part 2:** Develop formulas for incidence and efficiency costs of taxation that permit salience effects and other optimization errors

Chetty et al.: Empirical Framework

- Economy with two goods, x and y
- Prices: Normalize the price of y to 1 and let p denote the (fixed) pretax price of x.
- Taxes: y untaxed, x subject to an ad valorem sales tax τ (not included in posted price)
    - Tax-inclusive price of x is $q = p(1 + \tau)$.
- Let demand for good x be denoted by $x(p, \tau)$

Chetty et al.: Empirical Framework

- If agents optimize fully, demand should only depend on the total tax-inclusive price: $x(p,\tau) = x((1+\tau)p, 0)$

- Full optimization implies price elasticity equals gross-of-tax elasticity:

$$\varepsilon_{x,p} \equiv -\frac{\partial \log x}{\partial \log p} = \varepsilon_{x,1+\tau^S} \equiv -\frac{\partial \log x}{\partial \log(1+\tau)}$$

- To test this hypothesis, log-linearize demand fn. $x(p,\tau)$ to obtain estimating equation:

$$\log x(p,\tau) = \alpha + \beta \log p + \theta\beta \log(1+\tau)$$

- θ measures degree to which agents under-react to the tax:

$$\theta = \frac{\partial \log x}{\partial \log(1+\tau)} \Big/ \frac{\partial \log x}{\partial \log p} = \frac{\varepsilon_{x,1+\tau}}{\varepsilon_{x,p}}$$

Chetty et al.: Two Empirical Strategies

Two strategies to estimate θ:

1. **Manipulate tax salience:** make sales tax as visible as pre-tax price
   - Effect of intervention on demand:
   $$v = \log x((1+\tau)p, 0) - \log x(p, \tau)$$
   - Compare to effect of equivalent price increase to estimate θ:
   $$(1-\theta) = -\frac{v}{\varepsilon_{x,p}\log(1+\tau)}$$
2. **Manipulate tax rate:** compare $\varepsilon_{x,p}$ and $\varepsilon_{x,1+\tau}$
   $$\theta = \varepsilon_{x,1+\tau}/\varepsilon_{x,p}$$

Chetty et al.: Strategy 1

- Experiment manipulating salience of sales tax implemented at a supermarket that belongs to a major grocery chain
  - 30% of products sold in store are subject to sales tax
  - Posted tax-inclusive prices on shelf for subset of products subject to sales tax (7.375% in this city)
- Data: Scanner data on price and weekly quantity sold by product

Source: Chetty, Looney, and Kroft (2009)

TABLE 1

Evaluation of Tags: Classroom Survey

	Mean	Median	SD
Original Price Tags:			
Correct tax-inclusive price w/in $0.25	0.18	0.00	0.39
Experimental Price Tags:			
Correct tax-inclusive price w/in $0.25	0.75	1.00	0.43
T-test for equality of means: $p < 0.001$			
N=49			

Students were asked to choose two items from image.

Asked to report “Total bill due at the register for these two items.”

Source: Chetty, Looney, and Kroft (2009)

Chetty et al.: Research Design

- Quasi-experimental difference-in-differences
- Treatment group:
  - *Products*: Cosmetics, Deodorants, and Hair Care Accessories
  - *Store*: One large store in Northern California
  - *Time period*: 3 weeks (February 22, 2006 – March 15, 2006)
- Control groups:
  - *Products*: Other prods. in same aisle (toothpaste, skin care, shave)
  - *Stores*: Two nearby stores similar in demographic characteristics
  - *Time period*: Calendar year 2005 and first 6 weeks of 2006

Effect of Posting Tax-Inclusive Prices: Mean Quantity Sold

	TREATMENT STORE		
Period	Control Categories	Treated Categories	Difference
Baseline	26.48 (0.22)	25.17 (0.37)	-1.31 (0.43)
Experiment	27.32 (0.87)	23.87 (1.02)	-3.45 (0.64)
Difference over time	0.84 (0.75)	-1.30 (0.92)	**DD_{TS} = -2.14** (0.64)

Source: Chetty, Looney, and Kroft (2009)

Effect of Posting Tax-Inclusive Prices: Mean Quantity Sold

TREATMENT STORE

Period	Control Categories	Treated Categories	Difference
Baseline	26.48	25.17	-1.31
	(0.22)	(0.37)	(0.43)
Experiment	27.32	23.87	-3.45
	(0.87)	(1.02)	(0.64)
Difference over time	0.84	-1.30	**DD_{TS} = -2.14**
	(0.75)	(0.92)	(0.64)

CONTROL STORES

Period	Control Categories	Treated Categories	Difference
Baseline	30.57	27.94	-2.63
	(0.24)	(0.30)	(0.32)
Experiment	30.76	28.19	-2.57
	(0.72)	(1.06)	(1.09)
Difference over time	0.19	0.25	**DD_{CS} = 0.06**
	(0.64)	(0.92)	(0.90)

Source: Chetty, Looney, and Kroft (2009)

Effect of Posting Tax-Inclusive Prices: Mean Quantity Sold

TREATMENT STORE

Period	Control Categories	Treated Categories	Difference
Baseline	26.48	25.17	-1.31
	(0.22)	(0.37)	(0.43)
Experiment	27.32	23.87	-3.45
	(0.87)	(1.02)	(0.64)
Difference over time	0.84	-1.30	**DD_{TS} = -2.14**
	(0.75)	(0.92)	(0.64)

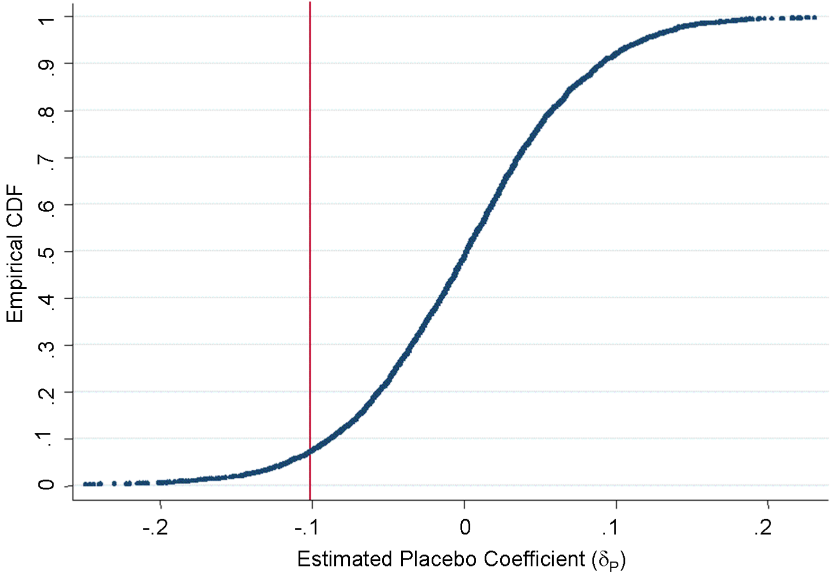
CONTROL STORES

Period	Control Categories	Treated Categories	Difference
Baseline	30.57	27.94	-2.63
	(0.24)	(0.30)	(0.32)
Experiment	30.76	28.19	-2.57
	(0.72)	(1.06)	(1.09)
Difference over time	0.19	0.25	**DD_{CS} = 0.06**
	(0.64)	(0.92)	(0.90)
		DDD Estimate	***-2.20***
			(0.58)

Source: Chetty, Looney, and Kroft (2009)

Figure 1

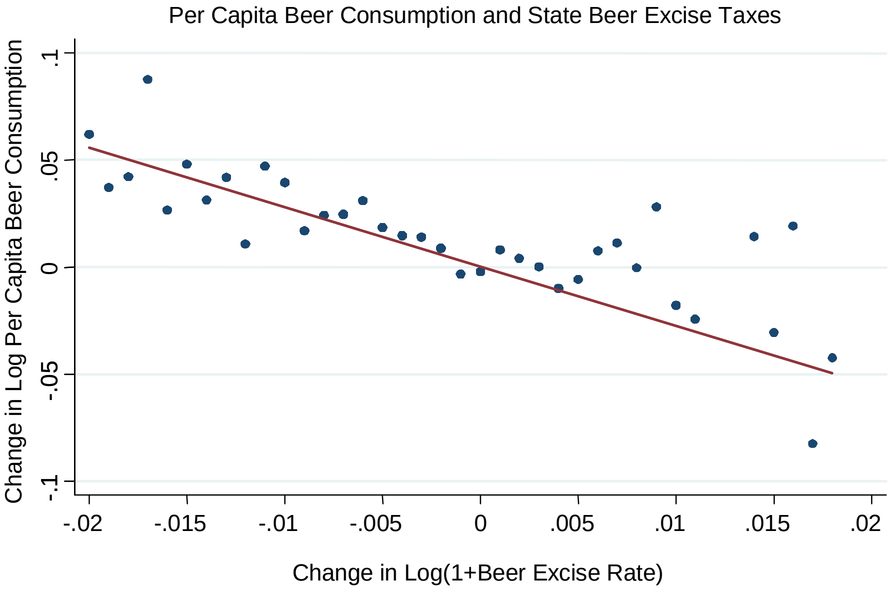
Distribution of Placebo Estimates: Log Quantity



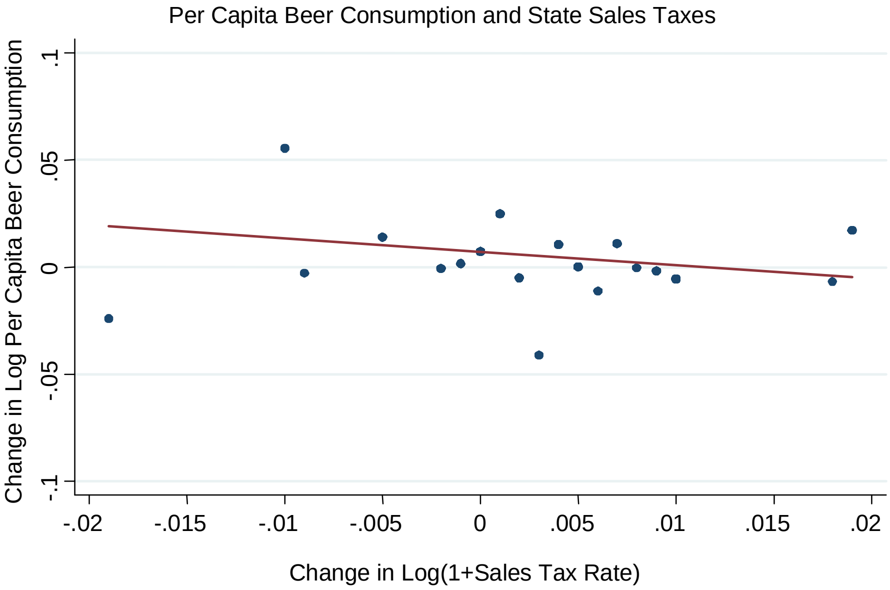
Source: Chetty, Looney, and Kroft (2009)

Chetty et al.: Strategy 2

- Compare effects of price changes and tax changes
- Alcohol subject to two state-level taxes in the U.S.:
  - Excise tax: included in price
  - Sales tax: added at register, not shown in posted price
- Exploiting state-level changes in these two taxes, estimate θ
  - Addresses concern that experiment may have induced a "Hawthorne effect"



Source: Chetty, Looney, and Kroft (2009)



Source: Chetty, Looney, and Kroft (2009)

Effect of Excise and Sales Taxes on Beer Consumption

Dependent Variable: ΔLog(per capita beer consumption)

	Baseline	Bus Cyc, Alc Regs.	3-Year Diffs	Food Exempt
	(1)	(2)	(3)	(4)
ΔLog(1+Excise Tax Rate)	**-0.87**	**-0.89**	**-1.11**	**-0.91**
	(0.17)***	(0.17)***	(0.46)**	(0.22)***
ΔLog(1+Sales Tax Rate)	**-0.20**	**-0.02**	**-0.00**	**-0.14**
	(0.30)	(0.30)	(0.32)	(0.30)
Business Cycle Controls		x	x	x
Alcohol Regulation Controls		x	x	x
Year Fixed Effects	x	x	x	x
F-Test for Equality of Coeffs.	0.05	0.01	0.05	0.04
Sample Size	1,607	1,487	1,389	937

Source: Chetty, Looney, and Kroft (2009)

Tax Incidence with Salience Effects

- Let $\{x(p,t,Z), y(p,t,Z)\}$ denote empirically observed demands

- Place no structure on these demand functions except for feasibility:

$$(p+t)x(p,t,Z) + y(p,t,Z) = Z$$

- Demand functions taken as empirically estimated objects rather than optimized demand from utility maximization

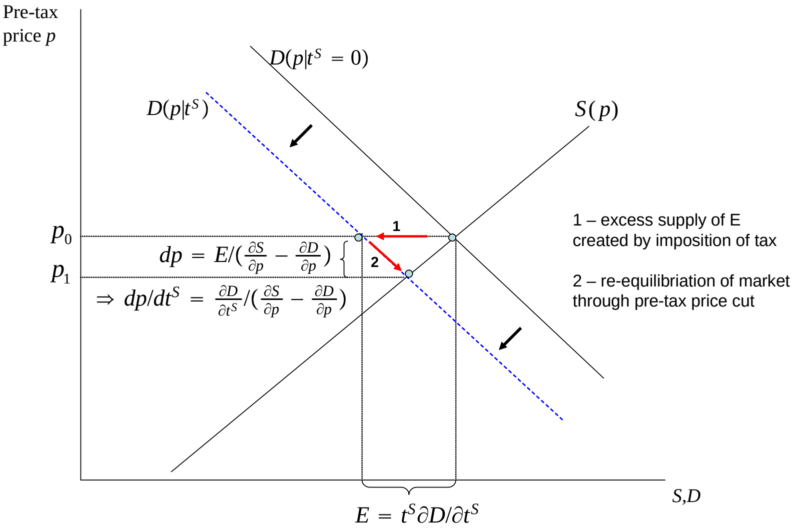
- Supply side model same as above

- Market clearing price p satisfies

$$D(p,t,Z) = S(p)$$

where $D(p,t,z) = x(p,t,z)$ is market demand for x.

Tax Incidence with Salience Effects



Source: Chetty, Looney, and Kroft (2009)

Tax Incidence with Salience Effects: Formula

- Incidence on producers of increasing t is

$$\frac{dp}{dt} = \frac{\partial D/\partial t}{\partial S/\partial p - \partial D/\partial p} = -\theta \frac{\varepsilon_D}{\varepsilon_S - \varepsilon_D}$$

1. Incidence on producers attenuated by θ

2. No tax neutrality: taxes on producers have greater incidence on producers than non-salient taxes levied on consumers

- Intuition: Producers need to cut pretax price less when consumers are less responsive to tax

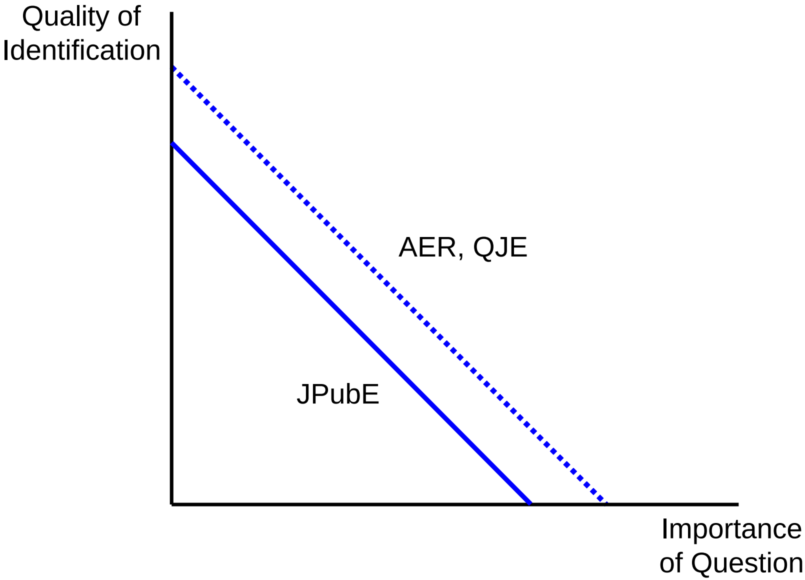
Empirical Applications

1. [Evans, Ringel, and Stech 1999]: Cigarette excise taxes
2. [Hastings and Washington 2008]: Food stamps
3. [Rothstein 2008]: Earned Income Tax Credit

Evaluating Empirical Studies

- Consider ideal experimental design first
- Then formulate a feasible design and analyze its flaws relative to ideal design
- Frontier for empirical papers: often face a trade-off between identification vs. importance/impact

Frontier of Empirical Research



Cigarette Taxation: Background

- Cigarettes are heavily taxed in many countries
- Generates around $15 billion in tax revenue in US, about as much as estate taxation
- Taxed at both federal and state levels
- Federal tax is $0.24 per pack with $7.3 billion raised in 1996
- Each state also applies specific excise taxes
- Variation among states: from 2.5 cents per pack in VA to $1.00 in AK

Cigarette Taxation: Background

- Since 1975, close to 200 state tax changes $\longrightarrow$ natural experiments to investigate tax incidence
- Note that over the last 50 years, many increases in taxes but real tax flat because of inflation erosion
- Controversial commodity due to health and paternalism concerns
- Policy question: How do tax increases affect prices? Do they take money from cigarette companies?
- Partial equilibrium is a plausible approximation for cigarettes $\longrightarrow$ good example with which to start

Evans, Ringel, and Stech (1999)

- Exploit state-level changes in excise tax rates to characterize aggregate market for cigarettes (prices, quantities)
- Provides a good introduction to standard diff-in-diff methods
- Idea: Suppose federal govt. implements a tax change. Compare cigarette prices before and after the change

$$D = [P_{A1} - P_{A0}]$$

- Underlying assumption: absent the tax change, there would have been no change in cigarette price.

Difference-in-Difference

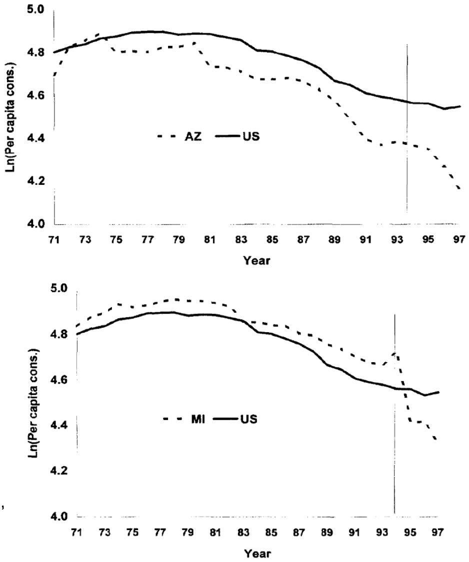
- But what if price fluctuates because of climatic conditions, or if there is an independent trend in demand?

  $\rightarrow$First difference (and time series) estimate biased

- Can improve on the difference by using diff-in-diff

$$DD = [P_{A1} - P_{A0}] - [P_{B1} - P_{B0}]$$

- State A: experienced a tax change (treatment)
- State B: does not experience any tax change (control)
- Identifying assumption: "parallel trends:" absent the policy change, $P_1 - P_0$ would have been the same for A and B



Source: Evans, Ringel, and Stech 1999

Parallel Trend Assumption

- Can use placebo DD to test parallel trend assumption
- Compute DD for prior periods$\longrightarrow$if not zero, then $DD_{t=1}$ prob. biased
  - Useful to plot long time series of outcomes for treatment and control
  - Pattern should be parallel lines, with sudden change just after reform
- Want treat. and cntrl. as similar as possible
- Can formalize this logic using a permutation test: pretend reform occurred at other points and replicate estimate

Triple Difference

- Some studies use a "triple difference" (DDD)
- Chetty, Looney, Kroft (2009): experiment using treatment/control products, treatment/control stores

$$DDD = DD_{TS} - DD_{CS}$$

- DD_{TS}: difference of treat., cntrl products in treat. store
- DD_{CS}: difference of treat., cntrl. products in cntrl. store
- DDD is mainly useful only as a robustness check:
  - $DD_{CS} \neq 0$, unconvincing that DDD removes all bias
  - $DD_{CS} = 0$, then $DD = DDD$ but DD has smaller s.e.

Fixed Effects

- ERS have data for 50 states, 30 years, and many tax changes
- Want to pool all this data to obtain single incidence estimate
- Fixed effects: generalize DD with $S > 2$ periods and $J > 2$ groups
- Suppose that group j in year t experiences policy T of intensity T_{jt}
- Want to identify effect of T on price P. OLS regression:

$$P_{jt} = \alpha + \beta T_{jt} + \epsilon_{jt}$$

- With no fixed effects, the estimate of β is biased if treatment T_{jt} is correlated with ϵ_{jt}
    - Often the case in practice - states with taxes differ in many ways (e.g. more anti-tobacco campaigns)

Fixed Effects

- Include time and state dummies as a way of solving this problem:

$$P_{jt} = \alpha + \gamma_t + \delta_j + \beta T_{jt} + \epsilon_{jt}$$

- Fixed effect regression is equivalent to partial regression

$$\hat{P}_{jt} = \beta \hat{T}_{jt} + \epsilon_{jt}$$

where $\hat{P}_{jt} = P_{jt} - P_j - P_t$ and $\widehat{T}_{jt}$ is defined analogously

- Identification obtained from within-state variation over time

- Note: common changes that apply to all groups (e.g. fed tax change) captured by time dummy; not a source of variation that identifies β

Fixed Effects vs. Difference-in-Difference

- Advantage relative to *DD*: more precise, robust results
- Disadvantage: fixed effects is a black-box regression, more difficult to check trends visually as can be done with a single change

  $\longrightarrow$ Combine it with simple, graphical, non-parametric evidence

- Same parallel trends identification assumption as DD
  - Potential violation: policy reforms may respond to trends in outcomes
  - Ex: tobacco prices increase $\longrightarrow$ state decides to lower tax rate

Evans, Ringel, and Stech (1999)

- Implement a fixed effects model for prices
- Regress price on state/year fixed effects, covariates, and tax rate (in cents)
- Also estimate demand elasticities using fixed effects estimator
- Regress log quantity consumed on state/year fixed effects, covariates, and real tax rate (in cents)

TABLE 2

OLS Estimates, Retail Price Model: Tobacco Institute Data

Independent variable	Average state retail price, 1985–1996 Nominal (1)	Average state retail price, 1985–1996 Real (2)	Net retail price in Tennessee, 1970–1994 Nominal (3)	Net retail price in Tennessee, 1970–1994 Real (4)
Nominal/real tax	1.01 (0.04)	0.92 (0.04)		
Nominal/real wholesale price			1.07 (0.02)	0.86 (0.04)
R^2	0.972	0.933	0.989	0.963
Observations	612	612	25	25

Standard errors in parentheses. Real prices in 1997 cents/pack. Models in columns (1) and (2) control for state effects.

Source: Evans, Ringel, and Stech 1999

Evans, Ringel, and Stech: Incidence Results

- 100% pass through implies supply elasticity of $\varepsilon_S = \infty$ at state level
  - Could be different at national level
  - Important to understand how the variation you are using determines what parameter you are identifying

TABLE 3

OLS Estimates, Log Per Capita Consumption Model, Tobacco Institute Data, 1985–1996

Independent variable	Coefficients (standard errors) on					
	Real tax			Real price		
	(1)	(2)	(3)	(4)	(5)	(6)
Current value	−0.254 (0.037)	−0.165 (0.040)	−0.173 (0.041)	−0.176 (0.027)	−0.176 (0.027)	−0.167 (0.029)
1-year lag		−0.215 (0.413)	−0.188 (0.047)		−0.027 (0.032)	−0.031 (0.032)
2-year lag			−0.061 (0.045)			−0.017 (0.033)
Price elasticity	−0.424 (0.062)	−0.635 (0.074)	−0.705 (0.090)	−0.294 (0.045)	−0.337 (0.058)	−0.359 (0.072)
R^2	0.975	0.977	0.977	0.975	0.975	0.976

Source: Evans, Ringel, and Stech 1999

Evans, Ringel, and Stech: Demand Elasticity

- Demand model estimate implies that: $\varepsilon_D = -0.42$

  $\rightarrow$ 10% increase in price induces a 4.2% reduction in consumption

- Tax passed 1-1 onto consumers, so we can compute ε_D from $\hat{\beta}$ in demand model:

$$\varepsilon_D = \frac{P}{Q}\frac{\Delta Q}{\Delta T} = \hat{\beta}/(\Delta T/P)$$

  taking P and Q average values in the data

- Can substitute $\Delta P = \Delta T$ here because of 1-1 pass through

IV Estimation of Price Elasticities

- How to estimate price elasticity of demand when tax and prices do not move together 1-1?
- Standard technique: instrument for prices using taxes
- First stage, taking note of F-stat:

$$P_{jt} = \alpha' + \gamma'_t + \delta'_j + \beta T_{jt} + \epsilon_{jt}$$

- Second stage:

$$Q_{jt} = \alpha + \gamma_t + \delta_j + \lambda \widehat{P}_{jt} + \epsilon_{jt}$$

- Reduced form, using T_{jt} as an instrument for P_{jt}:

$$Q_{jt} = \alpha + \gamma_t + \delta_j + \mu T_{jt} + \epsilon_{jt}$$

- 2SLS regression coefficient:

$$\hat{\lambda} = \hat{\mu} / \hat{\beta}$$

Evans, Ringel, and Stech: Long Run Elasticity

- *DD* before and after one year captures short term response: effect of current price P_{jt} on current consumption Q_{jt}
- F.E. also captures short term responses
- What if full response takes more than one period? Especially important considering nature of cigarette use
- F.E. estimate biased. One solution: include lags ($T_{j,t-1}, T_{j,t-2}, ...$).
- Are identification assumptions still valid here? Tradeoff between LR and validity of identification assumptions

Evans, Ringel, and Stech: Distributional Incidence

- Use individual data to see who smokes by education group and income level
- Spending per capita decreases with the income level
- Tax is regressive on an absolute level (not only that share of taxes relative to income goes down)
- Conclusion: Taxes/fines levied on cigarette companies lead to poor paying more for same goods, with no impact on companies!

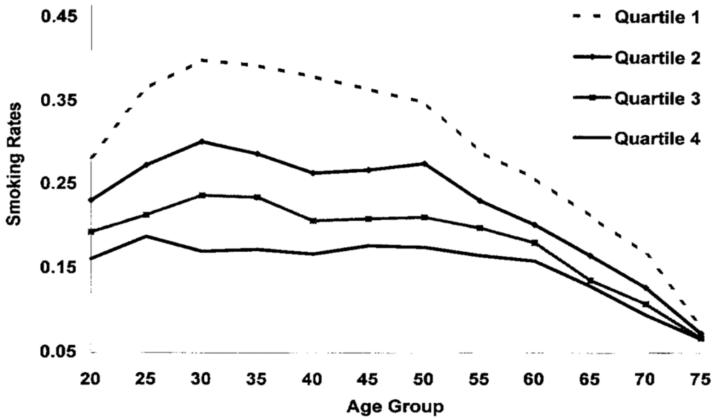


FIGURE 8. *Smoking Rates by Income Quartiles and Age—1992–1993 CPS TUS*

Source: Evans, Ringel, and Stech 1999

Cigarette Tax Incidence: Other Considerations

1. Lifetime vs. current incidence (Poterba 1989)
   - Finds cigarette, gasoline and alcohol taxation are less regressive (in statutory terms) from a lifetime perspective
   - High corr. between income and cons share in cross-section; weaker corr. with permanent income.
2. Behavioral models (Gruber and Koszegi 2004)
   - If agents have self control problems, incidence conc. on poor is beneficial to the extent that they smoke less
3. Intensive vs. extensive margin: Adda and Cornaglia (2006)
   - Use data on cotinine (biomarker) levels in lungs to measure inhalation
   - Higher taxes lead to fewer cigarettes smoked but no effect on cotinine in lungs, implying longer inhalation of each cigarette

Hastings and Washington 2008

- Question: How does food stamps subsidy affect grocery store pricing?
- Food stamps typically arrive at the same time for a large group of people, e.g. first of the month
- Use this variation to study:
  1. Whether demand changes at beginning of month (violating PIH)
  2. How much of the food stamp benefit is taken by firms by increased prices rather than consumers (intended recipients)

Hastings and Washington: Data

- Scanner data from several grocery stores in Nevada
- Data from stores in high-poverty areas ($>$15% food stamp recipients) and in low-poverty areas ($<$3%)
- Club card data on whether each individual used food stamps
- Data from other states where food stamps are staggered across month used as a control
- Research design: use variation across stores, individuals, and time of month to measure pricing responses

Table 2: Change in Expenditures across stores

	All	Store 1	Store 2	Store3
Benefit Household*Week2	-0.2053**	-0.2337**	-0.1439**	-0.1944**
	(0.007)	(0.010)	(0.015)	(0.014)
Benefit Household*Week3	-0.2907**	-0.3167**	-0.2027**	-0.2853**
	(0.008)	(0.010)	(0.017)	(0.015)
Benefit Household*Week4	-0.3352**	-0.3768**	-0.2293**	-0.3208**
	(0.008)	(0.011)	(0.017)	(0.015)
Week2	-0.0186**	-0.0104**	-0.0310**	-0.0223**
	(0.003)	(0.004)	(0.005)	(0.007)
Week3	-0.0062*	-0.0071	-0.0025	-0.0132
	(0.003)	(0.004)	(0.005)	(0.007)
Week4	-0.0061*	-0.0147**	-0.0019	0.0087
	(0.003)	(0.004)	(0.005)	(0.007)
Mean Expenditures By				
Benefit Households	58.19	63.8	59.98	47.86
Non-Benefit Households	30.66	31.13	32.76	24.56
Obs.	1398145	731353	404386	262406

Notes: Standard errors in parentheses, * significant at 5%; ** significant at 1%

Source: Hastings and Washington 2008.

- Expenditure is 20-30% higher in week 1 for food stamp recipients (red)
- But no change across weeks for non-food-stamp recipients (blue)

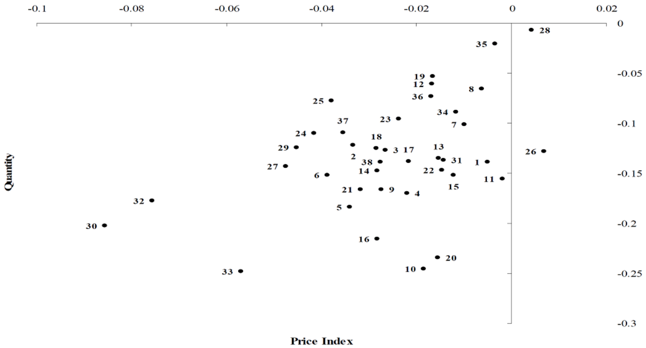
Table 8: Change in Log(Price Index) and Log (Price Index of The First Week of The Month) across different Stores

	NV High Poverty				NV Low Poverty			
	All	Store 1	Store 2	Store3	All	Store 1	Store 2	Store3
Log.(Price Index)								
Week2	-0.0195**	-0.0172**	-0.0212**	-0.0204**	-0.0012	-0.0011	-0.0038	0.0014
	(0.002)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)	(0.005)	(0.005)
Week3	-0.0243**	-0.0138**	-0.0268**	-0.0334**	0.0015	0.0021	-0.0005	0.0027
	(0.002)	(0.004)	(0.003)	(0.004)	(0.003)	(0.005)	(0.005)	(0.005)
Week4	-0.0266**	-0.0165**	-0.0293**	-0.0349**	-0.0046	-0.0075	-0.0036	-0.0027
	(0.002)	(0.005)	(0.003)	(0.004)	(0.003)	(0.005)	(0.005)	(0.005)

Source: Hastings and Washington 2008.

- Store raises prices by 2-3% in week 1 in high-poverty areas (red)
- But no change across weeks of the month in low-poverty areas (blue)

Figure 2: Scatterplot of Change in Quantities Purchased and Change in Price, by Product Category



Legend: 1. Candy, Gum and Mints; 2. Cookies/Crackers and Misc Snacks; 3. Jams, Jellies and Spreads; 4. Soft Beverages; 5. Salty Snacks; 6. Cereal and Breakfast; 7. Desserts; 8. Desserts and Baking Mixes; 9. Flour/Sugar/Corn Meal; 10. Shelf Stable Juices and Drinks; 11. Canned Vegetables; 12. Canned Fruits; 13. Canned Fish and Meat; 14. Ready-to-Eat Prepared Foods; 15. Soups; 16; Rice & Beans; 17. Pasta and Pasta Sauce; 18. Hispanic; 19. Tobacco and Smoking Needs. 20. Refrigerated Dairy; 21. Cheese; 22. Refrigerated Foods; 23. Refrigerated Juice and Drinks; 24. Ice Cram; 25. Frozen Dessert; 15. Frozen Fruits 27. Frozen Breakfast Items; 28. Frozen Juices; 29. Frozen Vegetables; 30. Prepared Frozen Foods; 31. Fresh Produce; 32. Fresh Produce

Source: Hastings and Washington 2008.

Hastings and Washington: Results

- Demand increases by 30% in 1st week, prices by about 3%
- Very compelling because of multiple dimensions of tests: cross-individual, cross-store, cross-category, and cross-state
- Areas for future work:
  1. Pricing outside of supermarkets; many other outlets where food stamps are used may change prices differently
  2. Incidence effects for goods other than groceries could be very different (car prices and EITC payments)
- Interesting theoretical implication: subisidies in markets where low-income recipients are pooled with others have better distributional effects
  - May favor food stamps as a way to transfer money to low incomes relative to subsidy such as EITC

Rothstein 2008

- How does EITC affect wages?
- EITC payments subsidize work and transfer money to low income working individuals ($50 bil/year)
- This subsidy could be taken by employers by shifting wage
  - Ex: inelastic demand for low-skilled labor and elastic supply $\longrightarrow$ wage rate adjusts 1-1 with EITC
- Policy question: are we actually transferring money to low incomes through this program or are we just helping business owners?

Rothstein 2008

- Rothstein considers a simple model of the labor market with three types of agents

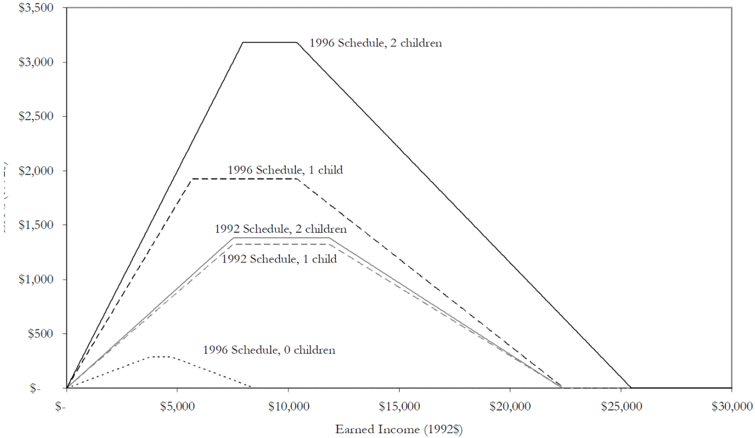
    1. Employers
    2. EITC-eligible workers
    3. EITC-ineligible workers

- Extends standard partial eq incidence model to allow for differentiated labor supply and different tax rates across demographic groups

- Heterogeneity both complicates the analysis and permits identification

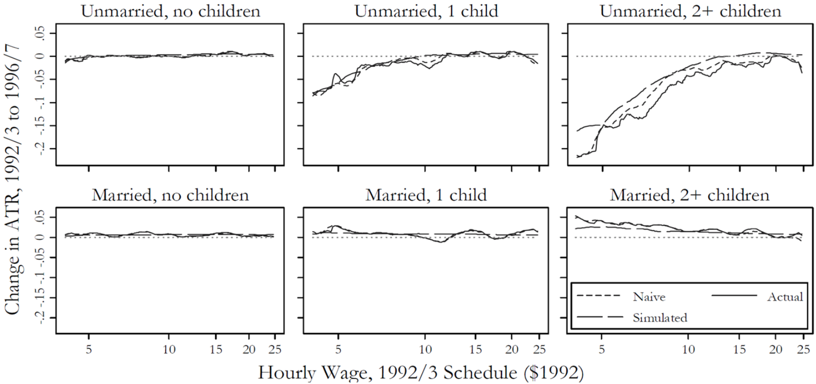
- Identification strategy: compare wage changes across groups who were affected differently by expansions of EITC program from 1992-94

Figure 1. EITC Schedule, 1992 and 1996 by number of children



Source: Rothstein 2008.

Figure 3.
Change in mean ATR among families with working women, by skill and group



Source: Rothstein 2008.

Rothstein: Empirical Strategy

- Two main challenges to identification:
  1. EITC 1992-1994 expansion when nation coming out of recession
     $\rightarrow$ Compare to other workers (EITC ineligible, slightly higher incomes)
  2. Violation of common trends assumption: technical change, more demand for low-skilled workers in 1990s.
     $\rightarrow$ Compare to trends in pre-period (essentially a *DDD* strategy)

- Two dependent variables of interest:
  1. [Prices] Measure how wages change for a worker of *given* skill
  2. [Quantities] Measure how demand and supply for workers of each skill type change because of EITC

- Basic concept: use two moments – wage and quantity changes to back out slopes of supply and demand curves

Rothstein: Empirical Strategy

- Ideal test: measure how wage of a given individual changes when EITC is introduced relative to a similar but ineligible individual
- Problem: data is CPS repeated cross-sections. Cannot track "same individual."
- Moreover, wage rigidities may prevent cuts for existing employees.
- Solution: reweighting procedure to track "same skill" worker over time (DiNardo, Fortin, and Lemieux 1996)

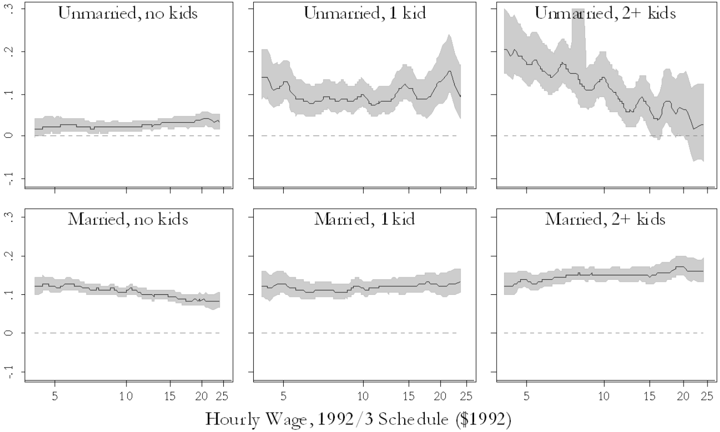
DFL Reweighting

- Widely used method that generalizes propensity score reweighting
- Used to examine changes in distributions over time semi-parametrically, conditioning on observables
- Example: suppose wages are a function purely of height
- When EITC is expanded, average observed height of workers falls because less-skilled (shorter) people enter the labor force
- We want to identify how wage distribution changes for people of given height
- Solution: hold “fixed” height semi-parametrically by reweighting the distribution of wages ex-post to match heights ex-ante.

DFL Reweighting

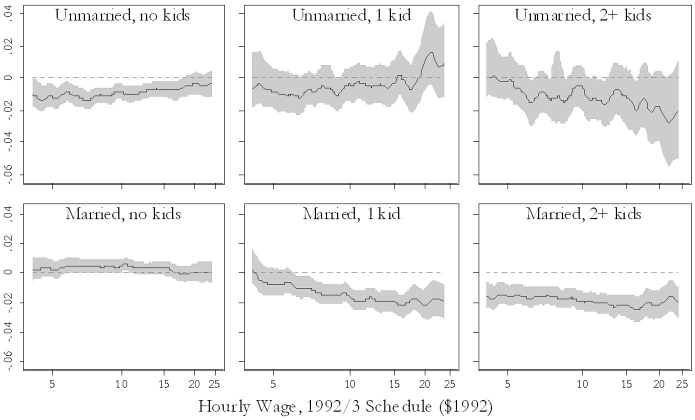
- Example: 100 short, 100 tall pre-reform and 200 short, 100 tall post-reform
- Then put 2/3 weight on tall and 1/3 on short when calculating wage distribution after reform
- Compare reweighted post-reform distribution to pre-reform distribution to assess effect of expansion on wages
- Key assumption for causal interpretation of changes: selection on observables
  - Here it is height; more generally, experience, age, demographics, etc.

Change in total labor supply



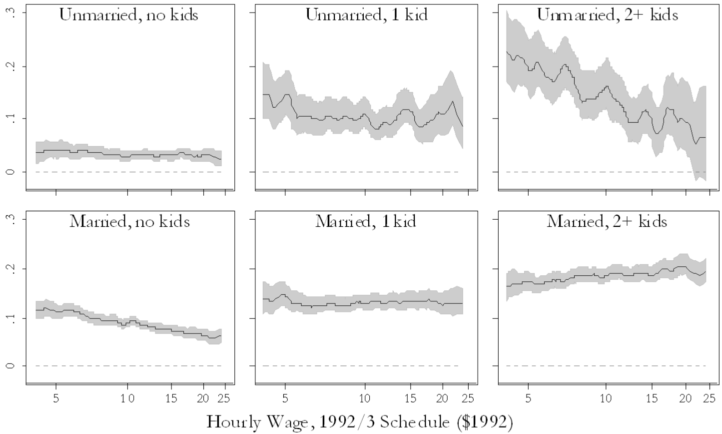
Source: Rothstein 2008.

Intensive margin: Change in weekly hours conditional on working



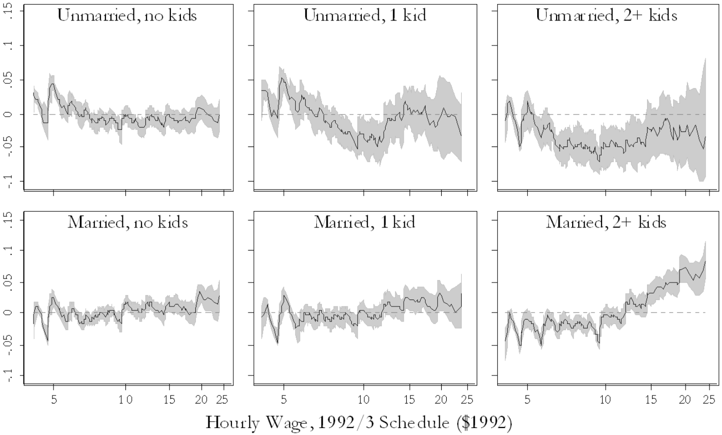
Source: Rothstein 2008.

Extensive margin: Change in employment-population ratio



Source: Rothstein 2008.

Wage changes



Source: Rothstein 2008.

Rothstein: Results

- Basic DFL comparisons yield perverse result: groups that benefited from EITC and started working more had more wage growth
- Potential explanation: demand curve shifted differentially – higher demand for low skilled workers in 1990s.
- To deal with this, repeats same analysis for 1989-1992 (no EITC expansion) and takes differences
- Changes sign back to expected, but imprecisely estimated

Table 5. Reduced form models for labor supply and wages

	Predicted	1992-1996 changes		1988-1992 & 1992-1996 changes	
		Single women	All women	Single women	All women
	(1)	(2)	(3)	(4)	(5)
Dependent variable: Change in labor supply, group g, skill s					
Change in ATR within group	$-\sigma$	**-0.73**	**-0.74**	**-1.06**	**-1.06**
		(0.25)	(0.25)	(0.23)	(0.23)
Change in avg. ATR (across groups)	$\frac{\sigma^2}{\sigma - \rho}$	-0.24	-1.02	0.00	-0.26
		(0.32)	(1.26)	(1.02)	(2.02)
Change in population size (avg. across groups)	$\frac{-\sigma}{\sigma - \rho}$	-0.10	-0.16	-0.20	-0.12
		(0.21)	(0.37)	(0.31)	(0.63)
Group dummies		y	y	y	y
Base log wage		y	y	n	n
Skill dummies		n	n	y	y
1992-1996 indicator		n	n	y	y
Base log wage * 1992-1996		n	n	y	y
Dependent variable: Change in wage, group g, skill s					
Change in avg. ATR (across groups)	$\frac{\sigma}{\sigma - \rho}$	**-1.26**	**-5.14**	2.45	**4.78**
		(0.39)	(1.75)	(1.33)	(2.28)
Change in population size (avg. across groups)	$\frac{-1}{\sigma - \rho}$	-0.38	-0.52	-0.22	-1.14
		(0.29)	(0.63)	(0.26)	(0.70)

Source: Rothstein 2008.

Rothstein: Results

- Ultimately uses quantity estimates and incidence formula to back out predicted changes
  - Wage elasticity estimates: 0.7 for labor supply, -0.3 for labor demand
- Implications using formulas from model:
  - EITC-eligible workers gain $0.70 per $1 EITC expansion
  - Employers gain about $0.70
  - EITC-ineligible low-skilled workers lose about $0.40
- On net, achieve only $0.30 of redistribution toward low income individuals for every $1 of EITC

Rothstein: Caveats

1. Identification heavily complicated by recession, trends (SBTC); no clean control group
2. Data limitations: no panel data; problems in measurement – no annual income, cannot measure MTR
3. Selection on endogenous variables
4. Short run vs. long run effects; important due to evidence of nominal wage rigidities.
5. Pure extensive-margin analysis. Intensive margin would go the other way b/c EITC is not a marginal subsidy to wage for a very large fraction of the population.
6. General equilibrium effects are not considered

Extensions of Basic Partial Equilibrium Analysis

1. **Market rigidities:**
   - With price floors, incidence can differ
   - Consider incidence of social security taxes with minimum wage
   - Statutory incidence: 6.2% on employer and 6.2% on employee
   - Share of each should not matter as long as total is constant because wages will fall to adjust
   - But with binding minimum wage, employers cannot cut wage further, so statutory incidence determines economic incidence on the margin

Extensions of Basic Partial Equilibrium Analysis

1. Market rigidities

2. **Imperfect competition**

    - Overshifting: possible to get an increase in after-tax price $>$ level of the tax
    - Ad valorem and excise taxation are no longer equivalent
    - See Salanie text

Extensions of Basic Partial Equilibrium Analysis

1. Market rigidities
2. Imperfect competition
3. **Effects on other markets:**
    - Increase in cigarette tax $\rightarrow$ substitute cigarettes for cigars, increasing price of cigars and shifting cigarette demand curve
    - Revenue effects on other markets: tax increases make agents poorer; less to spend on other markets
    - This motivates general equilibrium analysis of incidence

General Equilibrium Analysis

- Trace out full incidence of taxes back to original owners of factors
- Partial equilibrium: "producer" vs. consumer
- General equilibrium: capital owners vs. labor vs. landlords, etc.
- Two types of models:
  1. **Static**: many sectors or many factors of production
     - Workhorse analytical model: Harberger (1962): 2 sector and 2 factors of production
     - Computational General Equilibrium: many sectors, many factors of production model
  2. **Dynamic**
     - Intergenerational incidence: Soc Sec reform
     - Asset price effects: capitalization

Harberger 1962 Two Sector Model

1. Fixed total supply of labor L and capital K (short-run, closed economy)
2. Constant returns to scale in both production sectors
3. Full employment of L and K
4. Firms are perfectly competitive

Implicit assumption: no adjustment costs for capital and labor

Harberger Model: Setup

- Production in sectors 1 (bikes) and 2 (cars):

$$X_1 = F_1(K_1, L_1) = L_1 f(k_1)$$
$$X_2 = F_2(K_2, L_2) = L_2 f(k_2)$$

  with full employment conditions $K_1 + K_2 = K$ and $L_1 + L_2 = L$

- Factors w and L fully mobile $\rightarrow$ in eq., returns must be equal:

$$w = p_1 F_{1L} = p_2 F_{2L}$$
$$r = p_1 F_{1K} = p_2 F_{2K}$$

- Demand functions for goods 1 and 2:

$$X_1 = X_1(p_1/p_2) \text{ and } X_2 = X_2(p_1/p_2)$$

- Note: all consumers identical so redistribution of incomes via tax system does not affect demand via a feedback effect

- System of ten eq'ns and ten unknowns: K_i, L_i, p_i, X_i, w, r

Harberger Model: Effect of Tax Increase

- Introduce small tax $d\tau$ on rental of capital in sector 2 (K_2)
- All eqns the same as above except $r = (1 - d\tau)p_2 F_{2K}$
- Linearize the 10 eq'ns around initial equilibrium to compute the effect of $d\tau$ on all 10 variables (dw, dr, dL_1, ...)
- Labor income $= wL$ with L fixed, $rK =$ capital income with K fixed
- Therefore change in prices $dw/d\tau$ and $dr/d\tau$ describes how tax is shifted from capital to labor
- Changes in prices $dp_1/d\tau$, $dp_2/d\tau$ describe how tax is shifted from sector 2 to sector 1
- Kotlikoff and Summers (Section 2.2) state linearized equations as a fn. of substitution elasticities

Harberger Model: Main Effects

1. **Substitution effects**: capital bears incidence

   - Tax on K_2 shifts production in Sector 2 away from K so aggregate demand for K goes down

   - Because total K is fixed, r falls $\rightarrow$ K bears some of the burden

Harberger Model: Main Effects

2. **Output effects**: capital may not bear incidence

- Tax on K_2 implies that sector 2 output becomes more expensive relative to sector one
- Therefore demand shifts toward sector 1
- *Case 1:* $K_1/L_1 < K_2/L_2$ (1: bikes, 2: cars)
  - Sector 1 is less capital intensive so aggregate demand for K goes down
  - Output effect reinforces subst effect: K bears the burden of the tax
- *Case 2:* $K_1/L_1 > K_2/L_2$ (1: cars, 2: bikes)
  - Sector 1 is more capital intensive, aggregate demand for K increases
  - Subst. and output effects have opposite signs; labor may bear some or all the tax

Harberger Model: Main Effects

3. **Substitution + Output = Overshifting effects**

- *Case 1:* $K_1/L_1 < K_2/L_2$
  - Can get overshifting of tax, $dr < -d\tau$ and $dw > 0$
  - Capital bears more than 100% of the burden if output effect sufficiently strong
  - Taxing capital in sector 2 raises prices of cars $\rightarrow$ more demand for bikes, less demand for cars
  - With very elastic demand (two goods are highly substitutable), demand for labor rises sharply and demand for capital falls sharply
  - Capital loses more than direct tax effect and labor suppliers gain

Harberger Model: Main Effects

3. **Substitution + Output = Overshifting effects**

- *Case 2*: $K_1/L_1 > K_2/L_2$
  - Possible that capital is made better off by capital tax
  - Labor forced to bear more than 100% of incidence of capital tax in sector 2
  - Ex. Consider tax on capital in bike sector: demand for bikes falls, demand for cars rises
  - Capital in greater demand than it was before → price of labor falls substantially, capital owners actually gain
- Bottom line: taxed factor may bear less than 0 or more than 100% of tax.

Harberger Two Sector Model

- Theory not very informative: model mainly used to illustrate negative result that "anything goes"
- More interest now in developing methods to identify what actually happens
- Original Application of this framework by Harberger: sectors = housing and corporations
- Capital in these sectors taxed differently because of corporate income tax and many tax subsidies to housing
  - Ex: Deductions for mortgage interest and prop. tax are about $50 bn total
- Harberger made assumptions about elasticities and calculated incidence of corporate tax given potential to substitute into housing

Computable General Equilibrium Models

- Harberger analyzed two sectors; subsequent literature expanded analysis to multiple sectors
- Analytical methods infeasible in multi-sector models
- Instead, use numerical simulations to investigate tax incidence effects after specifying full model
- Pioneered by Shoven and Whalley (1972). See Kotlikoff and Summers section 2.3 for a review
- Produced a voluminous body of research in PF, trade, and development economics

CGE Models: General Structure

- N intermediate production sectors
- M final consumption goods
- J groups of consumers who consume products and supply labor
- Each industry has different substitution elasticities for capital and labor
- Each consumer group has Cobb-Douglas utility over M consumption goods with different parameters
- Specify all these parameters (calibrated to match some elasticities) and then simulate effects of tax changes

Criticism of CGE Models

- Findings very sensitive to structure of the model: savings behavior, perfect competition assumption
- Findings sensitive to size of key behavioral elasticities and functional form assumptions
- Modern econometric methods conceptually not suitable for GE problems, where the whole point is "spillover effects" (contamination)
- Need a new empirical paradigm to deal with these problems – a major open challenge

Open Economy Application

- Key assumption in Harberger model: both labor and capital perfectly mobile across sectors
- Now apply framework to analyze capital taxation in open economies, where capital is more likely to be mobile than labor
- See Kotlikoff and Summers section 3.1 for a good exposition

Open Economy Application: Framework

- One good, two-factor, two-sector model
- *Sector 1*: small open economy where L_1 is fixed and K_1 mobile
- *Sector 2*: rest of the world L_2 fixed and K_2 mobile
- Total capital stock $K = K_1 + K_2$ is fixed

Open Economy Application: Framework

- Small country introduces tax on capital income (K_1)

- After-tax returns must be equal:

$$r^* = F_{2K} = (1 - \tau)F_{1K}$$

- Capital flows from 1 to 2 until returns are equalized; if 2 is large relative to 1, no effect on r^*

- Wage rate $w_1 = F_{1L}(K_1, L_1)$ dec. when K_1 dec. b/c L_1 is fixed

- Return of capitalists in small country is unchanged; workers in home country bear the burden of the tax

    - Taxing capital is bad for workers!

Open Economy Application: Empirics

- Mobility of K drives the previous result
- Empirical question: is K actually mobile across countries?
- Two strategies:
  1. Test based on prices and equilibrium relationships [Macro finance]
  2. Look at mobility directly [Feldstein and Horioka 1980]

Strategy One: Macro-Finance approach

- Test based on prices and equilibrium relationships
- Check whether net returns r are equal across countries
- General finding - covered interest parity: obligations that are protected against fluctuations in inflation and exchange rates have the same returns across countries
- Difficulties in generalization: many assets yield different returns, unexpected inflation, changes in currency exchange rates
- Need models with uncertainty, risk aversion to deal with other assets
- Difficult to implement this test for risky assets

Feldstein and Horioka 1980

- Second strategy: look at capital mobility directly
- Feldstein and Horioka use data on OECD countries from 1960-74
- Closed economy: $S = I$; open economy: $S - I = X - M$
- Motivates regression:

$$I/GDP = \alpha + \beta S/GDP + ...$$

- Find $\beta = 0.89\ (0.07)$

Feldstein and Horioka 1980

- In closed economy, $\beta = 1$
- But do not know what β should be in an open economy
- β may be close to 1 in open economy if
    1. Policy objectives involving $S - I$ (trade deficit balance)
    2. Summing over all countries: $\bar{S} = \bar{I}$ as imports and exports cancel out
    3. Data problem: S constructed from I in some countries

Open Economy Applications: Empirics

- Large subsequent literature runs similar regressions and finds mixed results
  - Generally finds more flow of capital and increasing over time
- General view: cannot extract money from capital in small open economies
  - Ex. Europe: tax competition has led to lower capital tax rates
  - Could explain why state capital taxes are relatively low in the U.S.

General Equilibrium Incidence in Dynamic Models

- Static analysis above assumes that all prices and quantities adjust immediately
- In practice, adjustment of capital stock and reallocation of labor takes time
- Dynamic CGE models incorporate these effects; even more complex
  - Static model can be viewed as description of steady states
  - During transition path, measured flow prices (r, w) will not correspond to steady state responses
- How to measure incidence in dynamic models?

Capitalization and the Asset Price Approach

- Asset prices can be used to infer incidence in dynamic models (Summers 1983)
  - Study effect of tax changes on asset prices
  - Asset prices adjust *immediately* in efficient markets, incorporating the full present-value of subsequent changes
  - Efficient asset markets incorporate all effects on factor costs, output prices, etc.
- Limitation: can only be used to characterize incidence of policies on capital owners
  - There are no markets for individuals

Simple Model of Capitalization Effects

- Firms pay out profits as dividends
- Profits determined by revenues net of factor payments:

$$V = \sum \frac{D_t}{1+r} = \sum \frac{q_t X_t - w_{jt} L_{jt}}{1+r}$$

- Change in valuation of firm ($\frac{dV}{dt}$) reflects change in present value of profits
- $\frac{dV}{dt}$ is a sufficient statistic that incorporates changes in all prices
- Empirical applications typically use "event study" methodology
- Examine pattern of asset prices or returns over time, look for break at time of announcement of policy change
- Problem: clean shocks are rare; big reforms do not happen suddenly and are always expected to some extent

Empirical Applications

1. [Cutler 1988] Effect of Tax Reform Act of 1986 on corporations
2. [Linden and Rockoff 2008] Effect of a sex offender moving into neighborhood on home values
3. [Friedman 2008] Effect of Medicare Part D on drug companies

Cutler 1988

- Looks at the Tax Reform Act of 1986, which:
  1. Decreased the tax rate on corporate income
  2. Repealed the investment tax credit and reduced depreciation allowances
- These changes hurt companies with higher levels of current investment
- Examines daily returns of 350 firms, R_{it}

Cutler 1988

- First, compute excess return ($\hat{\epsilon}_{is}$) for each firm i by regressing:

$$R_{it} = \alpha + \beta_i R_{Mt} + \epsilon_{it}$$

- Obtain excess return $\hat{\epsilon}_{is}$: return purged of market trends

- Here, events are the dates when TRA was voted on in the House and Senate

- Compute the average excess return in a $\pm$ 10 day window for each firm $Excess_i = \hat{\epsilon}_{is}$ where s is the time of the event

- Second step regression:

$$Excess_i = a + b(Inv/K)_i + \nu_i$$

where $(Inv/K)_i$ is a measure of the rate of investment of firm i

- Theory predicts $b < 0$

TABLE 2—TAX CHANGES AND MARKET VALUE[a]

Period	Machinery and Equipment Share	Rate of Investment	Net Capital Stock	Average Tax Rate	Predicted Change In Cash Flow	Industry Dummies	R^2	N
One Day	.018	−.004	.037	−.002	−.045	No	.022	336
	(.012)	(.006)	(.033)	(.004)	(.064)			
One Day	.000	−.003	.032	−.003	−.044	Yes	.097	336
	(.015)	(.006)	(.036)	(.005)	(.066)			
Ten Days	.083	−.029	.068	−.009	−.126	No	.061	333
	(.025)	(.013)	(.070)	(.010)	(.141)			
Ten Days	.035	−.028	.044	−.011	−.166	Yes	.205	333
	(.031)	(.013)	(.075)	(.010)	(.139)			
One Month	.064	−.131	−.025	−.028	−.173	No	.052	330
	(.043)	(.042)	(.117)	(.016)	(.318)			
One Month	.024	−.158	−.027	−.028	.115	Yes	.174	330
	(.051)	(.044)	(.124)	(.016)	(.323)			

Source: Cutler 1988.

Cutler: Results

- Cutler finds $\hat{b} = -0.029(0.013)$
- This is consistent with expectations, but other findings are not:
  - Changes in future tax liabilities not correlated with stock value changes
  - Responses to two distinct events (passage of bill in House and Senate) not correlated
  - Were the votes really surprises? Need data on expectations
- Study is somewhat inconclusive because of noisy data
- But led to a subsequent better-identified literature

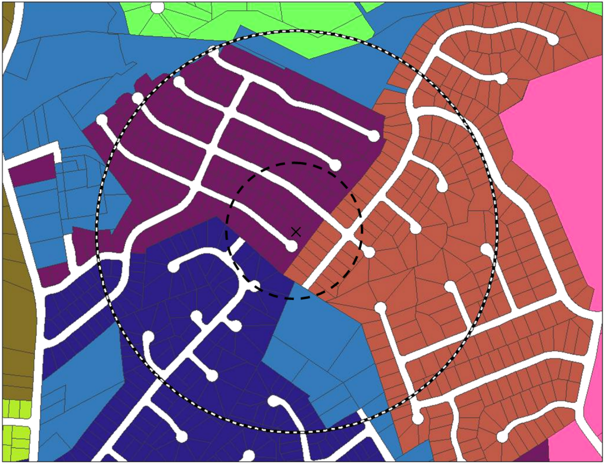
Linden and Rockoff 2008

- Another common application is to housing market to assess WTP for amenities
- Examples: pollution, schools, crime
- Rockoff and Linden (2008) estimate costs of crime using capitalization approach
- Identification strategy: look at how house prices change when a registered sex offender moves into a neighborhood
- Data: public records on offender's addresses and property values in North Carolina

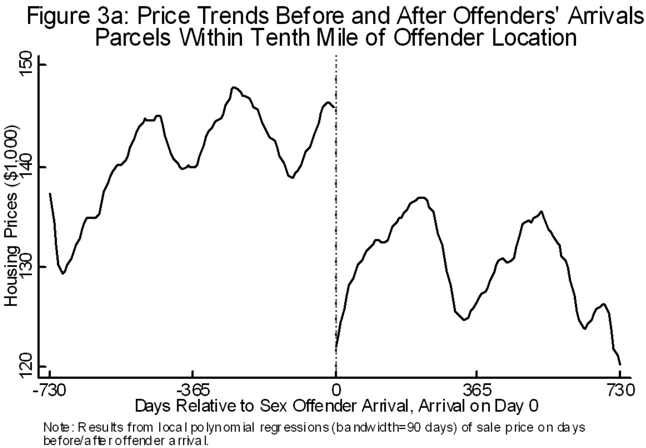
Illustration of Identification Strategy

Source: Linden and Rockoff 2008.

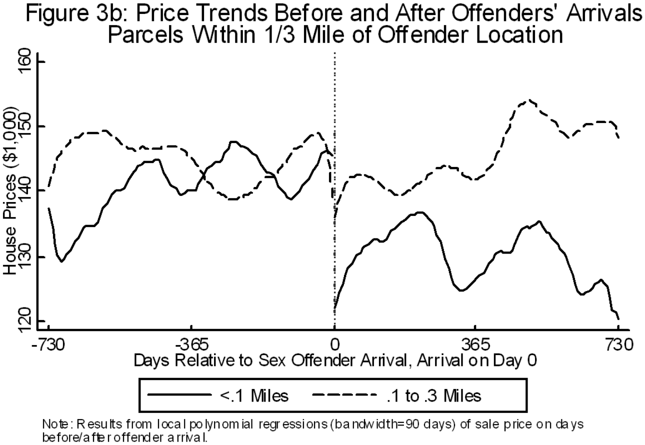
Illustration of Identification Strategy



Source: Linden and Rockoff 2008.



Source: Linden and Rockoff 2008.



Source: Linden and Rockoff 2008.

Linden and Rockoff: Results

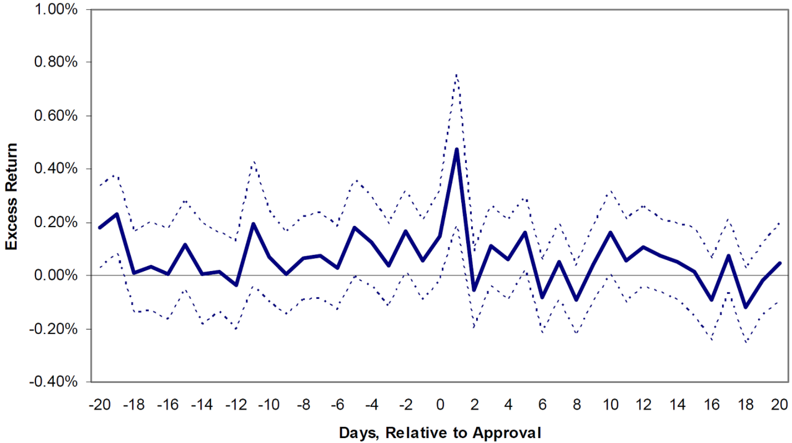
- Find house prices decline by about 4% ($5500) when a sex offender is located within 0.1 mile of the house
- Implied cost of a sexual offense given probabilities of a crime: $1.2 million
- This is far above what is used by Dept of Justice
- How to interpret evidence: true cost of crime or a behavioral effect?
  - Why does price fall only within 0.1 mile radius?

Friedman 2008

- Medicare part D passed by Congress in 2003; enacted in 2006
- Expanded Medicare coverage to include prescription drugs (provided coverage for 10 mil additional people)
- What is the incidence of Medicare part D? How much of the expenditure is captured by drug companies through higher profits?
- Event study: excess returns for drug companies around FDA approval of drugs
- Tests whether excess returns for high-Medicare share drugs is higher after Medicare Part D is passed
- Let MMS_i denote medicare market share drug class i. Second-stage estimating equation:

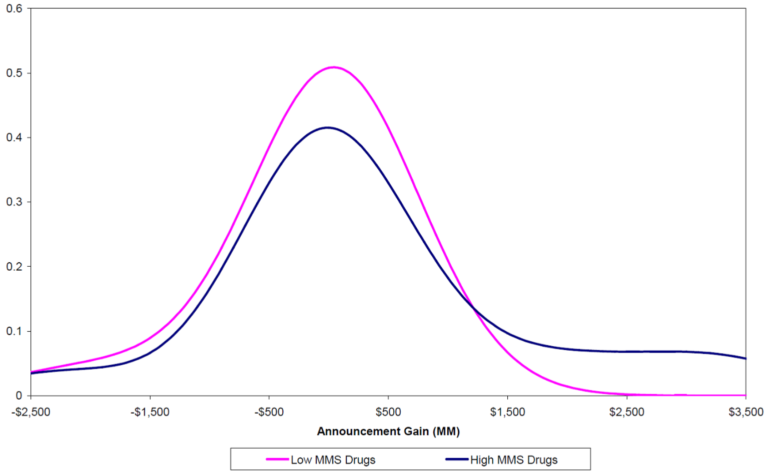
$$Excess_i = \alpha + \beta MMS_i + \gamma Post2003_t + \lambda Post2003_t \cdot MMS_i$$

Excess Returns Around Drug Approval Date



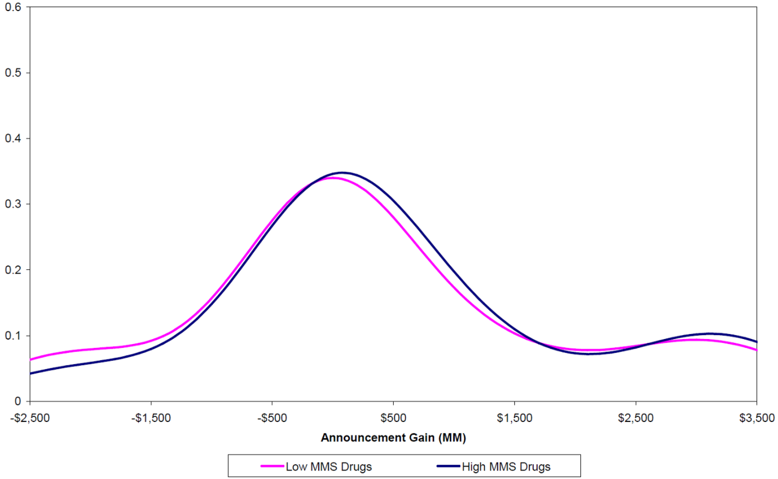
Source: Friedman 2008

Distribution of Excess Returns around Drug Approval: Post-Reform (2004-2007)



Source: Friedman 2008

Distribution of Excess Returns around Drug Approval: Pre-Reform (1999-2002)



Source: Friedman 2008

Friedman: Results

- Concludes that drug companies' profits increased by $250 bn in present value because of Medicare Part D
- Rough calibration suggests that drug companies capture about 1/3 of total surplus from program

Mandated Benefits

- We have focused until now on incidence of price interventions (taxes, subsidies)
- Similar incidence/shifting issues arise in analyzing quantity intervention (regulations)
- Leading case: mandated benefits – requirement that employers pay for health care, workers compensation benefits, child care, etc.
- Mandates are attractive to government because they are "off budget"; may reflect salience issues

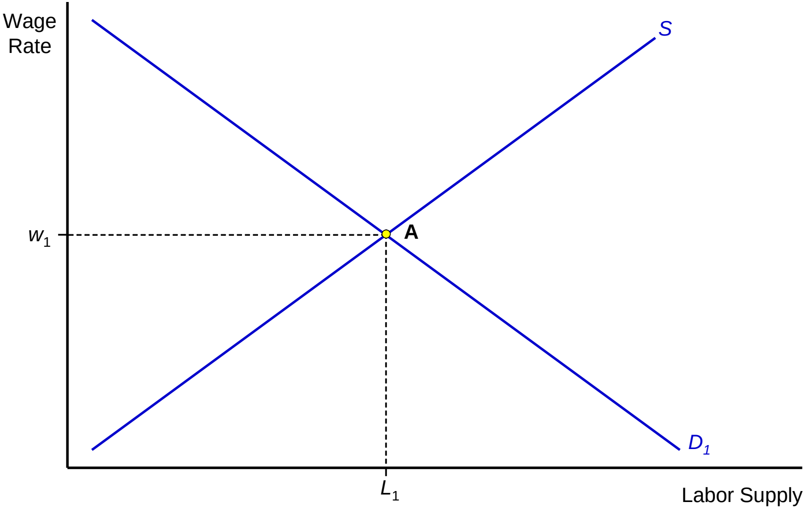
Mandated Benefits

- Tempting to view mandates as additional taxes on firms and apply same analysis as above
- But mandated benefits have different effects on equilibrium wages and employment differently than a tax (Summers 1989)
- Key difference: mandates are a benefit for the worker, so effect on market equilibrium depends on benefits workers get from the program
- Unlike a tax, may have no distortionary effect on employment and only an incidence effect (lower wages)

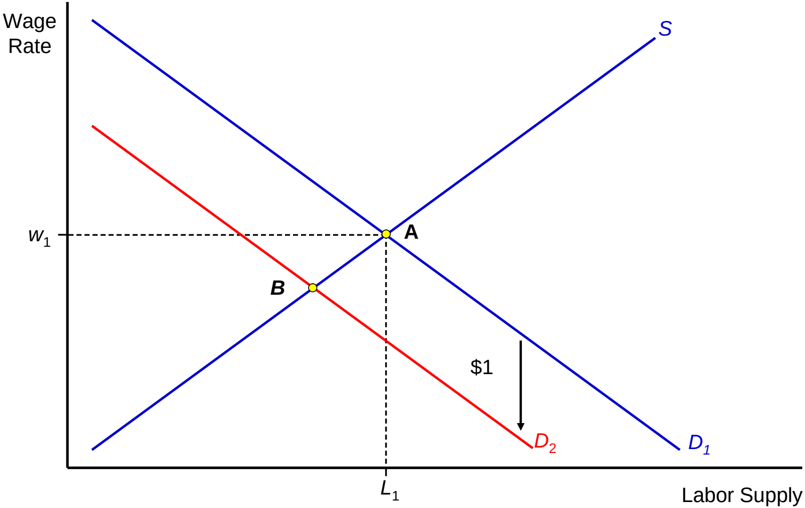
Mandated Benefits: Simple Model

- Labor demand (D) and labor supply (S) are functions of the wage, w
- Initial equilibrium:
$$D(w_0) = S(w_0)$$
- Now, govt mandates employers provide a benefit with cost t
- Workers value the benefit at αt dollars
- Typically $0 < \alpha < 1$ but $\alpha > 1$ possible with market failures
- Labor cost is now $w + t$, effective wage $w + \alpha t$
- New equilibrium:
$$D(w + t) = S(w + \alpha t)$$

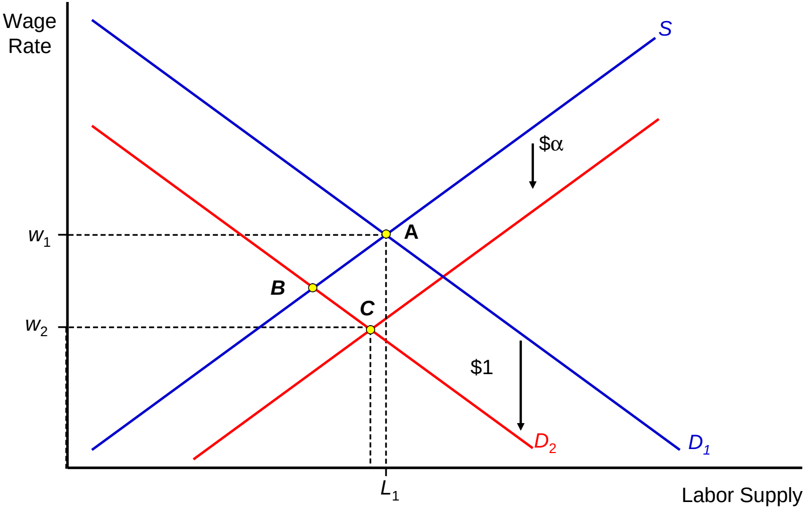
Mandated Benefit



Mandated Benefit



Mandated Benefit



Mandated Benefits: Incidence Formula

- Analysis for a small t: linear expansion around initial equilibrium

$$
\begin{aligned}
(dw/dt + 1)D' &= (dw/dt + \alpha)S' \\
dw/dt &= (D' - \alpha S')/(S' - D') \\
&= -1 + (1-\alpha)\frac{\eta_S}{\eta_S - \eta_D}
\end{aligned}
$$

where

$$
\begin{aligned}
\eta_D &= wD'/D < 0 \\
\eta_S &= wS'/S > 0
\end{aligned}
$$

- If $\alpha = 1$, $dw/dt = -1$ and no effect on employment

- More generally: $0 < \alpha < 1$ equivalent to a tax $1 - \alpha$ with usual incidence and efficiency effects

Empirical Applications

1. [Gruber 1994] Pregnancy health insurance costs
2. [Acemoglu and Angrist 2001] Americans with Disabilities Act

Gruber 1994

- Studies state mandates for employer-provided health insurance to cover pregnancy costs
- In 1990, expected cost for pregnancy about $500 per year for married women aged 20 to 40
- State law changes to mandate coverage of pregnancy costs in 1976

Gruber: Empirical Strategy

- Uses difference-in-difference estimator:

$$DD^T = [W_{YA} - W_{YB}] - [W_{NA} - W_{NB}]$$

- Time periods: before 1974-75 (B), after 1977-78 (A)
- Three experimental states (Y): IL, NJ, and NY
- Five nearby control states (N)
- Concern: differential evolution of wages in control vs. treatment states
- Placebo DD^C for control group: people over 40 and single males aged 20-40

$$DDD = DD^T - DD^C$$

Table 3—DDD Estimates of the Impact of State Mandates on Hourly Wages

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:	−0.062 (0.022)		
B. *Control Group: Over 40 and Single Males 20–40:*			
Experimental states	1.759 (0.007) [5,624]	1.748 (0.007) [5,407]	−0.011 (0.010)
Nonexperimental states	1.630 (0.007) [4,959]	1.627 (0.007) [4,928]	−0.003 (0.010)
Location difference at a point in time:	0.129 (0.010)	0.121 (0.010)	
Difference-in-difference:	−0.008: (0.014)		
DDD:	**−0.054** **(0.026)**		

Source: Gruber 1994.

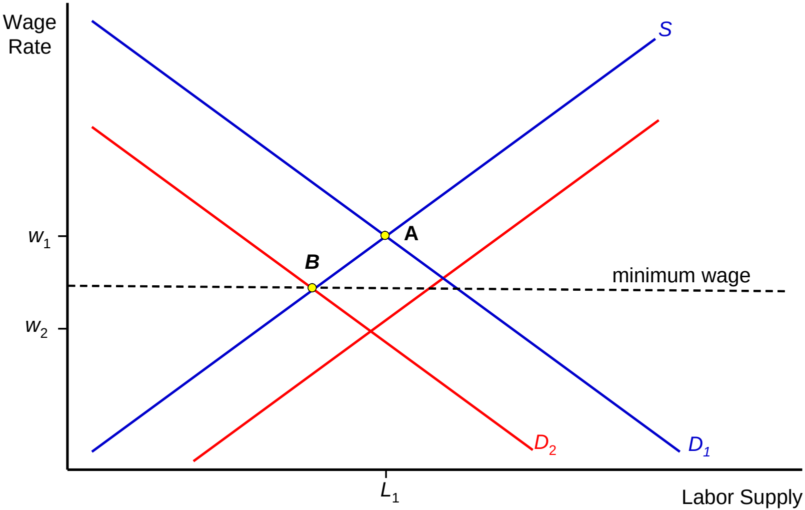
Gruber: Results

- Find $DD^T = -0.062(0.022)$, $DDD = -0.054(0.026)$
- Implies that hourly wage decreases by roughly the cost of the mandate (no distortion case, $\alpha = 1$).
- Indirect aggregate evidence also suggests that costs have been shifted on wages
  - Share of health care costs in total employee compensation has increased substantially over last 30 years
  - But share of total employee compensation as a share of national income roughly unchanged

Acemoglu and Angrist 2001

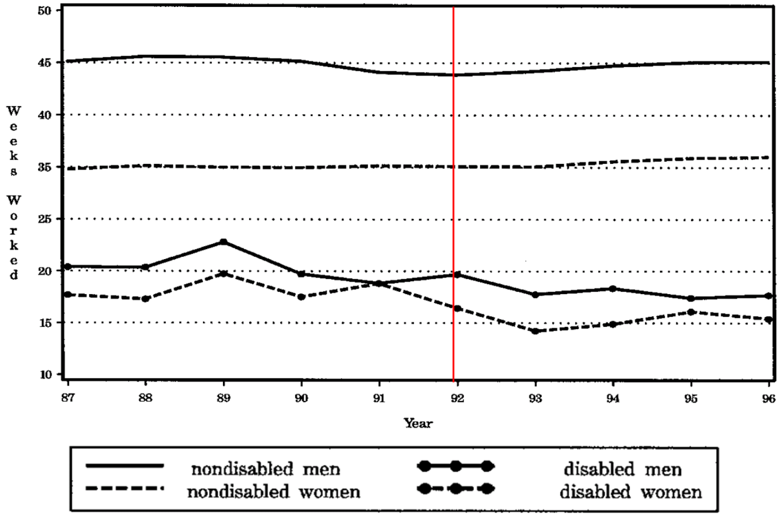
- Look at effect of ADA regulations on wages and employment of the disabled
- The 1993 Americans with Disabilities Act requires employers to:
  - Make accommodations for disabled employees
  - Pay same wages to disabled employees as to non-disabled workers
- Cost to accommodate disabled workers: $1000 per person on average
- Theory is ambiguous on net employment effect because of wage discrimination clause

Mandated Benefit with Minimum Wage



Acemoglu and Angrist 2001

- Acemoglu and Angrist estimate the impact of act using data from the Current Population Survey
- Examine employment and wages of disabled workers before and after the ADA went into effect



Source: Acemoglu and Angrist 2001.

Effect	Dependent Variable: Weeks Worked				Dependent Variable: Log of Weekly Earnings			
	Men		Women		Men		Women	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Age 21–39							
Disability × 1988	−.41		−1.08		−.052		−.055	
	(.71)		(1.07)		(.044)		(.052)	
Disability × 1989	2.00		.67		−.062		.050	
	(.71)		(1.06)		(.043)		(.050)	
Disability × 1990	−.79		−1.33		−.035		−.111	
	(.71)		(1.05)		(.044)		(.050)	
Disability × 1991	−.53	−.70	.08	1.09	−.020	.046	.105	.191
	(.70)	(.78)	(1.04)	(1.16)	(.043)	(.049)	(.050)	(.055)
Disability × 1992	.57	.40	−2.81	−1.56	−.00	.078	−.010	.100
	(.69)	(.95)	(1.02)	(1.40)	(.043)	(.059)	(.050)	(.067)
Disability × 1993	−1.44	−1.61	−4.37	−2.89	−.04	.050	−.058	.074
	(.69)	(1.14)	(1.02)	(1.69)	(.044)	(.072)	(.050)	(.082)
Disability × 1994	−1.63	−1.80	−5.00	−3.29	−.143	−.042	−.043	.112
	(.70)	(1.35)	(1.01)	(1.98)	(.044)	(.084)	(.050)	(.095)
Disability × 1995	−2.93	−3.09	−3.93	−1.99	−.098	.014	.005	.184
	(.72)	(1.56)	(1.01)	(2.29)	(.046)	(.098)	(.049)	(.110)
Disability × 1996	−2.68	−2.84	−4.41	−2.23	−.158	−.034	−.111	.090
	(.70)	(1.77)	(1.03)	(2.62)	(.044)	(.110)	(.050)	(.125)
Disability × linear trend		−.004		−.23		−.012		−.023
		(.22)		(.33)		(.014)		(.016)
Observations	193,317	193,317	211,910	211,910	167,974	167,974	157,688	157,688

Source: Acemoglu and Angrist 2001.

Acemoglu and Angrist: Results

- Employment of disabled workers fell after the reform:
  - About a 1.5-2 week drop in employment for males, roughly a 5-10% decline in employment
- Wages did not change
- Results consistent w/ labor demand elasticity of about -1 or -2
- Firms with fewer than 25 workers exempt from ADA regulations; no employment reduction for disabled at these firms
- ADA intended to help those with disabilities but appears to have hurt many of them because of wage discrimination clause
  - Underscores importance of considering incidence effects before implementing policies

Public Economics Lectures
Part 3: Efficiency Cost of Taxation

Raj Chetty and Gregory A. Bruich

Harvard University

Fall 2009

Outline

1. Marshallian surplus
2. Path dependence problem and income effects
3. Definitions of EV, CV, and excess burden with income effects
4. Harberger formula
5. Exact Consumer Surplus (Hausman 1981)
6. Empirical Applications
7. Welfare Analysis in Behavioral Models

Definition

- Incidence analysis: effect of policies on **distribution** of economic pie
- Efficiency or deadweight cost: effect of policies on **size** of the pie
- Focus in efficiency analysis is on quantities, not prices

References

- Auerbach (1985) handbook chapter
- Atkinson and Stiglitz, Chapters 6 and 7
- Chetty, Looney, Kroft (AER 2009)
- Chetty (Ann Review 2009)
- Hines (1999) for historical perspective
- For background on price theory concepts see: Mas-Colell, Whinston, Green Chapter 3 or Deaton and Muellbauer

Efficiency Cost: Introduction

- Government raises taxes for one of two reasons:

  1. To raise revenue to finance public goods
  2. To redistribute income

- But to generate \$1 of revenue, welfare of those taxed is reduced by more than \$1 because the tax distorts incentives and behavior

- Core theory of public finance: how to implement policies that minimize these efficiency costs

  - This basic framework for optimal taxation is adapted to study transfer programs, social insurance, etc.
  - Start with positive analysis of how to measure efficiency cost of a given tax system

Marshallian Surplus: Assumptions

- Most basic analysis of efficiency costs is based on Marshallian surplus
- Two critical assumptions:
  1. Quasilinear utility (no income effects)
  2. Competitive production

Partial Equilibrium Model: Setup

- Two goods: x and y
- Consumer has wealth Z, utility $u(x) + y$, and solves

$$\max_{x,y} u(x) + y \text{ s.t. } (p+t)x(p+t, Z) + y(p+t, Z) = Z$$

- Firms use $c(S)$ units of the numeraire y to produce S units of x
- Marginal cost of production is increasing and convex:

$$c'(S) > 0 \text{ and } c''(S) \geq 0$$

- Firm's profit at pretax price p and level of supply S is

$$pS - c(S)$$

Model: Equilibrium

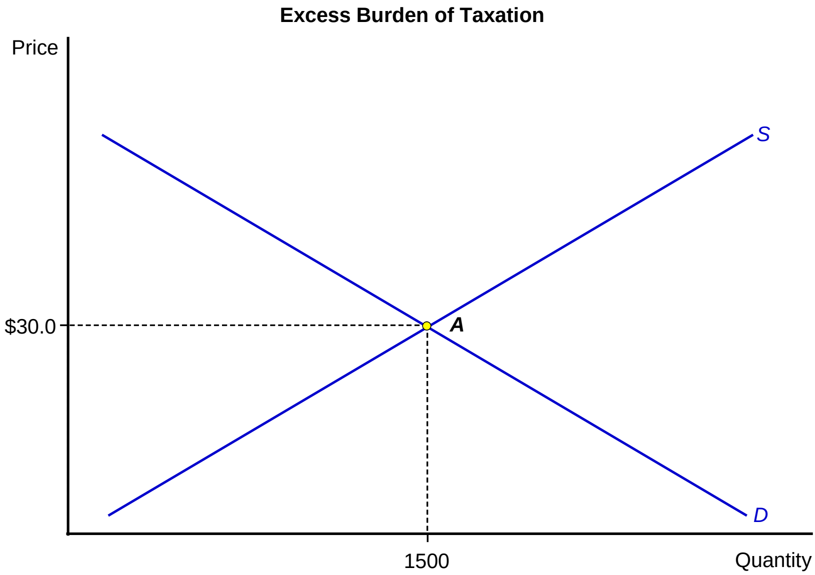
- With perfect optimization, supply fn for x is implicitly defined by the marginal condition

$$p = c'(S(p))$$

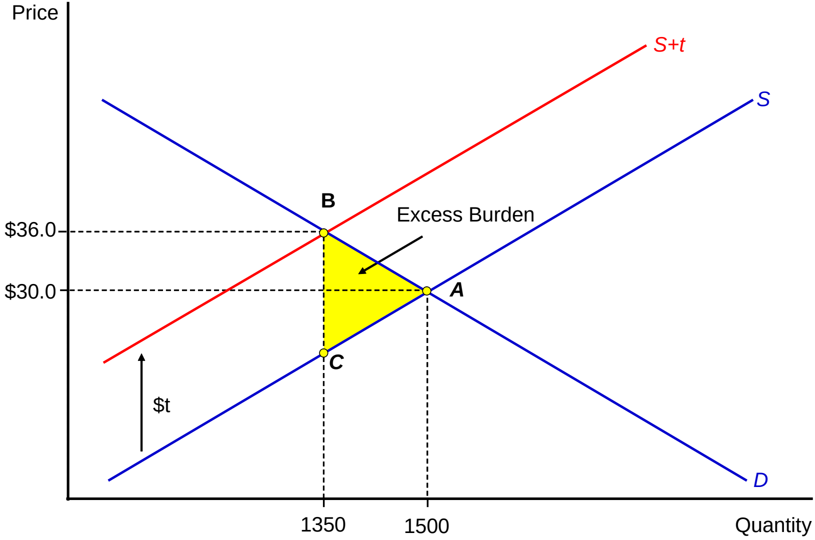
- Let $\eta_S = p\frac{S'}{S}$ denote the price elasticity of supply
- Let Q denote equilibrium quantity sold of good x
- Q satisfies:

$$Q(t) = D(p+t) = S(p)$$

- Consider effect of introducing a small tax $d\tau > 0$ on Q and surplus


Excess Burden of Taxation
Price
S
$30.0
A
D
1500
Quantity

Excess Burden of Taxation



Efficiency Cost: Marshallian Surplus

There are three empirically implementalbe ways to measure the area of the triangle:

1. In terms of supply and demand elasticities
2. In terms of total change in equilibrium quantity caused by tax
3. In terms of change in government revenue

Efficiency Cost: Marshallian Surplus

1. In terms of supply and demand elasticities:

$$EB = \frac{1}{2} dQ d\tau$$

$$EB = \frac{1}{2} S'(p) dp d\tau = (1/2)(pS'/S)(S/p)\frac{\eta_D}{\eta_S - \eta_D} d\tau^2$$

$$EB = \frac{1}{2}\frac{\eta_S \eta_D}{\eta_S - \eta_D} pQ(\frac{d\tau}{p})^2$$

- Note: second line uses incidence formula $dp = (\frac{\eta_D}{\eta_S - \eta_D})d\tau$
- Tax revenue $R = Q d\tau$
- Useful expression is deadweight burden per dollar of tax revenue:

$$\frac{EB}{R} = \frac{1}{2}\frac{\eta_S \eta_D}{\eta_S - \eta_D}\frac{d\tau}{p}$$

Efficiency Cost: Marshallian Surplus

2. In terms of total change in equilibrium qty caused by tax

- Define $\eta_Q = -\frac{dQ}{d\tau}\frac{p_0}{Q}$
- η_Q: effect of a 1% increase in price via a tax change on equilibrium quantity, taking into account the endogenous price change
- This is the coefficient β in a regression of the form

$$\log Q = \alpha + \beta \frac{\tau}{p_0} + \varepsilon$$

- Identify β using exogenous variation in τ. Then:

$$\begin{aligned} EB &= -(1/2)\frac{dQ}{d\tau}d\tau d\tau \\ &= -(1/2)\frac{dQ}{d\tau}(\frac{p}{Q})(\frac{Q}{p})d\tau d\tau \\ &= (1/2)\eta_Q pQ(\frac{d\tau}{p})^2 \end{aligned}$$

Efficiency Cost: Marshallian Surplus

3. In terms of change in government revenue

- Observe that with an initial tax rate of τ,

$$EB = \frac{1}{2}\frac{\eta_S\eta_D}{\eta_S - \eta_D}pQ(\frac{\tau}{p})^2$$

- Marginal excess burden (to a first-order approximation) is:

$$\frac{\partial EB}{\partial \tau} = \frac{\eta_S\eta_D}{\eta_S - \eta_D}Q\frac{\tau}{p} = \eta_Q Q\frac{\tau}{p}$$

- First-order approx includes only loss in govt revenue due to behavioral response
  - Rectangle in the Harberger trapezoid, proportional to τ
  - Does not include the second-order term (proportional to τ^2)

Efficiency Cost: Marshallian Surplus

- Alternative representation for marginal excess burden can be obtained using data on government budget $R(\tau) = Q\tau$

- MEB equals the difference between the "mechanical" revenue gain and the actual revenue gain

  - Mechanical revenue gain: $\frac{\partial R}{\partial \tau}|_Q = Q$
  - Actual revenue gain: $\frac{\partial R}{\partial \tau} = Q + \tau \frac{dQ}{dp}$

- Difference between mechanical and actual revenue gain:

$$
\begin{aligned}
\frac{\partial R}{\partial \tau}|_Q - \frac{dR}{d\tau} &= Q - [Q + \tau \frac{dQ}{dp}] = \tau \frac{dQ}{dp} \\
&= -\tau \frac{dQ}{dt} \frac{p}{Q} \frac{Q}{p} = \frac{\tau}{p} \eta_Q Q = \frac{\partial E}{\partial \tau}
\end{aligned}
$$

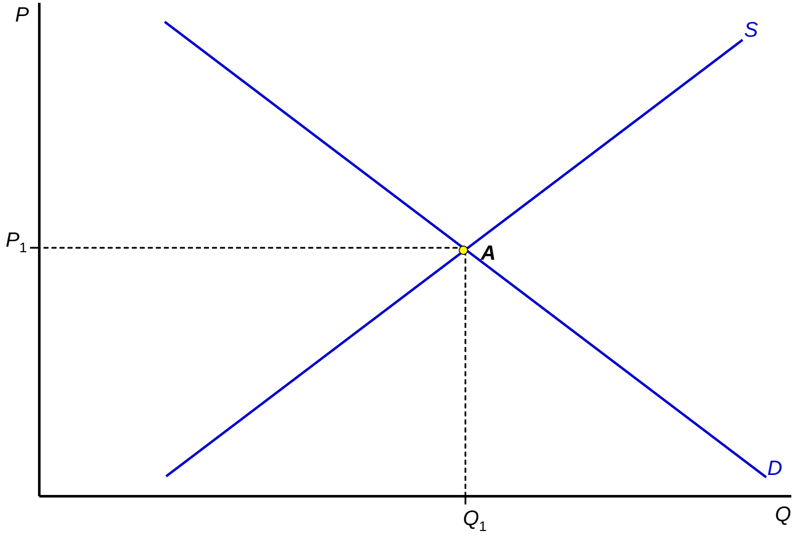
- Intuition: leakage in government revenue measures distortion

Efficiency Cost: Qualitative Properties

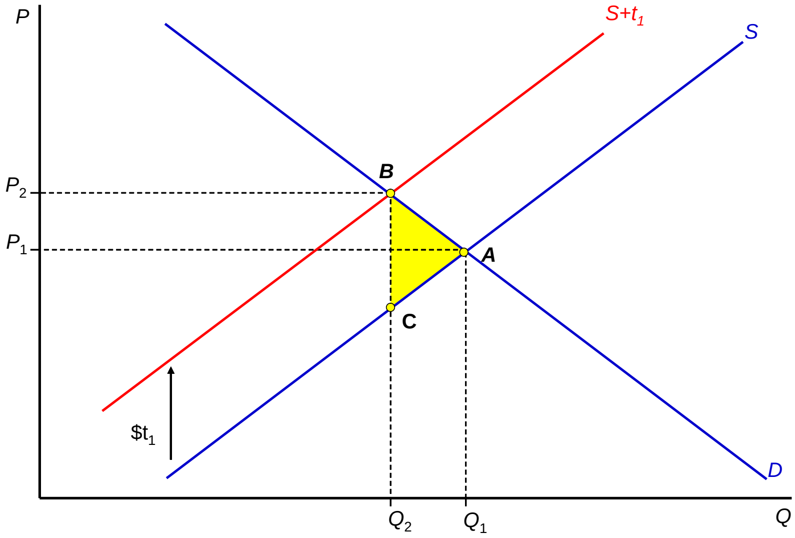
$$EB = \frac{1}{2}\frac{\eta_S \eta_D}{\eta_S - \eta_D} pQ(\frac{d\tau}{p})^2$$

1. Excess burden increases with square of tax rate
2. Excess burden increases with elasticities

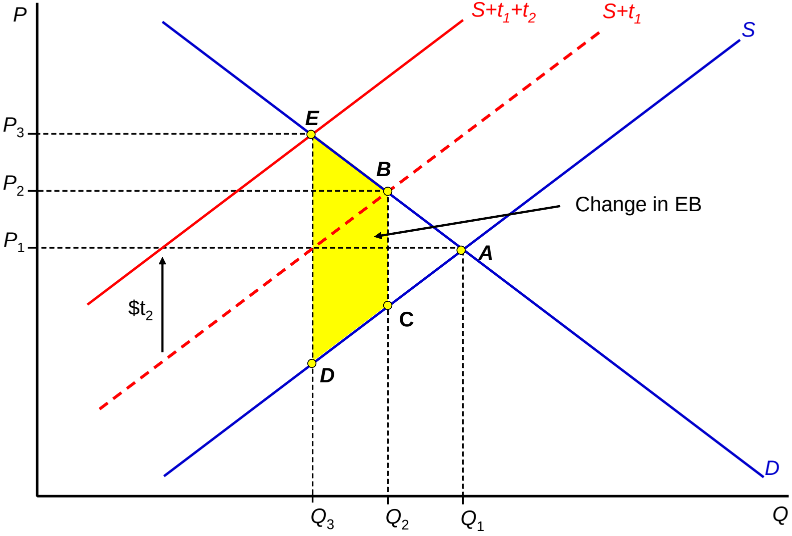
EB Increases with Square of Tax Rate



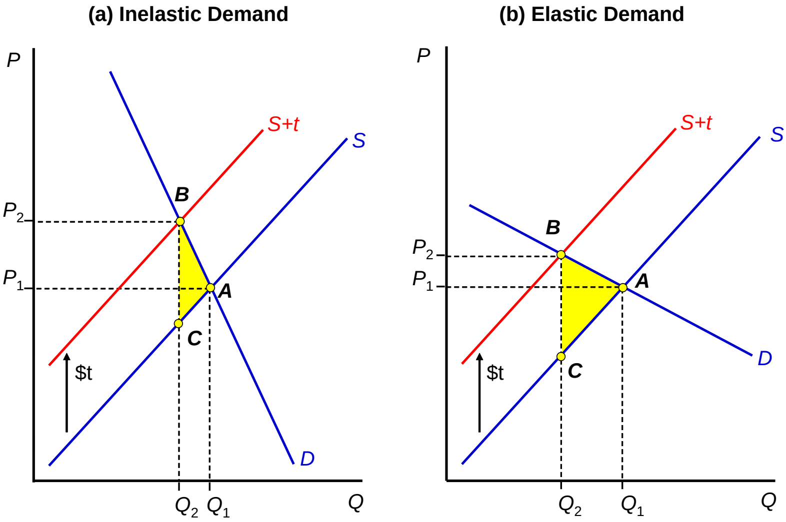
EB Increases with Square of Tax Rate



EB Increases with Square of Tax Rate



Comparative Statics



Tax Policy Implications

- With many goods, formula suggests that the most efficient way to raise tax revenue is:

    1. Tax relatively more the inelastic goods (e.g. medical drugs, food)
    2. Spread taxes across all goods so as to keep tax rates relatively low on all goods (broad tax base)

General Model with Income Effects

- Marshallian surplus is an ill-defined measure with income effects
- Drop quasilinearity assumption and consider an individual with utility

$$u(c_1, .., c_N) = u(c)$$

- Individual program:

$$\max_c u(c) \text{ s.t. } q \cdot c \leq Z$$

where $q = p + t$ denotes vector of tax-inclusive prices and Z is wealth

- Labor can be viewed as commodity with price w and consumed in negative quantity

General Model: Demand and Indirect Utility

- Multiplier on the budget constraint is λ
- First order condition in c_i:

$$u_{c_i} = \lambda q_i$$

- These conditions implicitly define:
  - $c_i(q, Z)$: the Marshallian (or uncompensated) demand function
  - $v(q, Z)$: the indirect utility function

Useful Properties of Demand and Utility

- Multiplier on budget constraint $\lambda = v_Z$ is the marginal utility of wealth
- Give wealth grant of dZ to consumer:

$$du = \sum_i u_{c_i} dc_i = \lambda \sum_i q_i dc_i = \lambda dZ$$

- Roy's identity: $v_{q_i} = -\lambda c_i$
  - Welfare effect of a price change dq_i same as reducing wealth by:

$$dZ = c_i dq_i$$

  - By Envelope Thm., adjustment of c_j does not produce a 1st order welfare effect

Path Dependence Problem

- Initial price vector q^0
- Taxes levied on goods $\rightarrow$price vector now q^1
- Change in Marshallian surplus is defined as the line integral:

$$CS = \int_{q^0}^{q^1} c(q, Z) dq$$

- With one price changing, this is area under the demand curve
- Problem: CS is **path dependent** with > 1 price changes
- Consider change from q^0 to $\tilde{q}$ and then $\tilde{q}$ to q^1:

$$CS(q^0 \rightarrow \tilde{q}) + CS(\tilde{q} \rightarrow q^1) \neq CS(q^0 \rightarrow q^1)$$

Path Dependence Problem

- Example of path dependence with taxes on two goods:

$$CS_1 = \int_{q_1^0}^{q_1^1} c_1(q_1, q_2^0, Z)dq_1 + \int_{q_2^0}^{q_2^1} c_2(q_1^1, q_2, Z)dq_2 \quad (1)$$

$$CS_2 = \int_{q_2^0}^{q_2^1} c_2(q_1^0, q_2, Z)dq_2 + \int_{q_1^0}^{q_1^1} c_1(q_1, q_2^1, Z)dq_1 \quad (2)$$

- For $CS_1 = CS_2$, need $\frac{dc_2}{dq_1} = \frac{dc_1}{dq_2}$

- With income effects, this symmetry condition is not satisfied in general

Consumer Surplus: Conceptual Problems

- Path-dependence problem reflects the fact that consumer surplus is an ad-hoc measure
  - It is not derived from utility function or a welfare measure
- Question of interest: how much utility is lost because of tax beyond revenue transferred to government?
- Need units to measure "utility loss"
- Introduce expenditure function to translate the utility loss into dollars (money metric)

Expenditure Function

- Fix utility at U and prices at q
- Find bundle that minimizes cost to reach U for q:

$$e(q, U) = \min_c q \cdot c \text{ s.t. } u(c) \geq U$$

- Let μ denote multiplier on utility constraint
- First order conditions given by:

$$q_i = \mu u_{c_i}$$

- These generate Hicksian (or compensated) demand fns:

$$c_i = h_i(q, u)$$

- Define individual's loss from tax increase as

$$e(q^1, u) - e(q^0, u)$$

- Single-valued function $\rightarrow$ coherent measure of welfare cost, no path dependence

Compensating and Equivalent Variation

- But where should u be measured?
- Consider a price change from q^0 to q^1
- Initial utility:

$$u^0 = v(q^0, Z)$$

- Utility at new price q^1:

$$u^1 = v(q^1, Z)$$

- Two concepts: compensating (CV) and equivalent variation (EV) use u^0 and u^1 as reference utility levels

Compensating Variation

- Measures utility at initial price level (u^0)
- Amount agent must be compensated in order to be indifferent about tax increase

$$CV = e(q^1, u^0) - e(q^0, u^0) = e(q^1, u^0) - Z$$

- How much compensation is needed to reach original utility level at *new* prices?
- CV is amount of ex-post cost that must be covered by government to yield same *ex-ante* utility:

$$e(q^0, u^0) = e(q^1, u^0) - CV$$

Equivalent Variation

- Measures utility at new price level
- Lump sum amount agent willing to pay to avoid tax (at pre-tax prices)

$$EV = e(q^1, u^1) - e(q^0, u^1) = Z - e(q^0, u^1)$$

- EV is amount extra that can be taken from agent to leave him with same *ex-post* utility:

$$e(q^0, u^1) + EV = e(q^1, u^1)$$

Efficiency Cost with Income Effects

- Goal: derive empirically implementable formula analogous to Marshallian EB formula in general model with income effects

- Existing literature assumes either

  1. Fixed producer prices and income effects
  2. Endogenous producer prices and quasilinear utility

- With both endogenous prices and income effects, efficiency cost depends on how profits are returned to consumers

- Formulas are very messy and fragile (Auerbach section 3.2)

Efficiency Cost Formulas with Income Effects

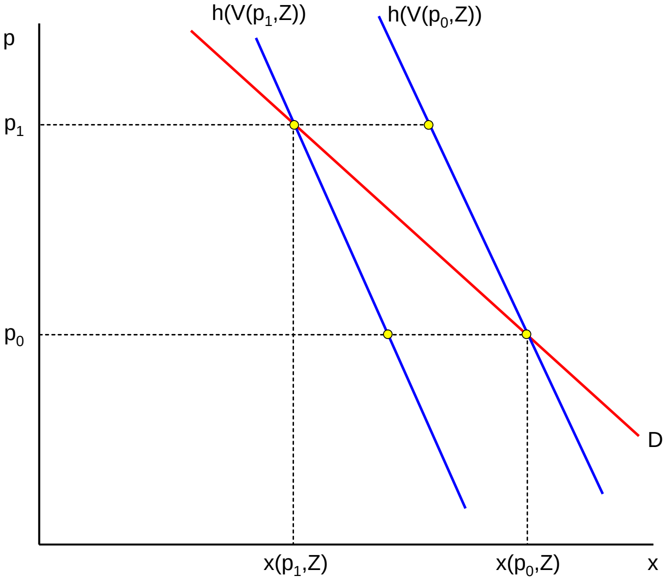
- Derive empirically implementable formulas using Hicksian demand (EV and CV)
- Assume p is fixed $\longrightarrow$ flat supply, constant returns to scale
- The envelope thm implies that $e_{q_i}(q,u) = h_i$, and so:

$$e(q^1,u) - e(q^0,u) = \int_{q^0}^{q^1} h(q,u)dq$$

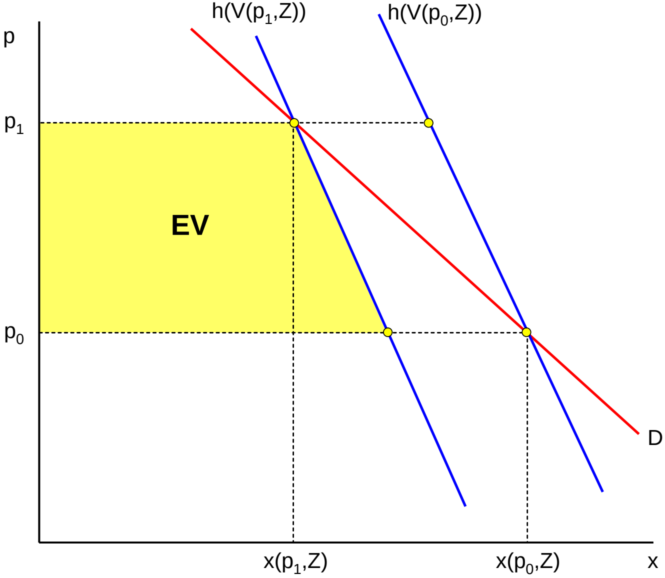
- If only one price is changing, this is the area under the Hicksian demand curve for that good
- Note that optimization implies that

$$h(q, v(q, Z)) = c(q, Z)$$

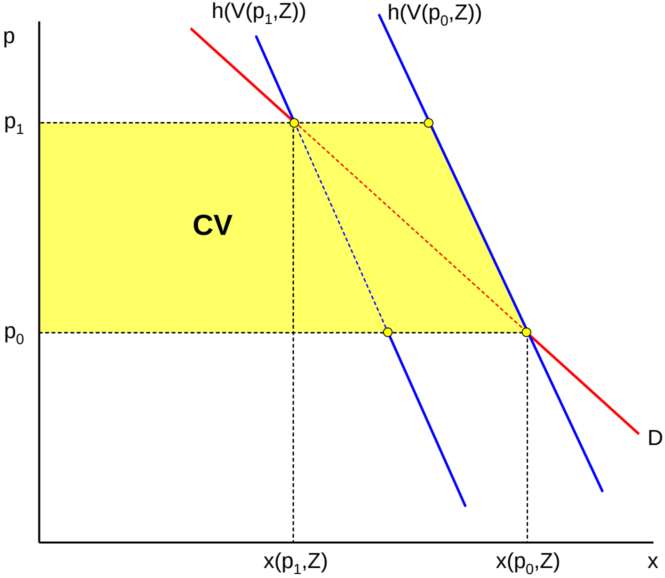
Compensating vs. Equivalent Variation



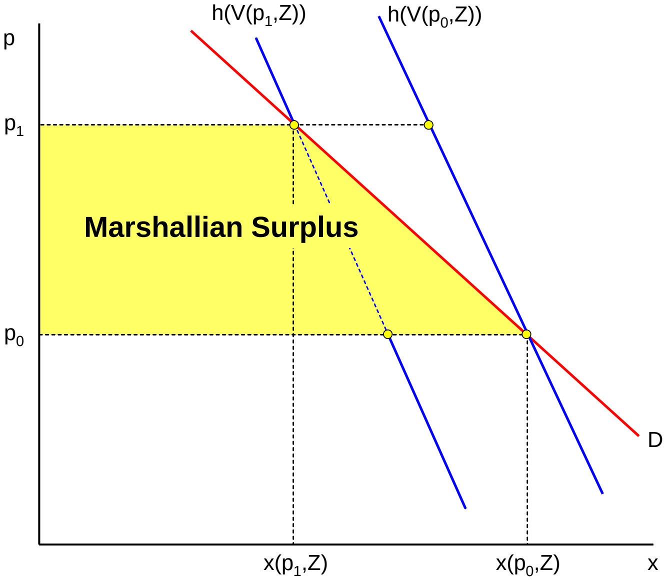
Compensating vs. Equivalent Variation



Compensating vs. Equivalent Variation



Marshallian Surplus



Path Independence of EV, CV

- With one price change:

$$EV < \text{ Marshallian Surplus} < \ CV$$

  but this is not true in general

- No path dependence problem for EV, CV measures with multiple price changes
  - Slutsky equation:

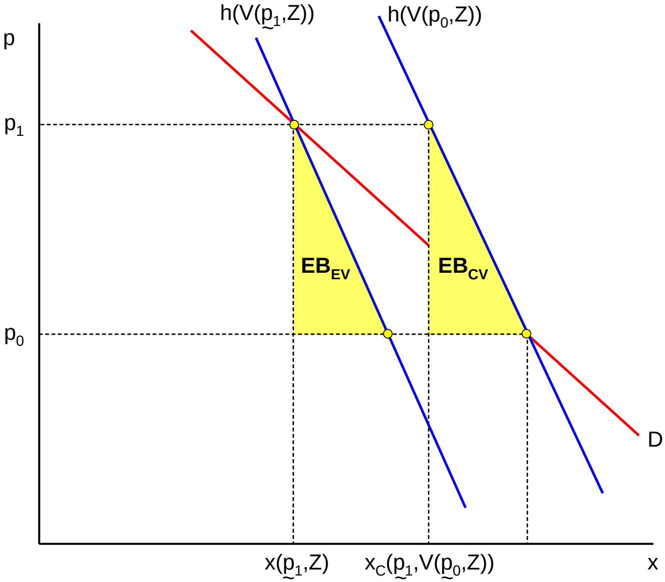
$$\underbrace{\frac{\partial h_i}{\partial q_j}}_{\text{Hicksian Slope}} = \underbrace{\frac{\partial c_i}{\partial q_j}}_{\text{Marshallian Slope}} + \underbrace{c_j \frac{\partial c_i}{\partial Z}}_{\text{Income Effect}}$$

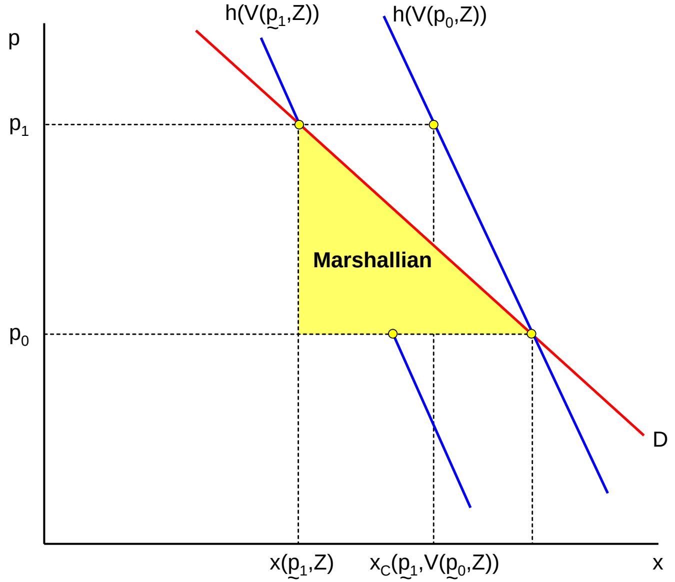
  - Optimization implies Slutsky matrix is symmetric: $\frac{\partial h_i}{\partial q_j} = \frac{\partial h_j}{\partial q_i}$
- Therefore the integral $\int_{q^0}^{q^1} h(q,u)dq$ is path independent

Excess Burden

- Deadweight burden: change in consumer surplus less tax paid
- Equals what is lost in excess of taxes paid
- Two measures, corresponding to EV and CV:

$$EB(u^1) = EV - (q^1 - q^0)h(q^1, u^1) \text{ [Mohring 1971]}$$
$$EB(u^0) = CV - (q^1 - q^0)h(q^1, u^0) \text{ [Diamond and McFadden 1974]}$$


h(V(p_1,Z))
h(V(p_0,Z))
p
p_1
p_0
EB_EV
EB_CV
D
x(p_1,Z)
x_C(p_1,V(p_0,Z))
x


h(V(p1,Z))
h(V(p0,Z))
p
p1
p0
Marshallian
D
x(p1,Z)
xC(p1,V(p0,Z))
x

Excess Burden

- In general, CV and EV measures of EB will differ
- Marshallian measure overstates excess burden because it includes income effects
  - Income effects are not a distortion in transactions
  - Buying less of a good due to having less income is not an efficiency loss; no surplus foregone b/c of transactions that do not occur
- Chipman and Moore (1980): CV = EV = Marshallian DWL only with quasilinear utility

Implementable Excess Burden Formula

- Consider increase in tax τ on good 1 to $\tau + \Delta\tau$
- No other taxes in the system
- Recall the expression for EB:

$$EB(\tau) = [e(p+\tau, U) - e(p, U)] - \tau h_1(p+\tau, U)$$

- Second-order Taylor expansion:

$$\begin{aligned} MEB &= EB(\tau + \Delta\tau) - EB(\tau) \\ &\simeq \frac{dEB}{d\tau}(\Delta\tau) + \frac{1}{2}(\Delta\tau)^2 \frac{d^2 EB}{d\tau^2} \end{aligned}$$

Harberger Trapezoid Formula

$$\begin{aligned}\frac{dEB}{d\tau} &= h_1(p+\tau,U) - \tau\frac{dh_1}{d\tau} - h_1(p+\tau,U)\\ &= -\tau\frac{dh_1}{d\tau}\\ \frac{d^2EB}{d\tau^2} &= -\frac{dh_1}{d\tau} - \tau\frac{d^2h_1}{d\tau^2}\end{aligned}$$

- Standard practice in literature: assume $\frac{d^2h_1}{d\tau^2}$ (linear Hicksian); not necessarily well justified because it does not vanish as $\Delta\tau \longrightarrow 0$

$$\Rightarrow MEB = -\tau\Delta\tau\frac{dh_1}{d\tau} - \frac{1}{2}\frac{dh_1}{d\tau}(\Delta\tau)^2$$

- Formula equals area of “Harberger trapezoid” using Hicksian demands

Harberger Formula

- Without pre-existing tax, obtain "standard" Harberger formula:

$$EB = -\frac{1}{2}\frac{dh_1}{d\tau}(\Delta\tau)^2$$

- Observe that first-order term vanishes when $\tau = 0$
- A new tax has second-order deadweight burden (proportional to $\Delta\tau^2$ not $\Delta\tau$)
- Bottom line: need compensated (substitution) elasticities to compute *EB*, not uncompensated elasticities
- Empirically, need estimates of income and price elasticities

Excess Burden with Taxes on Multiple Goods

- Previous formulas apply to case with tax on one good
- With multiple goods and fixed prices, excess burden of introducing a tax τ_k

$$EB = -\frac{1}{2}\tau_k^2 \frac{dh_k}{d\tau_k} - \sum_{i \neq k} \tau_i \tau_k \frac{dh_i}{d\tau_k}$$

- Second-order effect in own market, first-order effect from other markets with pre-existing taxes
- Hard to implement because we need all cross-price elasticities
- Complementarity between goods important for excess burden calculations
- Ex: with an income tax, minimize total DWL tax by taxing goods complementary to leisure (Corlett and Hague 1953)

Goulder and Williams 2003

- Show that ignoring cross effects by using one-good formula can be very misleading

- Differentiate multiple-good Harberger formula w.r.t. τ_k:

$$\frac{dEB}{d\tau_k} = -\tau_k \frac{dh_k}{d\tau_k} - \sum_{i \neq k} \tau_i \frac{dh_i}{d\tau_k}$$

- If τ_k is small (e.g. gas tax), what matters is purely distortion in other markets, e.g. labor supply

- As $\tau_k \to 0$, error in single-market formula approaches ∞

Goulder and Williams: Assumptions

- Make multiple-goods formula empirically implementable by making 3 assumptions/approximations:

  1. No income effects
  2. Ignore interactions with commodities other than labor (other taxes are small)
  3. Assume good is of "average" substitutability with labor: cross partial $\frac{\partial l}{\partial \tau_k}$ equals mean cross-partial across consumption goods

Goulder and Williams Formula

- Obtain following formula for marginal excess burden of raising tax on good k:

$$\frac{dEB}{d\tau_k} = \frac{\tau_k Q_k}{p_k}\eta_k - \frac{\tau_L L}{p_k}\eta_L s_k$$

  - τ_k, p_k, and Q_k are the tax, price, and quantity consumed of good k
  - η_k and η_L are own-price elasticity of good k and labor
  - $s_k = \frac{P_k Q_K}{wl(1-\tau_L)}$ is budget share of good k

- Only need estimates of own-price elasticities to implement this formula

- Why? Price increase in all consumption goods has the same effect on labor supply as an increase in tax on labor:

$$(1+t)\sum_k p_k c_k = wl$$

- Equivalence between consumption tax and labor income tax

Goulder and Williams Formula

- Rank goods according to complementarity with labor (i.e. cross-partial $\frac{dl}{d\tau_k}$)
- Find good at the mean level of $\frac{dl}{d\tau_k}$
- A tax increase on this good has same effect as an increase in sales tax t on all consumption goods scaled down by s_k
- Therefore cross-elasticity is equivalent to labor-supply elasticity times s_k
- Labor supply elasticity η_L sufficient to calculate cross-elasticity for good that has "average" level of substitutability

Goulder and Williams Results

- Calibrate formula using existing elasticity estimates
- Result: DWL of taxing goods such as gasoline is underestimated by a factor of 10 in practice because of income tax
- Caveat: is their approach and conclusion valid if there are salience effects?

Hausman 1981: Exact Consumer Surplus

- Harberger formulas: empirically implementable, but approximations (linearity, ignore cross-effects)
- Alternative approach: full structural estimation of demand model
- Start from observed market demand functions, finding the best fit
- Estimate regression of the form:

$$c(q, Z) = \gamma + \alpha q + \delta Z$$

- Then integrate to recover underlying indirect utility function $v(q, Z)$
- Inverting yields expenditure function $e(q, u)$; now compute "exact" EB
- Parametric approach: Hausman (AER 1981); non-parametric approach: Hausman and Newey (ECMA 1995)

Harberger vs. Hausman Approach

- Underscores broader difference between structural and quasi-experimental methodologies
- Public finance literature focuses on deriving "sufficient statistic" formulas that can be implemented using quasi-experimental techniques
- In IO, macro, trade, structural methods more common
- Now develop distinction between structural and sufficient statistic approaches to welfare analysis in a simple model of taxation
  - No income effects (quasilinear utility)
  - Constant returns to production (fixed producer prices)

Sufficient Statistics vs Structural Methods

- N goods: $x = (x_1, ..., x_N)$; Prices: $(p_1, ...p_N)$; Z = wealth
- Normalize $p_N = 1$ (x_N is numeraire)
- Government levies a tax t on good 1
- Individual takes t as given and solves

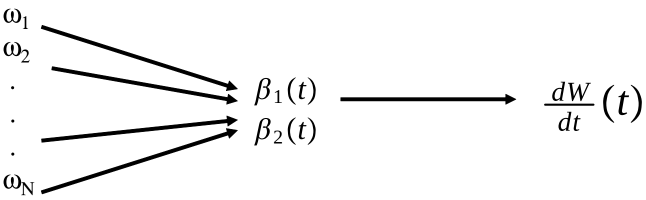
$$\max u(x_1, ..., x_{N-1}) + x_N \text{ s.t. } (p_1 + t)x_1 + \sum_{i=2}^{N} p_i x_i = Z$$

- To measure EB of tax, define social welfare as sum of individual's utility and tax revenue:

$$W(t) = \{\max_x u(x_1, ..., x_{N-1}) + Z - (p_1 + t)x_1 - \sum_{i=2}^{N-1} p_i x_i\} + tx_1$$

- Goal: measure $\frac{dW}{dt}$ = loss in social surplus caused by tax change

Primitives	Sufficient Stats.	Welfare Change



ω=preferences, constraints	$\beta = f(\omega, t)$ $y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon$	dW/dt used for policy analysis
ω not uniquely identified	β identified using program evaluation	

Source: Chetty (2009)

Sufficient Statistics vs Structural Methods

- Structural method: estimate N good demand system, recover u
  - Ex: use Stone-Geary or AIDS to recover preference parameters; then calculate "exact consumer surplus" as in Hausman (1981)
- Alternative: Harberger's deadweight loss triangle formula
  - Private sector choices made to maximize term in red (private surplus)

$$W(t) = \{\max_x u(x_1, ..., x_{N-1}) + Z - (p_1 + t)x_1 - \sum_{i=2}^{N-1} p_i x_i\} + tx_1$$

  - Envelope conditions for $(x_1, ..., x_N)$ allow us to ignore behavioral responses $(\frac{dx_i}{dt})$ in term in red, yielding

$$\frac{dW}{dt} = -x_1 + x_1 + t\frac{dx_1}{dt} = t\frac{dx_1}{dt}$$

$\rightarrow \frac{dx_1}{dt}$ is a "sufficient statistic" for calculating $\frac{dW}{dt}$

Heterogeneity

- Benefit of suff stat approach particularly evident with heterogeneity
- K agents, each with utility $u_k(x_1, ..., x_{N-1}) + x_N$
- Social welfare function under utilitarian criterion:

$$\begin{aligned} W(t) &= \{\max_x \sum_{k=1}^{K} [u_k(x_1^k, ..., x_{N-1}^k) + Z \\ &\quad -(p_1 + t)x_1^k - \sum_{i=2}^{N-1} p_i x_i^k]\} + \sum_{k=1}^{K} t x_1^k \end{aligned}$$

- Structural method: estimate demand systems for all agents
- Sufficient statistic formula is unchanged—still need only slope of aggregate demand $\frac{dx_1}{dt}$

$$\frac{dW}{dt} = -\sum_{k=1}^{K} x_1^k + \sum_{k=1}^{K} x_1^k + t\frac{d\sum_{k=1}^{K} x_1^k}{dt} = t\frac{dx_1}{dt}$$

Discrete Choice Model

- Two good model
- Agents have value V_k for good 1; can either buy or not buy
- Let $F(V)$ denote distribution of valuations
- Utility of agent k is

$$V_k x_1 + Z - (p+t)x_1$$

- Social welfare:

$$\begin{aligned} W(t) &= \{\int_{V_k} \max_{x_1^k}[V_k x_1^k + Z - (p_1+t)x_1^k] dF(V_k)\} \\ &\quad + \int_{V_k} t x_1^k dF(V_k) \end{aligned}$$

- This problem is not smooth at individual level, so cannot directly apply envelope thm. as stated

Discrete Choice Model

- Recast as planner's problem choosing threshold above which agents are allocated good 1:

$$
\begin{aligned}
W(t) &= \left\{ \max_{\bar{V}} \int_{\bar{V}}^{\infty} \left[V_k - (p_1 + t) \right] dF(V_k) + Z \right\} \\
&\quad + t \int_{\bar{V}}^{\infty} dF(V_k)
\end{aligned}
$$

- Again obtain Harberger formula as a fn of slope of aggregate demand curve $\frac{dx_1}{dt}$:

$$
\begin{aligned}
\frac{dW}{dt} &= -\left(1 - F\left(\bar{V}\right)\right) + \left(1 - F\left(\bar{V}\right)\right) + t \frac{d \int_{\bar{V}}^{\infty} dF(V_k)}{dt} \\
\Rightarrow \frac{dW}{dt} &= t \frac{dx_1}{dt}
\end{aligned}
$$

Economic Intuition for Robustness of Harberger Result

- Deadweight loss is fully determined by difference between marginal willingness to pay for good x_1 and its cost (p_1)
  - Recovering marginal willingness to pay requires an estimate of the slope of the demand curve because it coincides with marginal utility:

$$p = u'(x(p))$$

  - Slope of demand is therefore sufficient to infer efficiency cost of a tax, without identifying rest of the model

Efficiency Cost: Applications

1. [**Income Taxation**] Feldstein; Chetty; Gorodnichenko et al.
2. [**Housing Subsidy**] Poterba
3. [**Diesel Fuel Taxation**] Marion and Muehlegger

Feldstein 1995, 1999

- Following Harberger, large literature in labor estimated effect of taxes on hours worked to assess efficiency costs of taxation
- Feldstein observed that labor supply involves multiple dimensions, not just choice of hours: training, effort, occupation
- Taxes also induce inefficient avoidance/evasion behavior
- Structural approach: account for each of the potential responses to taxation separately and then aggregate
- Feldstein's alternative: elasticity of taxable income with respect to taxes is a sufficient statistic for calculating deadweight loss

Feldstein Model: Setup

- Government levies linear tax t on reported taxable income
- Agent makes N labor supply choices: $l_1, ... l_N$
- Each choice l_i has disutility $\psi_i(l_i)$ and wage w_i
- Agents can shelter $\$e$ of income from taxation by paying cost $g(e)$
- Taxable Income (TI) is

$$TI = \sum_{i=1}^{N} w_i l_i - e$$

- Consumption is given by taxed income plus untaxed income:

$$c = (1-t)TI + e$$

Feldstein Taxable Income Formula

- Agent's utility is quasi-linear in consumption:

$$u(c, e, l) = c - g(e) - \sum_{i=1}^{N} \psi_i(l_i)$$

- Social welfare:

$$W(t) = \{(1 - t)TI + e - g(e) - \sum_{i=1}^{N} \psi_i(l_i)\} + tTI$$

- Differentiating and applying envelope conditions for l_i ($(1 - t)w_i = \psi_i'(l_i)$) and e ($g'(e) = t$) implies

$$\frac{dW}{dt} = -TI + TI - t\frac{dTI}{dt} = t\frac{dTI}{dt}$$

- Intuition: marginal social cost of reducing earnings through each margin is equated at optimum $\rightarrow$ irrelevant what causes change in TI

Taxable Income Formula

- Simplicity of identification in Feldstein's formula has led to a large literature estimating elasticity of taxable income
- But since primitives are not estimated, assumptions of model used to derive formula are never tested
- Chetty (2009) questions validity of assumption that $g'(e) = t$
  - Costs of some avoidance/evasion behaviors are transfers to other agents in the economy, not real resource costs
  - Ex: cost of evasion is potential fine imposed by government

Chetty Transfer Cost Model: Setup

- Individual chooses e (evasion/shifting) and l (labor supply) to

$$\begin{aligned}\max_{e,l} u(c,l,e) &= c - \psi(l) \\ \text{s.t. } c &= y + (1-t)(wl - e) + e - z(e)\end{aligned}$$

- Social welfare is now:

$$\begin{aligned}W(t) = &\{y + (1-t)(wl - e) + e \\ &- z(e) - \psi(l)\} \\ &+ z(e) + t(wl - e)\end{aligned}$$

- Difference: $z(e)$ now appears twice in SWF, with opposite signs

Excess Burden with Transfer Costs

- Let $LI = wl$ be the total (pretax) earned income and $TI = wl - e$ denote taxable income

- Exploit the envelope condition for term in curly brackets:

$$\begin{aligned}
\frac{dW}{dt} &= -(wl - e) + (wl - e) + \frac{dz}{de}\frac{de}{dt} + t\frac{d[wl - e]}{dt} \\
&= t\frac{dTI}{dt} + \frac{dz}{de}\frac{de}{dt} \\
&= t\frac{dLI}{dt} - t\frac{de}{dt} + \frac{dz}{de}\frac{de}{dt}
\end{aligned}$$

- First-order condition for individual's choice of e:

$$\begin{aligned}
t &= \frac{dz}{de} \\
&\Rightarrow \frac{dW}{dt} = t\frac{dLI}{dt} \qquad (1)
\end{aligned}$$

- Intuition: MPB of raising e by \$1 (saving \$$t$) equals MPC

Chetty (2009) Formula

- With both transfer cost $z(e)$ and resource cost $g(e)$ of evasion:

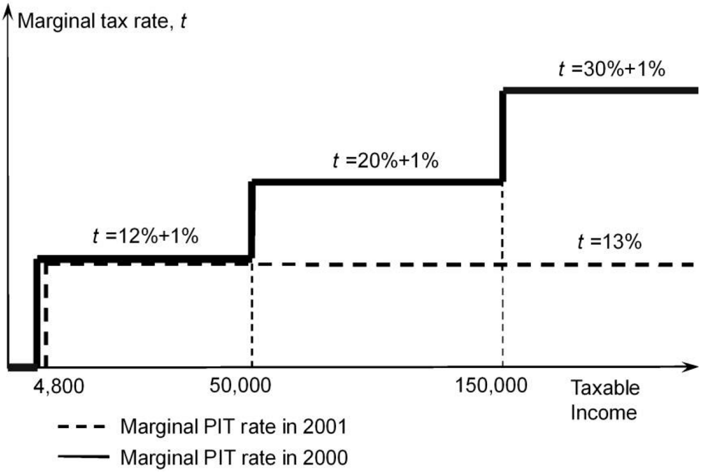
$$\begin{aligned}\frac{dW}{dt} &= t\frac{dLI}{dt} - g'(e)\frac{de}{dt} \\ &= t\{\mu\frac{dTI}{dt} + (1-\mu)\frac{dLI}{dt}\} \\ &= -\frac{t}{1-t}\{\mu TI\varepsilon_{TI} + (1-\mu)wl\varepsilon_{LI}\}\end{aligned}$$

- EB depends on weighted average of taxable income (ε_{TI}) and total earned income elasticities (ε_{LI})
  - Practical importance: even though reported taxable income is highly sensitive to tax rates for rich, efficiency cost may not be large!
- Most difficult parameter to identify: weight μ, which depends on marginal resource cost of sheltering, $g'(e)$

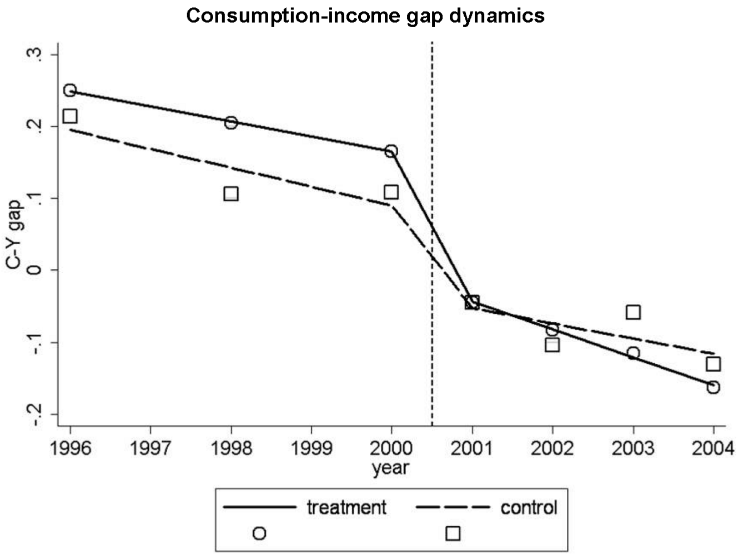
Gorodnichenko, Martinez-Vazquez, and Peter 2009

- Estimate ε_{LI} and ε_{TI} to implement formula that permits transfer costs
- Insight: consumption data can be used to infer ε_{LI}
- Estimate effect of 2001 flat tax reform in Russia on gap between taxable income and consumption, which they interpret as evasion

Marginal personal income tax rate before and after the reform



Source: Gorodnichenko, Martinez-Vazquez, and Peter 2009



Source: Gorodnichenko, Martinez-Vazquez, and Peter 2009

Gorodnichenko et al: Results

- Taxable income elasticity $\frac{dTI}{dt}$ is large, whereas labor income elasticity $\frac{dLI}{dt}$ is not

  $\longrightarrow$ Feldstein's formula overestimates the efficiency costs of taxation relative to more general measure for "plausible" $g'(e)$

- Question: could $g'(e)$ be estimated from consumption data itself?

Poterba 1992

- Estimates efficiency cost of subsidy for housing in the U.S. from mortgage interest deduction
- First need to define “cost” of owning \$1 of housing
- Definition: “user cost” – measures opportunity cost of living in home
- Could rent the house to someone else at percentage rate

$$r = \frac{\text{Rent}}{\text{Property Value}}$$

- With marginal income tax rate τ and nominal interest i, net user cost taking into account mortgage deduction is

$$c = r - \tau \times i$$

Porterba 1992

- Poterba first calculates changes in user cost over 1980s
- Tax reform in 1986 lowered tax rates for high income and raised user cost of housing sharply
  - Prior to 1986: very high tax rates on high incomes (60%)
  - In 1990, only 28%
- Nearly tripled the cost of housing

Variable	$30,000	$50,000	$250,000
User cost (percentage):			
1980	10.6	9.7	4.3
1985	13.1	11.8	8.4
1990	13.3	11.6	11.6
Uncompensated increase in housing demand (percentage):			
1980	20.8	27.6	67.8
1985	14.8	23.1	45.3
1990	12.4	23.2	23.2
Deadweight loss (1990 dollars):			
1980	137	404	12,262
1985	81	326	6,314
1990	53	326	1,631

Source: Poterba 1992

Porterba 1992

- Calculates compensated elasticity using estimates in literature and Slutsky eqn.

$$\begin{aligned} \text{Rosen (1982):} &\quad \varepsilon_{H,r} = -1 \\ \text{Income elasticity:} &\quad 0.75 \\ \text{Housing share:} &\quad 0.25 \\ \Rightarrow \text{Compensated elasticity:} &\quad -1 + \frac{3}{4} \times \frac{1}{4} \simeq -0.8 \end{aligned}$$

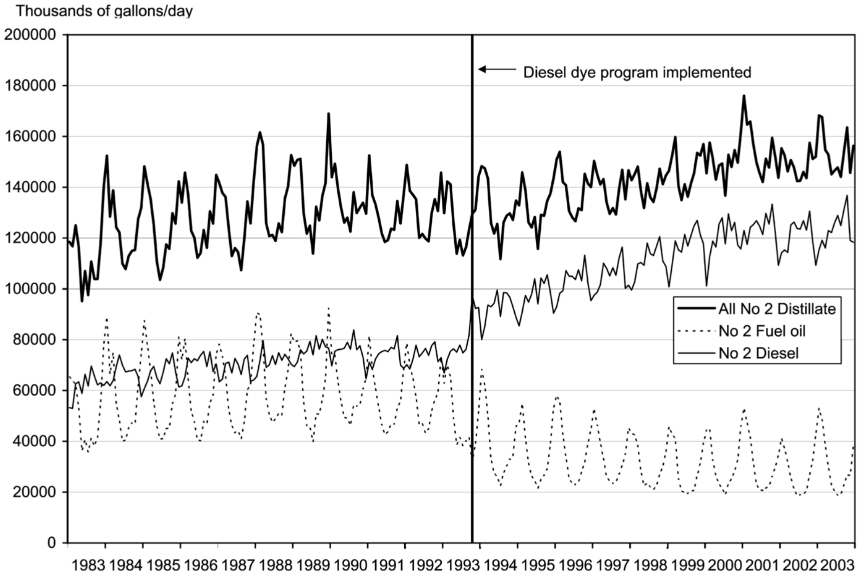
- Intuition for large elasticity: broker calculates "how much house you can afford" if they spend 30% of income
  - Can "afford" more with larger tax subsidy → tax is effectively salient
- Calculates amount of overconsumption of housing and efficiency cost of housing subsidy

Porterba: Results

- Tax reforms in 1980s reduced DWL from $12K to $2K for each household earning $250K
- Still have relatively large inefficiency from subsidizing mortgages
- This is why President Bush's Tax Panel recommended cap or elimination of subsidy for homeownership
- But hard to implement politically

Marion and Muehlegger 2008

- Study deadweight cost from taxing diesel fuels, focusing on evasion
- Diesel fuel used for business purposes (e.g. trucking) is taxed, but residential purposes (e.g. heating homes) is not
- Substantial opportunity to evade tax
- 1993: government added red dye to residential diesel fuel
    - Easy to monitor cheating by opening gas tank of a truck
- First document effect of dye reform on evasion

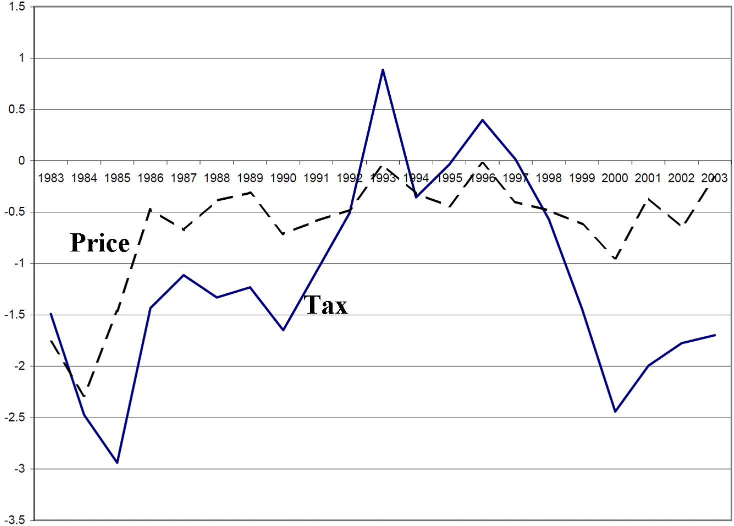


Source: Marion and Muehlleger 2008

Marion and Muehlegger: Excess Burden Calculations

- Use reform to assess deadweight costs of evasion and taxation
  - Harder to evade $\longrightarrow$ elasticity of behavior with respect to tax is much lower after reform
- Estimate price and tax elasticities before and after reform
  - Use cross-state variation in tax rates and price variation from world market
  - Note different interpretation of difference between price and tax elasticities in this study relative to tax salience papers

Price and Tax Elasticities By Year



Source: Marion and Muehlleger 2008

Marion and Muehlegger: Results

- Elasticities imply that 1% increase in tax rate raised revenue by 0.60% before dye reform vs. 0.71% after reform
- Reform reduced deadweight cost of diesel taxation
  - $MDWL = 40$ cents per dollar of revenue raised before dye reform
  - $MDWL = 30$ cents per dollar after reform
- Lesson: Deadweight cost depends not just on preferences but also on enforcement technology
- But again need to think carefully about marginal costs of evasion in this context: social or transfer?

Welfare Analysis in Behavioral Models

- Formulas derived thus far rely critically on full optimization by agents in private sector
- Now consider how efficiency cost calculations can be made in models where agents do not optimize perfectly
- Relates to broader field of behavioral welfare economics
- Focus on two papers here:
  1. Conceptual Issues: Bernheim and Rangel 2009
  2. Applied Welfare Analysis: Chetty, Looney, Kroft 2009

Behavioral Welfare Economics

- Abstractly, effect of policies on welfare are calculated in two steps
  1. Effect of policy on behavior
  2. Effect of change in behavior on utility
- Challenge: identifying (2) when agents do not optimize perfectly
  - How to measure objective function without tools of revealed preference?
  - Danger of paternalism

Behavioral Welfare Economics: Two Approaches

- Approach #1: Build a positive model of deviations from rationality
  - Ex: hyperbolic discounting, bounded rationality, reference dependence
  - Then calculate optimal policy within such models
- Approach #2: Choice-theoretic welfare analysis (Bernheim and Rangel 2009)
  - Do not specify a positive model to rationalize behavior
  - Instead map directly from observed choices to statements about welfare
  - Analogous to "sufficient statistic" approach

Behavioral Welfare Economics: Two Approaches

- Consider three different medicare plans with different copays: L, M, H and corresponding variation in premiums

- We have data from two environments:

  1. On red paper, $H > M > L$
  2. On blue paper, $M > H > L$

Behavioral Welfare Economics: Two Approaches

- Approach 1: build a model of why color affects choice and use it to predict which choice reveals "true" experienced utility
- Approach 2: Yields bounds on optimal policy
  - L cannot be optimal given available data irrespective of positive
  - Optimal copay bounded between M and H
- Key insight: no theory of choice needed to make statements about welfare (do not need to understand why color affects choice).

Bernheim and Rangel 2009: Setup

- Theory that delivers bounds on welfare based purely on choice data
- In standard model, agents choose from a choice set $x \in X$
- Goal of policy is to identify optimal x
- In behavioral models, agents choose from "generalized choice sets" $G = (X, d)$
- d is an "ancillary condition" – something that affects choice behavior but (by assumption) does not affect experienced utility
  - Ex: color of paper, salience, framing, default option

Bernheim and Rangel 2009: Choice Sets

- Let $C(X, d)$ denote choice made in a given GCS
- Choice inconsistency if $C(X, d) \neq C(X, d')$
- Define revealed preference relation P as xPy if x always chosen over y for any d
- Using P, can identify choice **set** that maximizes welfare instead of single point
- With continuous choices, effectively obtain bounds on welfare

Bernheim and Rangel 2009: Compensating Variation

- Consider a change in choice set from X to $X' \subset X$
  - Compute CV as amount needed to make agent indifferent to restriction of choice set for each d (standard calculation)
  - Lower bound on CV is minimum over all d's
  - Upper bound on CV is maximum over all d's

Bernheim and Rangel 2009: Compensating Variation

- Ex: suppose insurance plans are restricted to drop M option
- Under red paper condition, CV is 0 – no loss in welfare
- Under blue paper condition, calculate price cut $\$z$ on H needed to make agent indifferent between M and H.
- Bounds on CV: $(0, z)$
- If L option is dropped, bounds collapse to a singleton: $CV = 0$.

Bernheim and Rangel 2009: Refinements

- Problem: looseness of bounds
- Bounds tight when ancillary conditions do not lead to vast changes in choices
- That is, bounds tight when behavioral problems are small
- In cases where behavioral issues are important, this is not going to be a very informative approach

Bernheim and Rangel 2009: Refinements

- Solution: "refinements" – discard certain d's as being "contaminated" for welfare analysis
  - E.g. a neuroscience experiment shows that decisions made under red paper condition are more rational
  - Or assume that choice rational when incentives are more salient
- With fewer d's, get tighter bounds on welfare and policy
- "Refinements" require some positive theory of behavior
- Bernheim and Rangel approach provides a useful framework to organize problems but not sharp policy lessons

Applied Welfare Analysis with Salience Effects

- Chetty, Looney, and Kroft (2009) section 5
- Derive partial-equilibrium formulas for incidence and efficiency costs
- Focus here on efficiency cost analysis
- Formulas do not rely on a specific positive theory, in the spirit of Bernheim and Rangel (2009)

Welfare Analysis with Salience Effects: Setup

- Two goods, x and y; price of y is 1, pretax price of x is p.
- Taxes: y untaxed. Unit sales tax on x at rate t^S, which is not included in the posted price
- Tax-inclusive price of x: $q = p + t^S$

Welfare Analysis with Salience Effects: Setup

- Representative consumer has wealth Z and utility $u(x) + v(y)$
- Let $\{x^*(p, t^S, Z), y^*(p, t^S, Z)\}$ denote bundle chosen by a fully-optimizing agent
- Let $\{x(p, t^S, Z), y(p, t^S, Z)\}$ denote empirically observed demands
- Place no structure on these demand functions except for feasibility:

$$(p + t^S)x(p, t^S, Z) + y(p, t^S, Z) = Z$$

Welfare Analysis with Salience Effects: Setup

- Price-taking firms use y to produce x with cost fn. c
- Firms optimize perfectly. Supply function $S(p)$ defined by:

$$p = c'(S(p))$$

- Let $\varepsilon_S = \frac{\partial S}{\partial p} \times \frac{p}{S(p)}$denote the price elasticity of supply

Efficiency Cost with Salience Effects

- Define excess burden using EV concept
- Excess burden (EB) of introducing a revenue-generating sales tax t is:

$$EB(t^S) = Z - e(p, 0, V(p, t^S, Z)) - R(p, t^S, Z)$$

Preference Recovery Assumptions

A1 Taxes affect utility only through their effects on the chosen consumption bundle. Agent's indirect utility given taxes of (t^E, t^S) is

$$V(p, t^S, Z) = u(x(p, t^S, Z)) + v(y(p, t^S, Z))$$

A2 When tax inclusive prices are fully salient, the agent chooses the same allocation as a fully-optimizing agent:

$$x(p, 0, Z) = x^*(p, 0, Z) = \arg\max_x u(x) + v(Z - px)$$

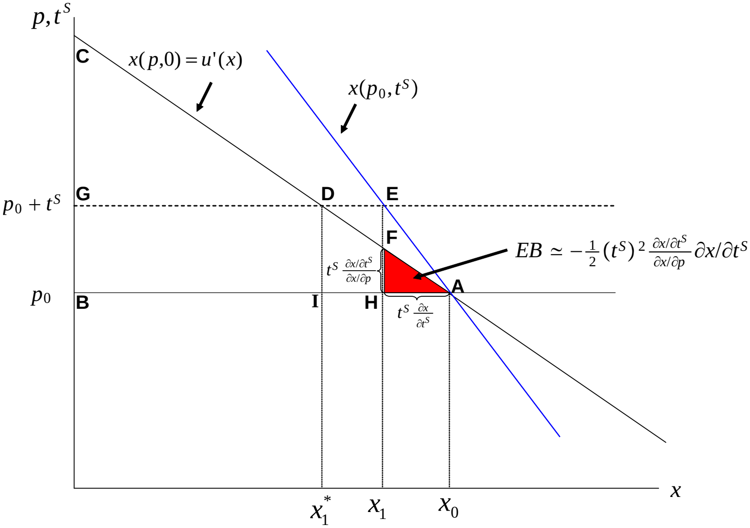
- A1 analogous to specification of ancillary condition; A2 analogous to refinement

Efficiency Cost with Salience Effects

- Two steps in efficiency calculation:

1. Use price-demand $x(p, 0, Z)$ to recover utility as in standard model
2. Use tax-demand $x(p, t^S, Z)$to calculate $V(p, t^S, Z)$ and EB

Excess Burden with No Income Effect for Good x ($\frac{\partial x}{\partial Z} = 0$)



Source: Chetty, Looney, and Kroft (2009)

Efficiency Cost: No Income Effects

- In the case without income effects ($\frac{\partial x}{\partial Z} = 0$), which implies utility is quasilinear, excess burden of introducing a small tax t^S is

$$\begin{aligned} EB(t^S) &\simeq -\frac{1}{2}(t^S)^2 \frac{\partial x / \partial t^S}{\partial x / \partial p} \partial x / \partial t^S \\ &= \frac{1}{2}(\theta t^S)^2 \frac{\varepsilon_D}{p + t^S} \end{aligned}$$

- Inattention reduces excess burden when $dx/dZ = 0$.
- Intuition: tax t^S induces behavioral response equivalent to a fully perceived tax of θt^S.
- If $\theta = 0$, tax is equivalent to a lump sum tax and $EB = 0$ because agent continues to choose first-best allocation.

Efficiency Cost with Income Effects

- Same formula, but all elasticities are now compensated:

$$
\begin{aligned}
EB(t^S) &\simeq -\frac{1}{2}(t^S)^2 \frac{\partial x^c/\partial t^S}{\partial x^c/\partial p} \partial x^c/\partial t^S \\
&= \frac{1}{2}(\theta^c t^S)^2 \frac{\varepsilon_D^c}{p+t^S}
\end{aligned}
$$

- Compensated price demand: $dx^c/dp = dx/dp + xdx/dZ$
- Compensated tax demand: $dx^c/dt^S = dx/dt^S + xdx/dZ$
- Compensated tax demand does not necessarily satisfy Slutsky condition $dx^c/dt^S < 0$ b/c it is not generated by utility maximization

Efficiency Cost with Income Effects

$$\begin{aligned} EB(t^S) &\simeq -\frac{1}{2}(t^S)^2 \frac{\partial x^c/\partial t^S}{\partial x^c/\partial p} \partial x^c/\partial t^S \\ &= \frac{1}{2}(\theta^c t^S)^2 \frac{\varepsilon_D^c}{p+t^S} \end{aligned}$$

- With income effects ($dx/dZ > 0$), making a tax less salient can **raise** deadweight loss.
  - Tax can generate $EB > 0$ even if $dx/dt^S = 0$
- Example: consumption of food and cars; agent who ignores tax on cars underconsumes food and has lower welfare.
- Intuition: agent does not adjust consumption of x despite change in net-of-tax income, leading to a positive compensated elasticity.

Directions for Further Work on Behavioral Welfare Analysis

1. Normative analysis of tax policy
    - Consumption taxation: VAT vs. sales tax
    - Tax smoothing
    - Value of tax simplification
2. Use similar approach to welfare analysis in other contexts
    - Design consumer protection laws and financial regulation in a less paternalistic manner by studying behavior in domains where incentives are clear.

Public Economics Lectures
Part 4: Optimal Taxation

Raj Chetty and Gregory A. Bruich

Harvard University
Fall 2009

Outline

1. Commodity Taxation I: Ramsey Rule
2. Commodity Taxation II: Production Efficiency
3. Income Taxation I: Mirrlees Model
4. Income Taxation II: Atkinson-Stiglitz
5. Capital Income Taxation: Chamley-Judd result
6. Optimal Transfer Programs

Optimal Commodity Taxation: Introduction

- Now combine lessons on incidence and efficiency costs to analyze optimal design of commodity taxes
- What is the best way to design taxes given equity and efficiency concerns?
- Optimal commodity tax literature focuses on linear $(t \cdot x)$ tax system
- Non-linear $(t(x))$ tax systems considered in income tax literature

Second Welfare Theorem

- Starting point: second-welfare theorem
- Can achieve any Pareto-efficient allocation as a competitive equilibrium with appropriate lump-sum transfers
- Requires same assumptions as first welfare theorem plus one more:
  1. Complete markets (no externalities)
  2. Perfect information
  3. Perfect competition
  4. Lump-sum taxes/transfers across individuals feasible
- If 1-4 hold, equity-efficiency trade-off disappears and optimal tax problem is trivial
  - Simply implement lump sum taxes that meet distributional goals given revenue requirement
- Problem: information

Second Welfare Theorem: Information Constraints

- To set the optimal lump-sum taxes, need to know the characteristics (ability) of each individual
- But no way to make people reveal their ability at no cost
  - Incentive to misrepresent skill level
- Tax instruments are therefore a fn. of economic outcomes
  - E.g. income, property, consumption of goods

$\rightarrow$ Distorts prices, affecting behavior and generating DWB

- Information constraints force us to move from the 1st best world of the second welfare theorem to the 2nd best world with inefficient taxation
  - Cannot redistribute or raise revenue for public goods without generating efficiency costs

Four Central Results in Optimal Tax Theory

1. Ramsey (1927): inverse elasticity rule
2. Diamond and Mirrlees (1971): production efficiency
3. Atkinson and Stiglitz (1976): no consumption taxation with optimal *non-linear* (including lump sum) income taxation
4. Chamley, Judd (1983): no capital taxation in infinite horizon models

Ramsey (1927) Tax Problem

- Government sets taxes on uses of income in order to accomplish two objectives:

  1. Raise total revenue of amount E
  2. Minimize utility loss for agents in economy

- Originally a problem set that Pigou assigned Ramsey

Ramsey Model: Key Assumptions

1. Lump sum taxation prohibited
2. Cannot tax all commodities (leisure untaxed)
3. Production prices fixed (and normalized to one):

$$\begin{aligned} p_i &= 1 \\ \Rightarrow q_i &= 1 + \tau_i \end{aligned}$$

Ramsey Model: Setup

- One individual (no redistributive concerns) with utility

$$u(x_1, .., x_N, l)$$

subject to budget constraint

$$q_1 x_1 + .. + q_N x_N \leq wl + Z$$

- Z = non wage income, w = wage rate
- Consumption prices are q_i

Ramsey Model: Consumer Behavior

- Lagrangian for individual's maximization problem:

$$\mathcal{L} = u(x_1, .., x_N, l) + \alpha(wl + Z - (q_1 x_1 + .. + q_N x_N))$$

- First order condition:

$$u_{x_i} = \alpha q_i$$

  Where $\alpha = \partial V / \partial Z$ is marginal value of money for the individual

- Yields demand functions $x_i(q, Z)$ and indirect utility function $V(q, Z)$ where $q = (w, q_1, .., q_N)$

Ramsey Model: Government's Problem

- Government solves either the maximization problem

$$\max V(q, Z)$$

subject to the revenue requirement

$$\tau \cdot x = \sum_{i=1}^{N} \tau_i x_i(q, Z) \geq E$$

- Or, equivalently, minimize excess burden of the tax system

$$\min EB(q) = e(q, V(q, Z)) - e(p, V(q, Z)) - E$$

subject to the same revenue requirement

Ramsey Model: Government's Problem

- For maximization problem, Lagrangian for government is:

$$\mathcal{L}_G = V(q, Z) + \lambda[\sum_i \tau_i x_i(q, Z) - E]$$

$$\Rightarrow \frac{\partial \mathcal{L}_G}{\partial q_i} = \underbrace{\frac{\partial V}{\partial q_i}}_{\text{Priv. Welfare Loss to Indiv.}} + \lambda[\underbrace{x_i}_{\text{Mechanical Effect}} + \underbrace{\sum_j \tau_j \partial x_j / \partial q_i}_{\text{Behavioral Response}}] = 0$$

- Using Roy's identity ($\frac{\partial V}{\partial q_i} = -\alpha x_i$):

$$(\lambda - \alpha) x_i + \lambda \sum_j \tau_j \partial x_j / \partial q_i = 0$$

- Note connection to marginal excess burden formula, where $\lambda = 1$ and $\alpha = 1$

Ramsey Optimal Tax Formula

- Optimal tax rates satisfy system of N equations and N unknowns:

$$\sum_j \tau_j \frac{\partial x_j}{\partial q_i} = -\frac{x_i}{\lambda}(\lambda - \alpha)$$

- Same formula can be derived using a perturbation argument, which is more intuitive

Ramsey Formula: Perturbation Argument

- Suppose government increases τ_i by $d\tau_i$
- Effect of tax increase on social welfare is sum of effect on government revenue and private surplus
- Marginal effect on government revenue:

$$dR = x_i d\tau_i + \sum_j \tau_j dx_j$$

- Marginal effect on private surplus:

$$\begin{aligned} dU &= \frac{\partial V}{\partial q_i} d\tau_i \\ &= -\alpha x_i d\tau_i \end{aligned}$$

- Optimum characterized by balancing the two marginal effects:

$$dU + \lambda dR = 0$$

Ramsey Formula: Compensated Elasticity Representation

- Rewrite in terms of Hicksian elasticities to obtain further intuition using Slutsky equation:

$$\partial x_j / \partial q_i = \partial h_j / \partial q_i - x_i \partial x_j / \partial Z$$

- Substitution into formula above yields:

$$(\lambda - \alpha) x_i + \lambda \sum_j \tau_j [\partial h_j / \partial q_i - x_i \partial x_j / \partial Z] = 0$$

$$\Rightarrow \frac{1}{x_i} \sum_j \tau_j \frac{\partial h_i}{\partial q_j} = -\frac{\theta}{\lambda}$$

where $\theta = \lambda - \alpha - \lambda \frac{\partial}{\partial Z} (\sum_j \tau_j x_j)$

Ramsey Formula: Compensated Elasticity Representation

- θ is independent of i and measures the value for the government of introducing a \$1 lump sum tax

$$\theta = \lambda - \alpha - \lambda \partial (\sum_j \tau_j x_j) / \partial Z$$

- Three effects of introducing a \$1 lumpsum tax:
  1. Direct value for the government is λ
  2. Loss in welfare for the individual is α
  3. Behavioral effect $\rightarrow$ loss in tax revenue of $\partial (\sum_j \tau_j x_j) / \partial Z$

- Can demonstrate that $\theta > 0 \Rightarrow \lambda > \alpha$ at the optimum using Slutsky matrix

Intuition for Ramsey Formula: Index of Discouragement

$$\frac{1}{x_i}\sum_j \tau_j \frac{\partial h_i}{\partial q_j} = -\frac{\theta}{\lambda}$$

- Suppose revenue requirement E is small so that all taxes are also small
- Then tax τ_j on good j reduces consumption of good i (holding utility constant) by approximately

$$dh_i = \tau_j \frac{\partial h_i}{\partial q_j}$$

- Numerator of LHS: total reduction in consumption of good i
- Dividing by x_i yields % reduction in consumption of each good i = "index of discouragement" of the tax system on good i
- Ramsey tax formula says that the indexes of discouragements must be equal across goods at the optimum

Special Case 1: Inverse Elasticity Rule

- Introducing elasticities, we can write formula as:

$$\sum_{j=1}^{N} \frac{\tau_j}{1+\tau_j}\varepsilon_{ij}^{c} = \frac{\theta}{\lambda}$$

- Consider special case where $\varepsilon_{\mathbf{ij}} = 0$ if $i \neq j$
  - Slutsky matrix is diagonal
- Obtain classic inverse elasticity rule:

$$\frac{\tau_i}{1+\tau_i} = \frac{\theta}{\lambda}\frac{1}{\varepsilon_{ii}}$$

Special Case 2: Uniform Taxation

- Suppose $\varepsilon_{\mathbf{ij}} = 0$ if $i \neq j$ *and* $\varepsilon_{x_i,w} = \frac{\partial h_i}{w} \frac{w}{h_i}$ constant
- Using following identity, $\sum_j \frac{\partial h_i}{\partial q_j} q_j + \frac{\partial h_i}{\partial w} w = 0$, we obtain

$$\frac{\partial h_i}{\partial q_i} q_i = -\frac{\partial h_i}{\partial w} w$$

- Proof of identity (J good economy, no labor):

$$\begin{aligned} \sum_j \frac{\partial h_i}{\partial q_j} q_j &= \sum_{j \neq i} \frac{\partial h_j}{\partial q_i} q_j + \frac{\partial h_i}{\partial q_i} q_i \\ &= \sum_{j \neq i} \frac{\partial h_j q_j}{\partial q_i} + \frac{\partial h_i q_i}{\partial q_i} - h_i \\ &= \frac{\partial e}{\partial q_i} - h_i = 0 \end{aligned}$$

Special Case 2: Uniform Taxation

- Then immediately obtain

$$\frac{1}{x_i}\tau_i = -\frac{\theta}{\lambda}\frac{1}{\frac{\partial h_i}{\partial q_i}} = \frac{\theta}{\lambda}\frac{1}{\frac{\partial h_i}{\partial w}w}$$

$$\frac{\tau_i}{q_i} = \frac{\theta}{\lambda}\frac{1}{\frac{\partial h_i}{\partial w}\frac{w}{x_i}} = \frac{\theta}{\lambda}\varepsilon_{x_i,w}$$

- With constant $\varepsilon_{x_i,w}$, $\frac{\tau_i}{q_i}$ is constant $\longrightarrow$ uniform taxation
- Corlett and Hague (1953): 3 good model, uniform tax optimal if all goods are equally complementary with labor (and labor is untaxed)
- More generally, lower taxes for goods complementary to labor
- Different intuition than Goulder and Williams (2003) argument for why taxing goods complementary with labor is undesirable
- Here, higher substitutability with labor $\Rightarrow$ higher own price elasticity; no pre-existing tax on labor

Ramsey Formula: Limitations

- Ramsey solution: tax inelastic goods to minimize efficiency costs
- But does not take into account redistributive motives
- Presumably necessities are more inelastic than luxuries
- Therefore, optimal Ramsey tax system is likely to be regressive
- Diamond (1975) extends Ramsey model to take redistributive motives into account

Diamond 1975: Many-Person Model

- H individuals with utilities $u^1, .., u^h, .., u^H$
- Aggregate consumption of good i is

$$X_i(q) = \sum_h x_i^h$$

- Govt. chooses tax rates τ_i and a lump sum transfer $T \geq 0$ to maximize social welfare:

$$\max W(V^1, .., V^H) \text{ s.t. } \sum_{i=1}^{N} \tau_i X_i \geq E + T$$

- Consider effect of increasing tax on good i by $d\tau_i$

Diamond: Effect of Tax Increase

- Effect of perturbation on revenue:

$$dE = X_i d\tau_i + \sum_j \tau_j dX_j = d\tau_i [X_i + \sum_j \tau_j \frac{\partial X_j}{\partial q_i}]$$

- Effect on individual h's welfare:

$$dU^h = \frac{\partial V^h}{\partial q_i} d\tau_i = -\alpha^h x_i^h d\tau_i$$

- Effect on total private welfare:

$$dW = \sum_h -(\partial W / \partial V^h) \alpha^h x_i^h d\tau_i = -d\tau_i [\sum_h \beta^h x_i^h]$$

where $\beta^h = \partial W / \partial V^h \alpha^h$ is h's *social* marginal utility of wealth

- At optimum:

$$dW + \lambda dE = 0$$

Diamond: Many-Person Optimal Tax Formula

- Solving yields formula for optimal tax rates:

$$-\sum_j \tau_j \frac{\partial X_j}{\partial q_i} = \frac{X_i}{\lambda}[\lambda - \frac{\sum_h \beta^h x_i^h}{X_i}]$$

- With no redistributive tastes (Ramsey case): $\beta^h = \alpha$ constant
  - Obtain same formula as before (in terms of uncompensated elasticities)
- With redistributive tastes, β^h lower for higher income individuals
  - New term $\frac{\sum_h \beta^h x_i^h}{X_i}$ is average social marginal utility, weighted by consumption of good i

Diamond Formula: Special Case

- When uncompensated cross price elasticities are zero, optimal tax rates satisfy
$$\frac{\tau_i}{1+\tau_i} = \frac{1}{\epsilon^u_{ii}}\left(1 - \frac{\sum_h \beta^h x^h_i}{\lambda X_i}\right)$$
- τ_i still inversely proportional to the elasticity but term in brackets no longer constant across goods
- For goods that are consumed by the poor $(\sum_h \beta^h x^h_i)/(\lambda X_i)$ is large
- Optimal tax rate for these goods is lower (elasticities being the same)
- Opposite for goods consumed by the rich

Diamond: Optimal Transfer

- In this model, optimal for the government to pay a uniform transfer T on top of tax rates
- With redistributive tastes, $T > 0$
- With no redistributive tastes, ideally set $T = -E$
- This is ruled out by constraint $T \geq 0$
  - Constraint arises because poor cannot afford to pay lump sum tax

Diamond and Mirrlees (1971)

- Previous analysis assumed fixed producer prices
- Diamond and Mirrlees (1971) relax this assumption by modelling production
- Two major results
  1. Production efficiency: even in an economy where first-best is unattainable, optimal policy maintains production efficiency
  2. Characterize optimal tax rates with endogenous prices and show that Ramsey rule can be applied

Lipsey and Lancaster (1956): Theory of the Second Best

- Standard optimal policy results only hold with single deviation from first best
  - Ex: Ramsey formulas invalid if there are pre-existing distortions, imperfect competition, etc.
- In second-best, anything is possible
  - Policy changes that would increase welfare in a model with a single deviation from first best need not do so in second-best
  - Ex: tariffs can improve welfare by reducing distortions in other part of economy
- Destructive result for welfare economics

Diamond and Mirrlees: Production Efficiency

- Diamond and Mirrlees result was an advance because it showed a general policy lesson even in second-best environment
- Example: Suppose government can tax consumption goods and also produces some goods on its own (e.g. postal services)
- May have intuition that government should try to generate profits in postal services by increasing the price of stamps
- This intuition is wrong: optimal to have no distortions in production of goods
- Bottom line: only tax goods that appear directly in agent's utility functions
- Should not distort production decisions via taxes on intermediate goods, tariffs, etc.

Diamond and Mirrlees Model

- Two good (labor, consumption), one consumer model
- Begin with this case because results easily seen graphically
- In one consumer case, restrict attention to situation where cannot impose lump sum tax
- Corresponding case in many consumer case: permit only uniform lump sum taxation

Diamond and Mirrlees Model: Setup

- Government directly chooses allocations and production subject to requirement that allocation must be supported by an equilibrium price vector
- Government levies tax τ on consumption to fund revenue requirement E
- Individual budget constraint: $(1+\tau)c \leq I$
- First trace out demand as a function of tax rates: the offer curve

Consumer's Offer Curve

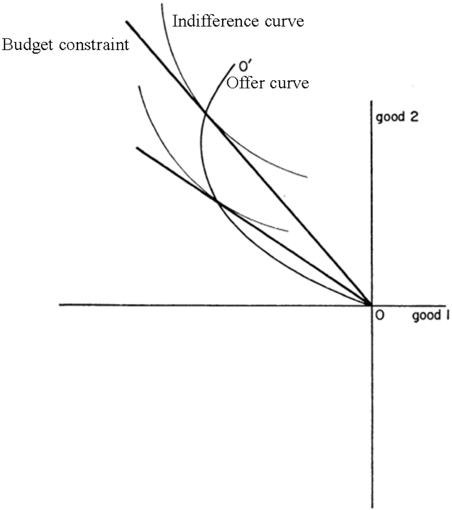


FIGURE 3

Source: Diamond and Mirrlees (1971)

Diamond and Mirrlees: Social Planner's problem

- Government's problem is to

$$\max_{\tau} V(q) = u(x(q), l(q))$$

subject to two constraints

1. Revenue constraint: $\tau c \geq E$
2. Production constraint: $x = f(l)$

- Replace these constraints by $(l, c) \in H$ where H is feasible production set taking into account the tax revenue needed

Production Set with Revenue Requirement

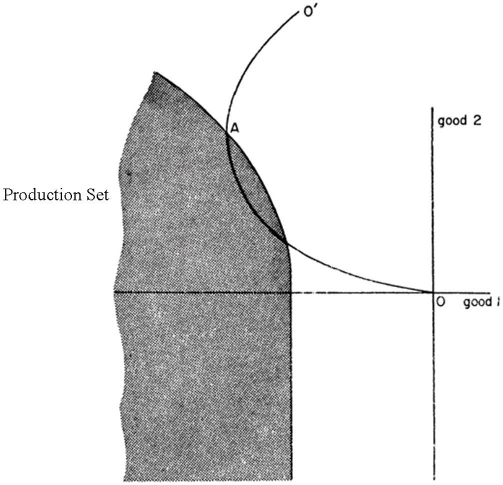


FIGURE 4

Source: Diamond and Mirrlees (1971)

First Best: Optimal Lump Sum Tax

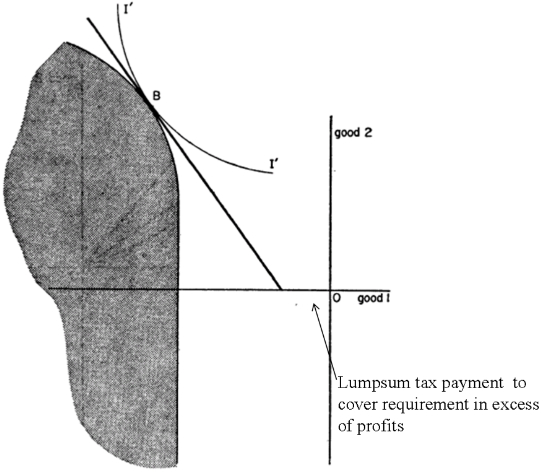


FIGURE 6

Source: Diamond and Mirrlees (1971)

Second Best: Optimal Distortionary Tax

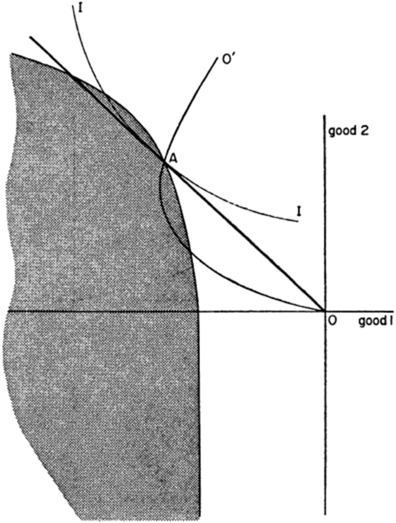


FIGURE 5

Source: Diamond and Mirrlees (1971)

Production Efficiency Result in One-Consumer Model

- Key insight: allocation with optimal distortionary tax is still on PPF
- Equilibrium price vector q places consumer on PPF, subject to revenue requirement
- With lump sum tax, tangency between PPF and consumer's indifference curve, yielding higher welfare

Diamond and Mirrlees: General Model

- Many consumers, many goods and inputs
- Important assumption: either constant returns to scale in production (no profits) or pure profits can be fully taxed
- With this assumption, profits do not enter social welfare fn

Diamond and Mirrlees: General Model

- Government chooses the vector $q = p + \tau$ to

$$\max W(V^1(q), .., V^H(q)) \text{ s.t. } \sum_i \tau_i \cdot X_i(q) \geq E$$

where $X_i(q) = \sum_h x_i^h(q)$, sum of individual demands given after-tax prices q

- Constraint can be replaced by

$$X(q) = \sum_h x^h(q) \in H$$

where H is the production set which takes into account the government requirement E of the government

- **Efficiency result:** at the optimum q^*, $X(q^*)$ is on the boundary of H

Proof of Production Efficiency Result

- Suppose $X(q^*)$ is in the interior of H
- Then take a commodity i that is desired by everybody, and decrease tax on i by $d\tau_i$
- Then $X(q^* - d\tau_i) \in H$ for $d\tau_i$ small by continuity of demand functions; so it is a feasible point
- Everybody is better because of that change:

$$dV^h = -V^h_{q_i} d\tau_i = V^h_R x^h_i d\tau_i$$

- This implies that q^* is not the optimum. Q.E.D.

Production Efficiency Result

- Result can be stated algebraically using MRS and MRT
- Consider two industries, x and y and two inputs, K and L
- Then with the optimal tax schedule, production is efficient:

$$MRTS_{KL}^{x} = MRTS_{KL}^{y}$$

- This is true even though allocation is inefficient:

$$MRT_{xy} \neq MRS_{xy}$$

Policy Consequences: Public Sector Production

- Public sector production should be efficient
- If there is a public sector producing some goods, it should:
  - Face the same prices as the private sector
  - Choose production with the unique goal of maximizing profits, not generating government revenue
- Ex. postal services, electricity, health care, ...

Policy Consequences: No Taxation of Intermediate Goods

- **Intermediate goods:** goods that are neither direct inputs or outputs to indiv. consumption
- Taxes on transactions between firms would distort production

Policy Consequences: No Taxation of Intermediate Goods

- Consider two industries, with labor as the primary input
- Intermediate good A, final good B
- Industry A:
  - Uses labor l_A to produce good A
  - One for one technology
- Industry B:
  - Uses good A and labor l_B to produce good B $x_B = F(l_B, x_A)$
  - Constant returns to scale
- With wage rate w, the producer price of good A is $p_A = w$
- Suppose that good A is taxed at rate τ
- Then the cost for firm B of good A is $w + \tau$

Policy Consequences: No Taxation of Intermediate Goods

- Firm B chooses l and x_A to max

$$F(l_B, x_A) - wl - (w + \tau)x_A$$
$$\Rightarrow \quad F_l = w \text{ and } F_{x_A} = w + \tau > F_l$$

- Aggregate production is inefficient:
  - Decrease l_B and increase l_A a small amount
  - Then x_A increases
  - Total production of good B increases
  - And tax revenue rises (government budget constraint satisfied)

Policy Consequences: No Taxation of Intermediate Goods

- Computers:
  - Sales to firms should be untaxed
  - But sales to consumers should be taxed
- In practice, tax policy often follows precisely the opposite rule
- Ex. Diesel fuel tax studied by Marion and Muehlegger (2008)

Policy Consequences: Tariffs

- In open economy, the production set is extended because it is possible to trade at linear prices (for a small country) with other countries
- Diamond-Mirrlees result: small open economy should be on the frontier of the extended production set
- Implies that no tariffs should be imposed on goods and inputs imported or exported by the production sector
- Ex. sales of IBM computers to other countries should be untaxed
- Ex. purchases of oil by oil companies should be untaxed
- Ex. should be no special tariff on imported cars from Japan, but should bear same commodity tax as cars made in US

Diamond and Mirrlees: Optimal Tax Rates

- Optimal tax formulas take the same form as the solution to Ramsey many-persons problem

- Result holds even where producer prices are not constant

- Same formulas as in Ramsey just by replacing the p's by the actual p's that arise in equilibrium

- Key point: Incidence in the production sector and GE responses can be completely ignored in formulas

Diamond and Mirrlees Model: Key Assumptions

- Result hinges on key assumptions about govt's ability to:
  1. Set a full set of differentiated tax rates on each input and output
  2. Tax away fully pure profits (or production is constant-returns-to-scale)
- A2 rules out improving welfare by taxing profitable industries to improve distribution at expense of prod. eff.
- These assumptions effectively separate the production and consumption problems

Diamond and Mirrlees Result: Limitations

- **Practical relevance** of the result is a bit less clear
- Ex. Assumption 1 is not realistic (Naito 1999)
- Skilled and unskilled labor inputs ought to be differentiated
- Not the case in current income tax system
- In such cases, may be optimal to:
    1. Subsidize low skilled intensive industries
    2. Set tariffs on low skilled intensive imported goods (to protect domestic industry)

Optimal Income Taxation: Outline

1. Optimal Static Income Taxation: Mirrlees (1971)
2. Empirical Implementation of Mirrlees Model: Saez (2001)
3. Income and Commodity Taxation: Atkinson and Stiglitz (1976)
4. Optimal Transfer Programs: Saez (2002)

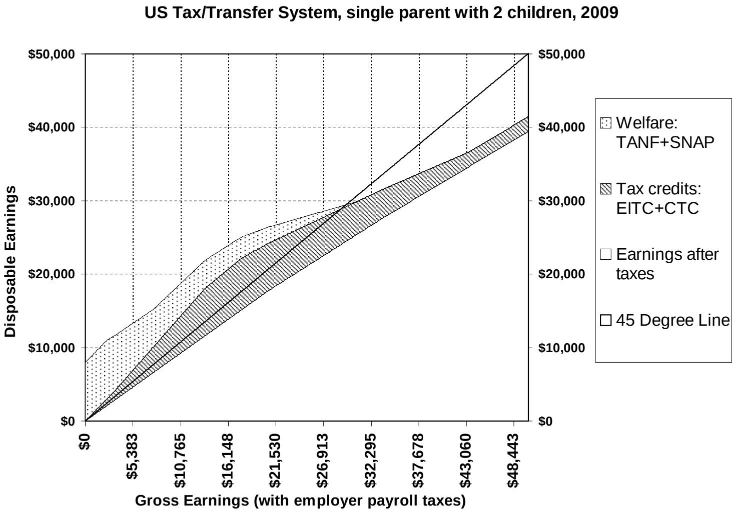
Key Concepts for Taxes/Transfers

- Let $T(z)$ denote tax liability as a function of earnings z

1. Transfer benefit with zero earnings $-T(0)$ [sometimes called demogrant or lumpsum grant]
2. Marginal tax rate $T'(z)$: individual keeps $1 - T'(z)$ for an additional $1 of earnings (relevant for intensive margin labor supply responses)
3. Participation tax rate $\tau_p = [T(z) - T(0)]/z$: individual keeps fraction $1 - \tau_p$ of earnings when moving from zero earnings to earnings z:

$$z - T(z) = -T(0) + z - [T(z) - T(0)] = -T(0) + z \cdot (1 - \tau_p)$$

   Relevant for extensive margin labor supply responses
4. Break-even earnings point z^*: point at which $T(z^*) = 0$



Source: Saez 2010 AEA Clark Lecture

Optimal Income Tax with No Behavioral Responses

- Utility $u(c)$ strictly increasing and concave
- Same for everybody where c is after tax income
- Income is z and is fixed for each individual, $c = z - T(z)$ where $T(z)$ is tax on z
- Government maximizes Utilitarian objective:

$$\int_0^\infty u(z - T(z))h(z)dz$$

- Subject to budget constraint $\int T(z)h(z)dz \geq E$ (multiplier λ)

Optimal Income Tax without Behavioral Responses

- Lagrangian for this problem is:

$$\mathcal{L} = [u(z - T(z)) + \lambda T(z)]h(z)$$

- First order condition:

$$\begin{aligned} T(z) \quad : \quad & 0 = \partial L/\partial T(z) = [-u'(z - T(z)) + \lambda]h(z) \\ \Rightarrow \quad & u'(z - T(z)) = \lambda \\ \Rightarrow \quad & z - T(z) = c \text{ constant for all } z \\ \Rightarrow \quad & c = \bar{z} - E \end{aligned}$$

where $\bar{z} = \int zh(z)dz$ average income

- 100% marginal tax rate; perfect equalization of after-tax income

- Utilitarianism with diminishing marginal utility leads to egalitarianism

Mirrlees 1971: Incorporating Behavioral Responses

- Standard labor supply model: Individual maximizes

$$u(c, l) \text{ s.t. } c = wl - T(wl)$$

where c is consumption, l labor supply, w wage rate, $T(.)$ income tax

- Individuals differ in ability w distributed with density $f(w)$

- Govt social welfare maximization: Govt maximizes

$$\begin{aligned} SWF &= \int G(u(c,l))f(w)dw) \\ \text{s.t. resource constraint} & \quad \int T(wl)f(w)dw \geq E \\ \text{and individual FOC} & \quad w(1-T')u_c + u_l = 0 \end{aligned}$$

where $G(.)$ is increasing and concave

Mirrlees 1971: Results

- Optimal income tax trades-off redistribution and efficiency
  - $T(.) < 0$ at bottom (transfer)
  - $T(.) > 0$ further up (tax) [full integration of taxes/transfers]
- Mirrlees formulas are a complex fn. of primitives, with only a few general results
  1. $0 \leq T'(.) \leq 1$, $T'(.) \geq 0$ is non-trivial and rules out EITC [Seade 1976]
  2. Marginal tax rate $T'(.)$ should be zero at the top if skill distribution bounded [Sadka-Seade]

Mirrlees: Subsequent Work

- Mirrlees model had a profound impact on information economics
  - Ex. models with asymmetric information in contract theory
- But until late 1990s, Mhad little impact on practical tax policy
- Recently, Mirrlees model connected to empirical literature
  - Diamond (1998), Piketty (1997), and Saez (2001)
  - Sufficient statistic formulas in terms of labor supply elasticities instead of primitives

Optimal Income Taxation: Sufficient Statistic Formulas

1. Revenue-maximizing linear tax (Laffer curve)
2. Top income tax rate (Saez 2001)
3. Full income tax schedule (Saez 2001)

- See also section 4 of Chetty (Ann. Rev. 2009)

Revenue-Maximizing Tax Rate: Laffer Curve

- With a constant tax rate τ, reported income z depends on $1-\tau$ (net-of-tax rate)
- Tax Revenue $R(\tau) = \tau \cdot z(1-\tau)$ is inverse-U shaped:
  - $R(\tau = 0) = 0$ (no taxes) and $R(\tau = 1) = 0$ (nobody works)
- Tax rate τ^* that maximizes R:

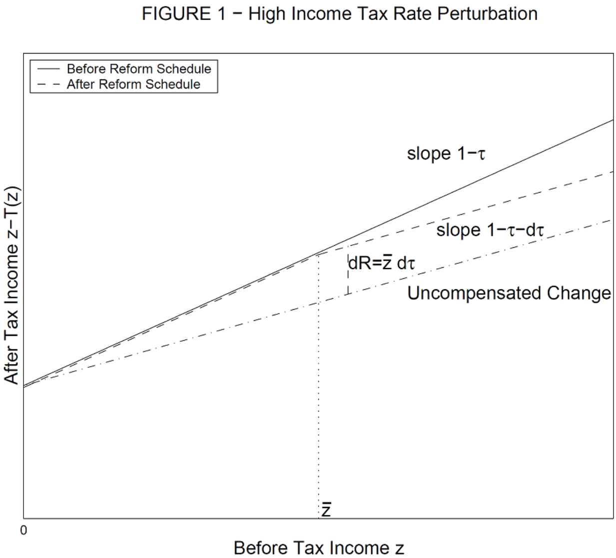
$$\begin{aligned} 0 &= R'(\tau^*) = z - \tau^* dz/d(1-\tau) \\ &\Rightarrow \tau^*_{\text{MAX}} = 1/(1+\varepsilon) \end{aligned}$$

  where $\varepsilon = [(1-\tau)/Z]dz/d(1-\tau)$ is the taxable income elasticity

- Strictly inefficient to have $\tau > \tau^*$

Optimal Top Income Tax Rate

- Now consider constant mtr τ above fixed income threshold $\bar{z}$
- Derive optimal τ using perturbation argument
- Assume away income effects $\varepsilon^c = \varepsilon^u = \varepsilon$
    - Diamond (1996) shows this is a key theoretical simplification
- Assume that there are N individuals above $\bar{z}$
- Denote by $z^m(1-\tau)$ their average income, which depends on net-of-tax rate $1-\tau$



Source: Saez 2001

Optimal Top Income Tax Rate

- Three effects of small $d\tau > 0$ reform above $\bar{z}$
- **Mechanical increase** in tax revenue:

$$dM = N \cdot [z^m - \bar{z}] d\tau$$

- **Behavioral response:**

$$\begin{aligned} dB &= N\tau dz^m = -N\tau \frac{dz^m}{d(1-\tau)} d\tau \\ &= -N\frac{\tau}{1-\tau} \cdot \bar{\varepsilon} \cdot z^m d\tau \end{aligned}$$

- **Welfare effect:** money-metric utility loss is dM by envelope theorem:
    - If govt. values marginal consumption of rich at $\bar{g} \in (0,1)$

    $$dW = -\bar{g} dM$$

    - $\bar{g}$ depends on curvature of $u(c)$ and SWF

Optimal Top Income Tax Rate

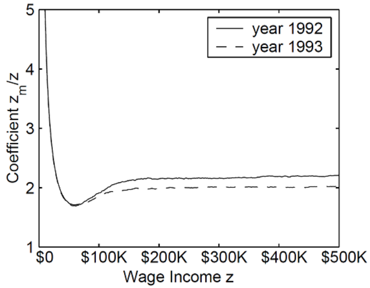
$$dM + dW + dB = Nd\tau \left\{ (1 - \bar{g})[z^m - \bar{z}] - \bar{\varepsilon} \frac{\tau}{1-\tau} z^m \right\}$$

- Optimal τ such that $dM + dW + dB = 0 \Rightarrow$

$$\frac{\tau^*_{\text{TOP}}}{1 - \tau^*_{\text{TOP}}} = \frac{(1 - \bar{g})(z_m/\bar{z} - 1)}{\bar{\varepsilon} \cdot z_m/\bar{z}}$$

- τ^*_{TOP} decreases with $\bar{g}$ [redistributive tastes]
- τ^*_{TOP} decreases with $\bar{\varepsilon}$ [efficiency]
- τ^*_{TOP} increases with $z_m/\bar{z}$ [thickness of top tail]
- Note: this is not an explicit formula for top tax rate because $z_m/\bar{z}$ is a fn. of τ

FIGURE 2 − Ratio mean income above z divided by z, z_m/z, years 1992 and 1993



Source: Saez 2001

Optimal Top Income Tax Rate

- In US tax return data, $z^m/\bar{z}$ very stable above $\bar{z} = \$200K$ with $\frac{z_m}{\bar{z}} = 2$

- With Pareto distribution $(f(z) = a \cdot k^a / z^{1+a})$, $\frac{a}{a-1} = \frac{z_m}{\bar{z}} \Rightarrow a = 2$

$$\Rightarrow \tau^*_{\text{TOP}} = \frac{1 - \bar{g}}{1 - \bar{g} + a \cdot \bar{\varepsilon}}$$

- Ex: $\bar{\varepsilon} = 0.5$, $\bar{g} = 0.5$, $a = 2 \Rightarrow \tau^*_{\text{TOP}} = 33\%$

Zero Top Rate with Bounded Distribution

- Suppose top earner earns z^T, and second earner earns z^S
- Then $z^m = z^T$ when $\bar{z} > z^S \Rightarrow z^m/\bar{z} \to 1$ when $\bar{z} \to z^T \Rightarrow$

$$dM = Nd\tau[z^m - \bar{z}] \to 0 < dB = Nd\tau\bar{\varepsilon}\frac{\tau}{1-\tau}z^m$$

- Optimal τ is zero for $\bar{z}$ close to z^T
  - Sadka-Seade zero top rate result
- Result applies literally only to top earner: if $z^T = 2 \cdot z^S$ then $z^m/\bar{z} = 2$ when $\bar{z} = z^S$
  - Zero at top no longer considered to be of practical relevance

Connection to Revenue Maximizing Tax Rate

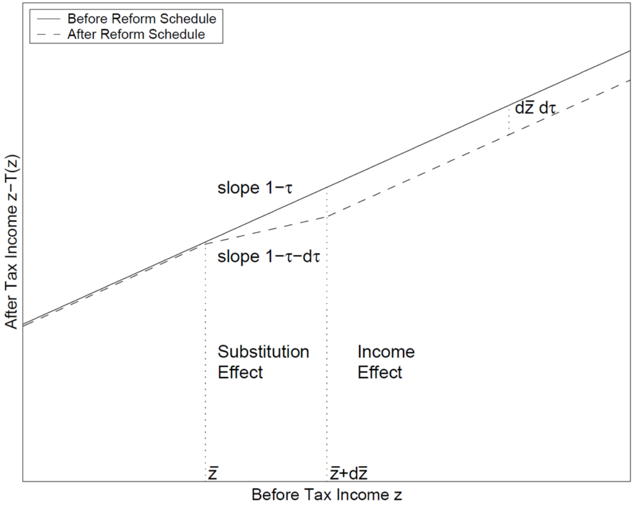
- Revenue maximizing top tax rate can be calculated by putting 0 weight on welfare of top incomes
  - Utilitarian SWF $\Rightarrow \bar{g} = u_c(z^m) \rightarrow 0$ when $\bar{z} \rightarrow \infty$
  - Rawlsian SWF $\Rightarrow \bar{g} = 0$ for any $\bar{z} > \min(z)$
- If $\bar{g} = 0$, we obtain $\tau_{TOP} = \tau_{MAX} = 1/(1 + a \cdot \bar{\varepsilon})$
- Example: $a = 2$ and $\bar{\varepsilon} = 0.5 \Rightarrow \tau = 50\%$
- Laffer linear rate is a special case where $\bar{z} = 0$

$$\Rightarrow z^m/\bar{z} = \infty = a/(a-1) \Rightarrow a = 1 \Rightarrow \tau_{MAX} = 1/(1+\bar{\varepsilon})$$

Optimal Non-Linear Income Tax

- Now consider general problem of setting optimal $T(z)$
- Let $H(z) =$ CDF of income [population normalized to 1] and $h(z)$ its density [endogenous to $T(.)$]
- Let $g(z) =$ social marginal value of consumption for taxpayers with income z in terms of public funds
- Let $G(z)$ be the **average** social marginal value of consumption for taxpayers with income above z $[G(z) = \int_z^\infty g(s)h(s)ds/(1 - H(z))]$

FIGURE 3 − Local Marginal Tax Rate Perturbation



Source: Saez 2001

General Non-Linear Income Tax

- Consider small reform: increase T' by $d\tau$ in small band $(z, z + dz)$

- Mechanical revenue effect

$$dM = dzd\tau(1 - H(z))$$

- Mechanical welfare effect

$$dW = -dzd\tau(1 - H(z))G(z)$$

- Behavioral effect: substitution effect δz inside small band $[z, z + dz]$:

$$dB = h(z)dz \cdot T' \cdot \delta z = h(z)dz \cdot T' \cdot d\tau \cdot \varepsilon(z) \cdot z/(1 - T')$$

- Optimum $dM + dW + dB = 0$

General Non-Linear Income Tax

- Optimal tax schedule satisfies:

$$\frac{T'(z)}{1-T'(z)} = \frac{1}{\varepsilon(z)}\left(\frac{1-H(z)}{zh(z)}\right)[1-G(z)]$$

- $T'(z)$ decreasing in $g(z')$ for $z' > z$ [redistributive tastes]
- $T'(z)$ decreasing in $\varepsilon_{(z)}$ [efficiency]
- $T'(z)$ decreasing in $h(z)/(1-H(z))$ [density]
- Connection to top tax rate: consider $z \to \infty$
  - $G(z) \to \bar{g}$, $(1-H(z))/(zh(z)) \to 1/a$
  - $\varepsilon_{(z)} \to \bar{\varepsilon} \Rightarrow T'(z) = (1-\bar{g})/(1-\bar{g}+a\cdot\bar{\varepsilon}) = \tau_{\text{TOP}}$

Negative Marginal Tax Rates Never Optimal

- Suppose $T' < 0$ in band $[z, z + dz]$
- Increase T' by $d\tau > 0$ in band $[z, z + dz]$
  - $dM + dW > 0$ because $G(z) < 1$ for any $z > 0$ (with declining $g(z)$ and $G(0) = 1$)
  - $dB > 0$ because $T'(z) < 0$ [smaller efficiency cost]
- Therefore $T'(z) < 0$ cannot be optimal
  - Marginal subsidies also distort local incentives to work
  - Better to redistribute using lump sum grant

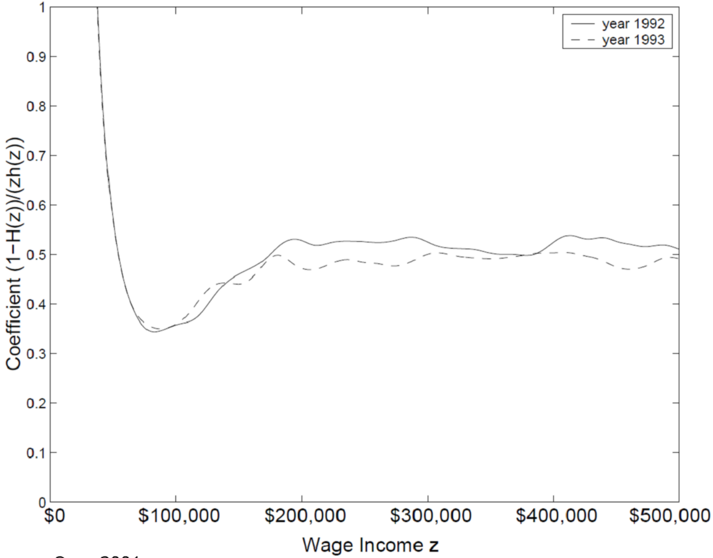
Numerical Simulations of Optimal Tax Schedule

- Formula above is a condition for optimality but not an explicit formula for optimal tax schedule
  - Distribution of incomes $H(z)$ endogenous to $T(.)$
- Therefore need to use structural approach (specification of primitives) to calculate optimal $T(.)$
- Saez (2001) specifies utility function (e.g. constant elasticity):

$$\begin{aligned} u(c,l) &= c-(l)^{1+\frac{1}{\varepsilon}} \\ &\Rightarrow l^{*}=[(1-T')w]^{\varepsilon} \end{aligned}$$

- Calibrate the exogenous skill distribution $F(w)$ such that actual $T(.)$ yields **empirical** $H(z)$

FIGURE 4 – Hazard Ratio (1−H(z))/(zh(z)), years 1992 and 1993



Source: Saez 2001

Numerical Simulations

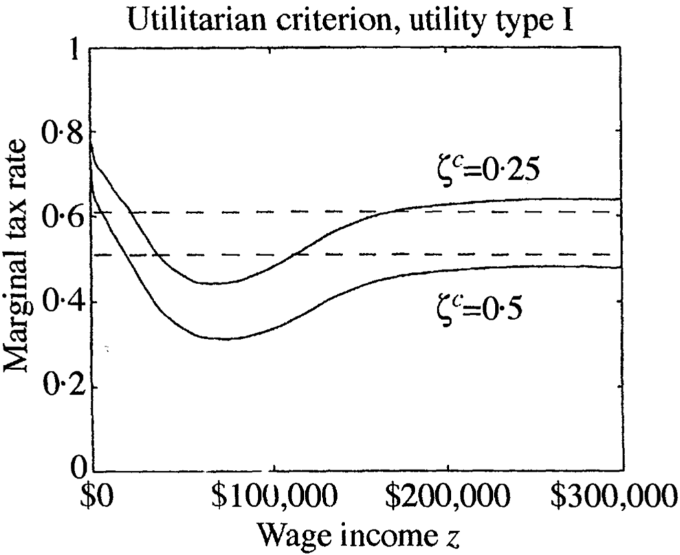
- Use formula expressed in terms of $F(w)$ to solve for optimal $T(z)$:

$$\frac{T'(z(w))}{1-T'(z(w))} = \left(1+\frac{1}{\varepsilon}\right)\left(\frac{1}{wf(w)}\right)\int_w^\infty \left[1-\frac{G'(u(s))}{p}\right] f(s)ds,$$

  where $p = \int G'(u(s))f(s)ds$ is marginal value of public funds

- Iterative fixed point method to solve for $T(z)$:
  - Start with initial MTR schedule T_0' and compute incomes $z^0(w)$ using individual FOCs
  - Get $T^0(0)$ using govt budget constraint, compute utilities $u^0(w)$
  - Compute $p_0 = \int G'(u^0(s))f(s)ds$

- Use formula to calculate T_1' and iterate until convergence (Brewer, Saez, Shephard 2009)



Source: Saez 2001

Commodity vs. Income Taxation

- Now combine commodity tax and income tax results to analyze optimal combination of policies

- In practice, government levies differential commodity taxes along with non-linear income tax

  1. Exempts some goods (food, education, health) from sales tax

  2. Imposes additional excise taxes on some goods (cars, gasoline, luxury goods)

  3. Imposes capital income taxes

- What is the best combination of taxes?

Commodity vs. Income Taxation: Model

- K consumption goods $c = (c_1, .., c_K)$ with pre-tax price $p = (p_1, .., p_K)$
- Individual h has utility $u^h(c_1, .., c_K, z)$
- Can govt increase welfare using commodity taxes $t = (t_1, .., t_K)$ in addition to nonlinear optimal income tax on earnings z?
- We know that more instruments cannot hurt:

$$\max_{t, T(.)} SWF \geq \max_{t=0, T(.)} SWF$$

Atkinson and Stiglitz: Commodity Taxation is Superfluous

- Atkinson and Stiglitz (1976) show that

$$\max_{t,T(.)} SWF = \max_{t=0,T(.)} SWF$$

- Commodity taxes not useful under two assumptions on utility functions $u^h(c_1, .., c_K, z)$

  1. Separability between $(c_1, .., c_K)$ and z in utility
  2. Homogeneity across individuals in the sub-utility of consumption:

$$u^h(c_1, .., c_K, z) = U^h(v(c_1, .., c_K), z)$$

- Original proof was based on optimality conditions
  - More straightforward proof by Laroque (2005) and Kaplow (2006)

Atkinson-Stiglitz: Proof

- Let $V(y, q) = \max_c v(c_1, .., c_K)$ st $qc \leq y$ be the indirect utility of consumption given post-tax earnings y and price q
  - This function is common across all individuals under assumptions above
- Start with any tax system $(T(.), t)$
- Replace $(T(.), t)$ with $(\bar{T}(.), t = 0)$ where $\bar{T}(z)$ is such that

$$V(z - T(z), p + t) = V(z - \bar{T}(z), p)$$

- Utility $U^h(V, z)$ unchanged for all individuals
- Labor supply choices z unchanged as well because return to work $V'(z)$ unchanged

Atkinson-Stiglitz: Proof

- Revenue under original tax system: $T(z) + t \cdot c(t)$
- Revenue under new tax system: $\bar{T}(z)$
- Claim: $\bar{T}(z) \geq T(z) + t \cdot c(t)$
  - Conditional on z, $T(z)$ is a lump sum tax whereas t is distortionary
  - For a given utility level, can extract more using lump sum tax than distortionary tax

Atkinson-Stiglitz: Proof

- Algebraic proof of claim
  - Let $c(t)$ denote optimal bundle with tax $(t, T(z))$ and $c(0)$ denote optimal bundle with tax $(0, \bar{T}(z))$
  - Both bundles yield same utility V by construction
  - Optimization implies

$$\begin{aligned} p \cdot c(t) &= z - T(z) - t \cdot c(t) \geq p \cdot c(0) = z - \bar{T}(z) \\ &\Rightarrow \bar{T}(z) \geq T(z) + t \cdot c(t) \end{aligned}$$

- Government collects more taxes with $(\bar{T}(.), t = 0)$ and utility is unchanged
  - Therefore system without commodity taxes yields higher welfare

Atkinson-Stiglitz: Intuition

- With separability and homogeneity, conditional on earnings z, consumption choices $c = (c_1, .., c_K)$ do not provide any information on ability
- Differentiated commodity taxes $t_1, .., t_K$ create a tax distortion with no benefit
  - Better to do all the redistribution with the individual income tax
- With only linear income taxation (Diamond-Mirrlees 1971, Diamond 1975), diff. commodity taxation can be useful to "non-linearize" the tax system
  - But not if Engel curves for each c_k are linear in y (Deaton 1981)

Failures of A-S Assumptions

- If higher ability consume more of good k than lower ability people, then taxing good k is desirable. Examples:

  1. High ability people have a relatively higher taste for good k (at a given income)
     - Luxury chocolates or museums; violates homogeneous $v(c)$ assumption
  2. Good k is positively related to leisure (consumption of k increases when leisure increases at a given income)
     - Tax on travel, subsidy on computers and work related expenses

- In general Atkinson-Stiglitz assumptions are viewed as a good starting place for most goods

Atkinson-Stiglitz: Implications for Capital Taxation

- Two period model: wage rate w in period 1, retired in period 2
- Let $\delta =$ discount rate, $\psi(.)$ disutility of effort, and utility

$$u^h(c_1, c_2, z) = u(c_1) + \frac{u(c_2)}{1+\delta} - \psi(z/w)$$

- The budget constraint is

$$c_1 + c_2/(1 + r(1 - t_K)) \leq z - T(z)$$

- Tax on savings t_K is equivalent to tax on c_2
- Atkinson-Stiglitz implies that $t_K^* = 0$ in the presence of an optimal income tax
  - Very sharp policy prediction

Atkinson-Stiglitz: Implications for Capital Taxation

- If low ability people have higher δ then capital income tax $t_K > 0$ is desirable (Saez 2004)
  - Violates homogeneous utility assumption
  - Savings are equivalent to luxury chocolates or museums
- Saez (2004) restricts capital tax to be linear and income-independent
- With non-linear, income-dependent taxes, optimal t_K may be lower for high incomes than low incomes (Golosov, Tsyvinski, Weinzierl 2009)
  - No longer a justification for redistribution via capital income taxation

Chamley-Judd: Capital Taxation

- Judd (1985) and Chamley (1986) give a different argument against capital taxation
- Consider a Ramsey model where govt. is limited to linear distortionary taxes
- Result: optimal capital tax converges to zero in long run
- Intuition: DWL rises with square of tax rate
  - With non-zero capital tax, have an infinite price distortion between c_0 and c_t as $t \longrightarrow \infty$
  - Undesirable to have such large distortions on some margins

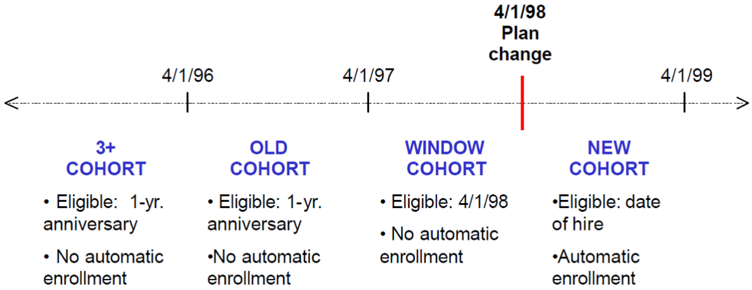
Chamley-Judd vs. Atkinson-Stiglitz

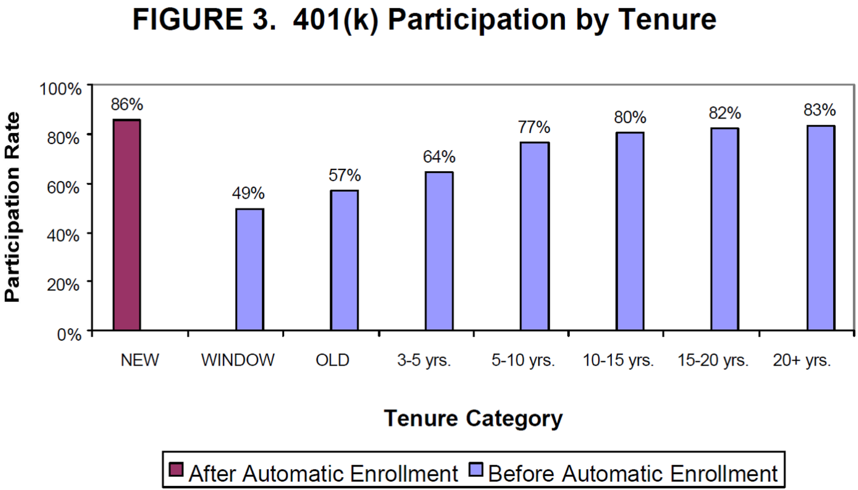
- Chamley-Judd: constrained policy instruments (linear taxes) but dynamic
- Atkinson-Stiglitz: full set of policy instruments (non linear income tax) but static
- New dynamic public finance literature: full set of instruments in dynamic model
- Key result: in dynamic Mirrlees models, optimal capital tax is not zero (Golosov, Kocherlekota, and Tsyvinski 2003)
    - Optimum satisfies Inverse Euler eqn., resulting in a wedge between MRS and MRTS
    - Intuition: payoff to distorting savings decisions relaxes IC constraints in optimal income tax problem in next period
    - Does not emerge in Atkinson-Stiglitz because all income is earned in first period

Taxation and Savings: Evidence

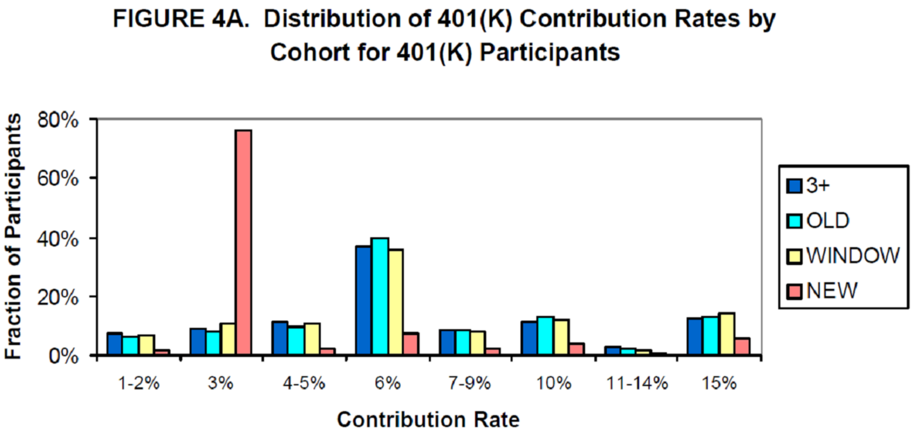
- Key assumption in Chamley-Judd and Atkinson-Stiglitz results: people optimize their savings decisions
- Recent evidence challenges this assumption
- Madrian and Shea (2001) study employee 401(k) enrollment decisions and contribution rates at a U.S. corporation:
    - Most people adhere to company defaults and do not make active savings choices
    - Suggests that defaults may have much bigger impacts on savings decision than net-of-tax returns

Madrian and Shea 2001: Defaults and Savings Behavior





Source: Madrian and Shea 2001



Source: Madrian and Shea 2001

Optimal Transfer Programs

- Several types of transfer programs are used in practice, each justified by a different theory and set of assumptions
- **Option 1**: Negative Income Tax: TANF (Mirrlees 1971)
  - Benefits: no one omitted; low admin costs; no stigma
  - Costs: efficiency loss from less work
- **Option 2**: Work-for-welfare: EITC (Saez 2002)
  - Benefits: more incentive to work; low admin costs
  - Costs: efficiency loss in phaseout range, no coverage of non-workers

Optimal Transfer Programs

- **Option 3**: Categorical anti-poverty programs: assistance for blind (Akerlof 1978)
  - Benefits: tagging relaxes incentive constraint by tying tax rate to immutable qualities
  - Costs: not always feasible and limited coverage
- **Option 4**: In-kind transfers: food stamps, public housing (Nichols and Zeckhauser 1982)
  - Benefits: Efficiency gains from relaxing IC for high-types via ordeals
  - Costs: Paternalism (spend on the right things), inefficient ordeal cost

Optimal Transfers: Mirrless Model

- Mirrlees model predicts that optimal transfer at bottom takes the form of a Negative Income Tax
  - Lump sum grant $-T(0)$ for those with no earnings
  - High MTRs $T'(z)$ at the bottom to phase-out the lumpsum grant quickly
- Intuition: NIT optimal because
  - Targets transfers to the most needy
  - Earnings at the bottom are low to start with so intensive response to high MTRs does not generate large output losses

Optimal Transfers: Participation Responses and EITC

- Mirrlees result predicated on assumption that all individuals are at an interior optimum in choice of labor supply
  - Rules out extensive-margin responses
  - But empirical literature shows that participation labor supply responses are most important especially for low incomes
- Diamond (1980), Saez (2002), Laroque (2005) incorporate such extensive labor supply responses into optimal income tax model
- Generate extensive margin by introducing fixed job packages (cannot smoothly choose earnings)

Saez 2002: Participation Model

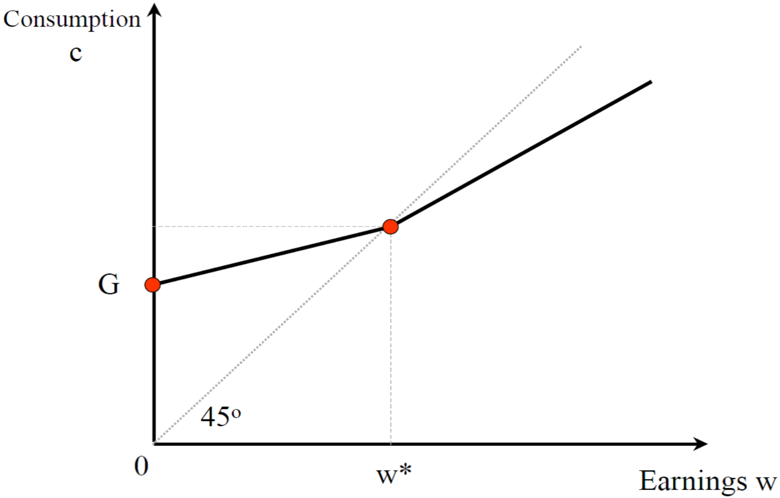
- Model with discrete earnings outcomes: $w_0 = 0 < w_1 < ... < w_I$
- Tax/transfer T_i when earning w_i, $c_i = w_i - T_i$
- Pure participation choice: skill i individual compares c_i and c_0 when deciding to work
- With participation tax rate τ_i, $c_i - c_0 = w_i \cdot (1 - \tau_i)$
- In aggregate, fraction $h_i(c_i - c_0)$ of population earns w_i, so $\sum_i h_i = 1$
- Participation elasticity is

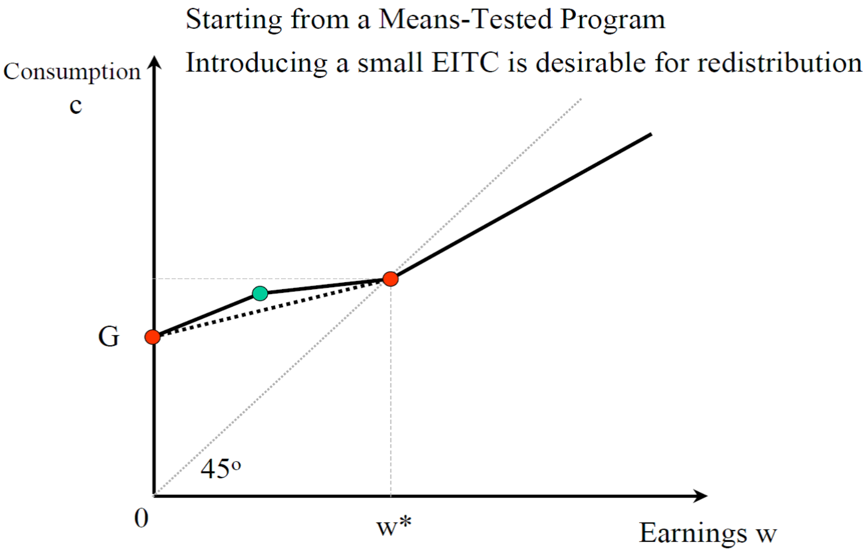
$$e_i = (c_i - c_0)/h_i \cdot \partial h_i / \partial (c_i - c_0)$$

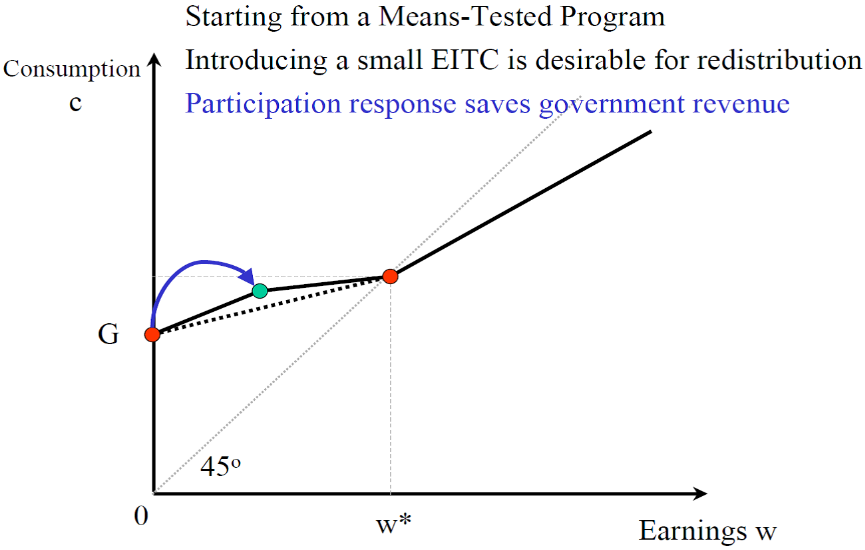
Saez 2002: Participation Model

- Social Welfare function is summarized by social marginal welfare weights at each earnings level g_i
- No income effects $\rightarrow \sum_i g_i h_i = 1$
- Main result: work subsidies with $T'(z) < 0$ (such as EITC) optimal
- Key requirements in general model with intensive+extensive responses
  - Responses are concentrated primarily along extensive margin
  - Social marginal welfare weight on low skilled workers > 1

Starting from a Means-Tested Program




Starting from a Means-Tested Program
Introducing a small EITC is desirable for redistribution
Consumption
c
G
45o
0
w*
Earnings w



Starting from a Means-Tested Program
Introducing a small EITC is desirable for redistribution
Participation response saves government revenue
Consumption
c
G
45o
0
w*
Earnings w

Mirrlees 1971 vs. Saez 2002

- EITC is desirable in Saez extensive-margin model because it
  - Redistributes more money to low incomes
  - Saves the government money by getting people off of welfare
- In Mirrlees intensive-margin model, second effect is shut down
  - Creating an EITC would always cost government more through intensive responses
  - Always preferable to redistribute by giving more money to lowest income

Saez 2002: Optimal Tax Formula

- Small reform $dc_i = -dT_i > 0$. Three effects:

  1. Mechanical loss of tax revenue $dM = h_i dT_i$
  2. Welfare Effect: each worker in job i gains dT_i so welfare gain $dW = -g_i h_i dT_i$
     - No first order welfare loss for switchers
  3. Behavioral Effect: $dh_i = -e_i h_i dT_i / (c_i - c_0)$
     $\rightarrow$Tax loss: $dB = -(T_i - T_0) dh_i = -e_i h_i dT_i (T_i - T_0)/(c_i - c_0)$

- FOC: $dM + dB + dW = 0 \Rightarrow$

$$\frac{\tau_i}{1-\tau_i} = \frac{T_i - T_0}{c_i - c_0} = \frac{1}{e_i}(1 - g_i)$$

- $g_1 > 1 \Rightarrow T_1 - T_0 < 0 \Rightarrow$ work subsidy

Figure 3a: Optimal Tax/Transfer Derivation

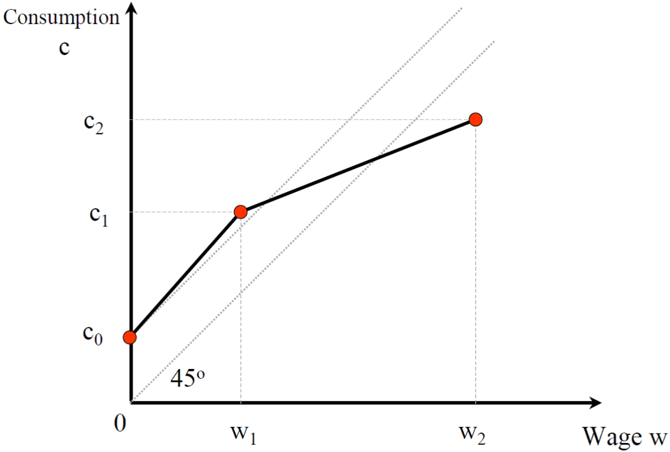


Figure 3a: Optimal Tax/Transfer Derivation (assuming $g_1>1$)

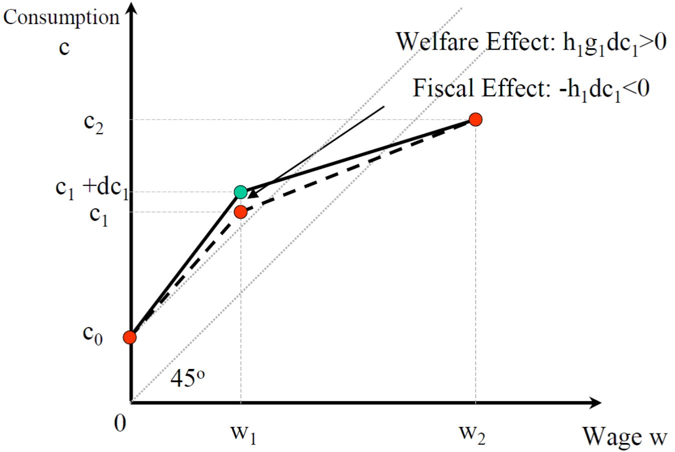


Figure 3a: Optimal Tax/Transfer Derivation (assuming $g_1 > 1$)

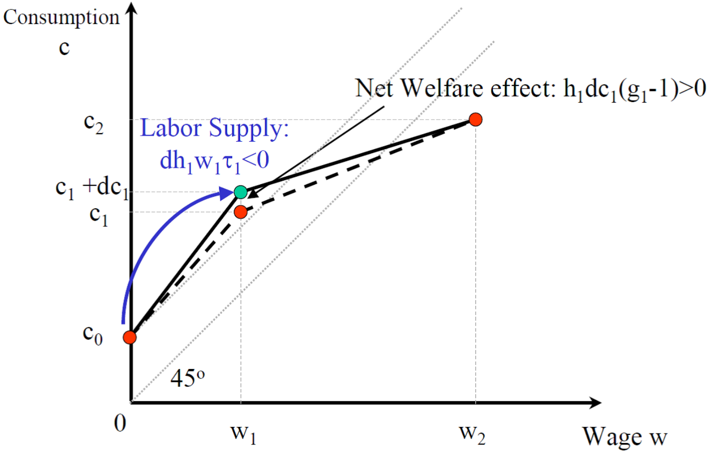
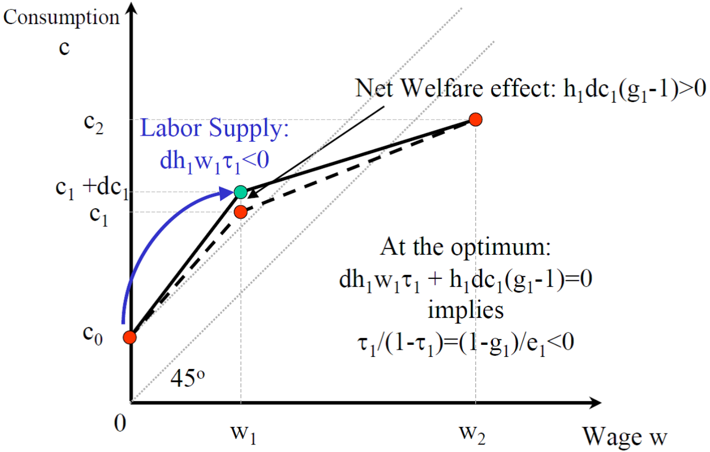
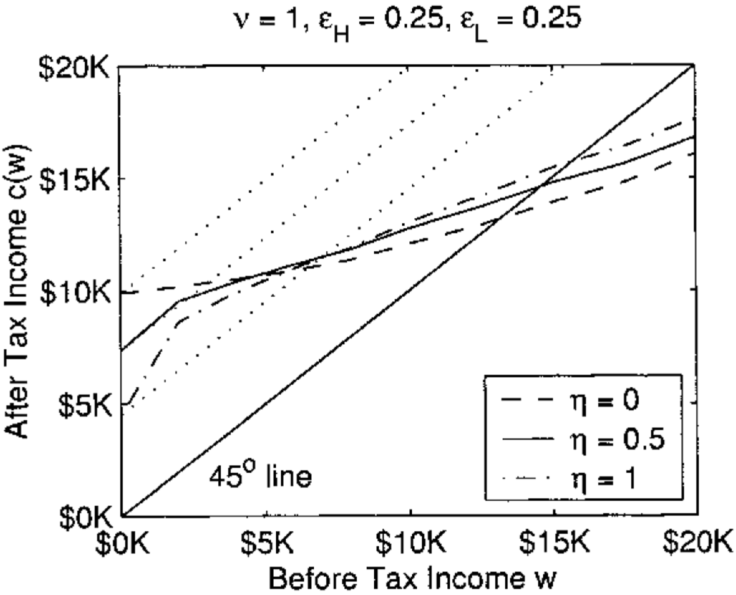


Figure 3a: Optimal Tax/Transfer Derivation (assuming $g_1>1$)



Saez 2002: General Model

- Model can be extended to allow both intensive and extensive responses
  - Allow higher types to switch to lower jobs
- General formula for optimal tax is a fn of both intensive and extensive margin elasticity
- Can be calibrated using empirical estimates of these elasticities



Source: Saez 2002

Tagging: Akerlof 1978

- We have assumed that $T(z)$ depends only on earnings z
- In reality, govt can observe many other characteristics X also correlated with ability and set $T(z, X)$
  - Ex: gender, race, age, disability, family structure, height,...
- Two major results:
  1. If characteristic X is **immutable** then redistribution across the X groups will be complete [until average social marginal welfare weights are equated across X groups]
  2. If characteristic X can be manipulated but X correlated with ability then taxes will depend on both X and z

Mankiw and Weinzierl 2009

- Tagging with Immutable Characteristics
- Consider a binary immutable tag: **T**all vs. **S**hort
- 1 inch = 2% higher earnings on average (Postlewaite et al. 2004)
- Average social marginal welfare weights $\bar{g}^T < \bar{g}^S$ because tall earn more
- Lump sum transfer from Tall to Short is desirable
- Optimal transfer should be up to the point where $\bar{g}^T = \bar{g}^S$
- Calibrations show that average tall person ($>$ 6ft) should pay \$4500 more in tax

Problems with Tagging

- Height taxes seem implausible, challenging validity of tagging model
- What is the model missing?
  1. Horizontal Equity concerns impose constraints on feasible policies:
     - Two people earning same amount but of different height should be treated the same way
  2. Height does not *cause* high earnings
     - In practice, tags used only when **causally** related to ability to earn [disability status] or welfare [family structure, # kids, medical expenses]
- Lesson: Mirrlees analysis $[T(z)]$ may be most sensible even in an environment with immutable tags

Nichols and Zeckhauser 1982: In-Kind Redistribution

- In first-best full information model, no reason for in-kind transfers
  - In-kind transfer is tradeable at market price $\longrightarrow$ in-kind equivalent to cash
  - In-kind transfer non-tradeable $\longrightarrow$ in-kind inferior to cash
- Nichols and Zeckhauser: potential rationale for in-kind transfers emerges in Mirrlees-type model with informational constraints
  - With heterogeneity in preferences, may be able to relax IC constraints using in-kind transfers

Nichols and Zeckhauser: Simple Illustration

- Consider a soup kitchen as an in-kind transfer policy
- Let $S =$ soup and $W =$ wait in minutes
- Two agents: poor (P) and rich (R)
- Utility functions are increasing in S and decreasing in W:

$$\begin{aligned} U_p &= 2S - .5W \\ U_r &= S - 1W \end{aligned}$$

- R has higher disutility from waiting and lower utility from soup
- Social welfare

$$SWF = U_p + U_r$$

Soup Kitchen without Wait: Cash Transfer

- With a total of \$100 in soup to give away and no wait times, the soup will be split between the two agents
- Both get some utility from soup, so both will claim it
- Assume that they split it equally, resulting in

$$\begin{aligned} U_p &= 100 \\ U_r &= 50 \\ SWF &= 150 \end{aligned}$$

- Equivalent to a cash-transfer program that pays each agent \$50

Soup Kitchen with Wait Times: In-Kind Transfer

- Now suppose we impose wait time of 51 minutes
  - R leaves - not worth it to him for \$50 in food - gets $U_p = 0$
  - P gets utility of $200 - 25.5 = 174.5$
- Social welfare with in-kind transfer (wait time) greater than cash transfer (no wait time)
- Targeting gains outweighing efficiency losses from ordeal
- Scope for such targeting depends upon degree of heterogeneity in preferences

Income Taxation as Insurance (Varian 1980)

- Important limitiation of Mirrlees model: no ex-post uncertainty
  - Once skill type is revealed, agent controls income perfectly
- In practice, there is considerable ex-post uncertainty in incomes (e.g. unemployment shocks)
- In this case, a progressive tax system could provide insurance
  - Do not want 100% insurance for moral hazard reasons
  - But some insurance desirable if individuals are risk averse

Varian: Taxation as Insurance

- Income $z = e + \epsilon$ where e is effort and ϵ is a random noise
- Government observes only z and sets a tax schedule based on z
- Individual utility
$$U = Eu(z - T(z)) - e$$
- Chooses $e = e^*$ to maximize this utility
- Effort e low if tax schedule very redistributive
- Government chooses $T(.)$ to maximize indirect utility: trade-off insurance vs incentives
- Optimal tax system depends on parameters similar to those in Mirrlees model

Varian Model: Private Insurance

- Varian model has received less attention than Mirrlees model
- One reason: government is not better than private market in providing such insurance
- In adverse selection (e.g. Mirrlees) models, only government can improve redistributive outcomes once skills are revealed to agents
- Agents cannot write contracts behind veil of ignorance
- In pure moral hazard model with ex-post information revelation, private markets should in principle reach optimum themselves
- In practice, firms offer wage contracts that provide some insurance against bad luck
    - Ex: tenure system in universities, increase of pay with job tenure, severance payments

Income Taxation and Social Insurance

- Two potential approaches to addressing private insurance provision

1. Optimal taxation with endogenous private insurance
    - Not clear how to model and measure endogenous private insurance
    - See Golosov and Tsyvinski (2007) and Chetty and Saez (2009) for some attempts
2. Focus on specific shocks where private markets are thought to be quite limited
    - Unemployment, disability, injury on the job
    - Not just general insurance against wage earnings fluctuations
    - Motivates literature on optimal social insurance

Public Economics Lectures
Part 5: Income Taxation and Labor Supply

Raj Chetty and Gregory A. Bruich

Harvard University

Fall 2009

Outline

1. Labor Supply Elasticity Estimation: Methodological Issues
2. Non-linear budget set methods
3. Estimates of hours/participation elasticities
4. Responses to low-income transfer programs
5. Intertemporal Labor Supply Models
6. Elasticity of Taxable Income
7. Macro Estimates
8. Implications for Preference Parameters

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
- Surveys in public economics:
    - Hausman (1985) *Handbook of Public Economics* vol 1
    - Moffitt (2003) *Handbook of Public Economics* vol 4
    - Saez, Slemrod, and Giertz (2009) survey in prep for JEL

Theoretical Issues in Estimation

- Labor supply elasticity is a parameter of fundamental importance for income tax policy
  - Optimal tax rate depends inversely on $\varepsilon^c = \frac{\partial \log l}{\partial \log w}_{U=\overline{U}}$, the compensated wage elasticity of labor supply
- First discuss econometric issues that arise in estimating ε^c
- Baseline model: (1) static, (2) linear tax system, (3) pure intensive margin choice, (4) single hours choice, (5) no frictions

Baseline Labor-Leisure Choice Model: Key Assumptions

1. Static model
2. Intensive-margin, one dimensional choice
3. No frictions or adjustment costs
4. Linear tax system

Static Model: Setup

- Let c denote consumption and l hours worked
- Normalize price of c to one
- Agent has utility $u(c, l) = c - a\frac{l^{1+1/\varepsilon}}{1+1/\varepsilon}$
- Agent earns wage w per hour worked and has y in non-labor income
- With tax rate τ on labor income, Individual solves

$$\max u(c, l) \text{ s.t. } c = (1-\tau)wl + y$$

Labor Supply Behavior

- First order condition
$$(1-\tau)w = al^{1/\varepsilon}$$
- Yields labor supply function
$$l = \alpha + \varepsilon \log(1-\tau)w$$
- Here y does not matter because u is quasilinear
- Log-linearization of general utility $u(c,l)$ would yield a labor supply fn of the form:
$$l = \alpha + \varepsilon \log(1-\tau)w - \eta y$$
- Can recover ε^c from ε and η using Slutsky equation

Problems with OLS Estimation of Labor Supply Equation

1. Econometric issues
   - Unobserved heterogeneity [tax instruments]
   - Measurement error in wages and division bias [tax instruments]
   - Selection into labor force [panel data]
2. Extensive vs. intensive margin responses [participation models]
3. Non-hours responses [taxable income]
4. Incorporating progressive taxes [non-linear budget set methods]

Econometric Problem 1: Unobserved Heterogeneity

- Early studies estimated elasticity using cross-sectional variation in wage rates
- Problem: unobserved heterogeneity
- Those with high wages also have a high propensity to work
- Cross-sectional correlation between w and h likely to yield an upward biased estimate of ε
- Solution: use taxes as instruments for $(1-\tau)w$

Econometric Problem 2: Measurement Error/Division Bias

- Wage w is typically not observed; backed out from dividing earnings by reported hours
- When hours are measured with noise, this can lead to "division bias"
- Let l^* denote true hours, l observed hours
- Compute $w = \frac{e}{l}$ where e is earnings

$$\Rightarrow \quad \log l = \log l^* + \mu$$
$$\Rightarrow \quad \log w = \log e - \log l = \log e - \log l^* - \mu = \log w^* - \mu$$

Measurement Error and Division Bias

- Mis-measurement of hours causes a spurious link between hours and wages

- Estimate a regression of the following form:

$$\log l = \beta_1 + \beta_2 \log w + \varepsilon$$

- Then

$$E\widehat{\beta}_2 = \frac{cov(\log l, \log w)}{var(\log w)} = \frac{cov(\log l^* + \mu, \log w^* - \mu)}{var(\log w) + var(\mu)}$$

- Problem: $E\widehat{\beta}_2 \neq \varepsilon$ because orthogonality restriction for OLS violated

- Ex. workers with high mis-reported hours also have low imputed wages, biasing elasticity estimate downward

- Solution: tax instruments again

Econometric Problem 3: Selection into Labor Force

- Consider model with fixed costs of working, where some individuals choose not to work
- Wages are unobserved for non-labor force participants
- Thus, OLS regression on workers only includes observations with $l_i > 0$
- This can bias OLS estimates: low wage earners must have very high unobserved propensity to work to find it worthwhile
- Requires a selection correction (e.g. Heckit, Tobit, or ML estimation)
- See Killingsworth and Heckman (1986) for implementation
- Current approach: use panel data to distinguish entry/exit from intensive-margin changes

Extensive vs. Intensive Margin

- Related issue: want to understand effect of taxes on labor force participation decision
- With fixed costs of work, 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:

$$P = \phi(\alpha + \varepsilon \log(1-\tau) - \eta y)$$

where $P \in \{0,1\}$ is an indicator for whether the individual works

- Function ϕ typically specified as logit, probit, or linear prob model
- Note: here it is critical to have tax variation; regression cannot be run with wage variation

Non-Hours Responses

- Traditional literature focused purely on hours of work and labor force participation
- Problem: income taxes distort many margins beyond hours of work
  - More important responses may be on those margins
  - Hours very hard to measure (most ppl report 40 hours per week)
- Two solutions in modern literature:
  - Focus on taxable income (wl) as a broader measure of labor supply (Feldstein 1995)
  - Focus on subgroups of workers for whom hours are better measured, e.g. taxi drivers

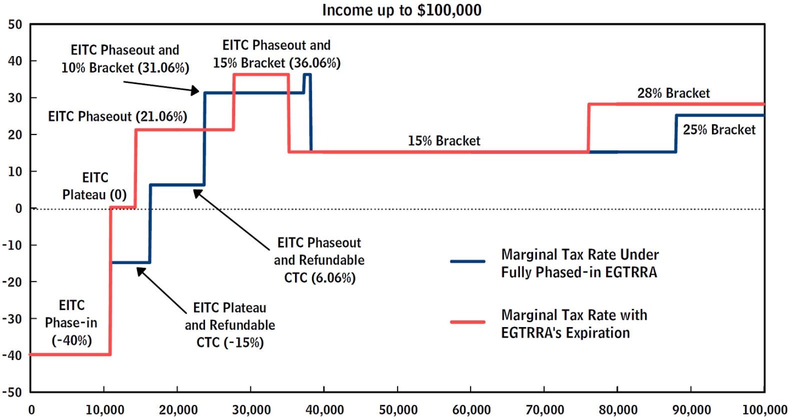
Progressive Taxes and Labor Supply

- OLS regression specification is derived from model with a single linear tax rate
- In practice, income tax systems are non-linear
- Consider effect of US income tax code on budget sets

Figure 8.

Effective Marginal Federal Income Tax Rates for a Married Couple with Two Children Before and After the Expiration of EGTRRA

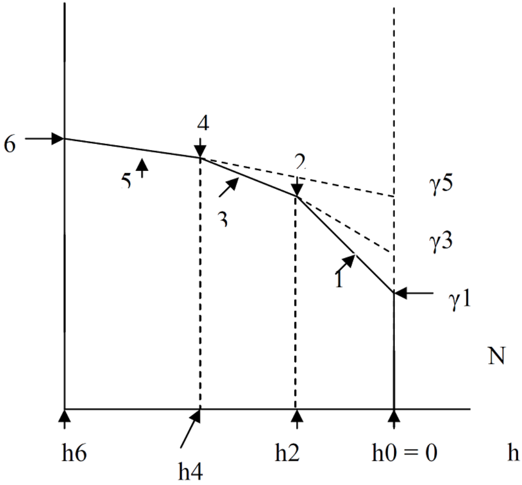
(Percent)



Note: EGTRRA is the 2001 Bush Tax Reform

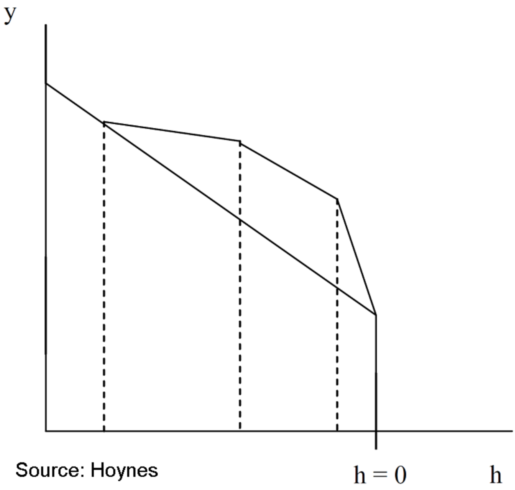
Source: Congressional Budget Office 2005

Example 1: Progressive Income Tax



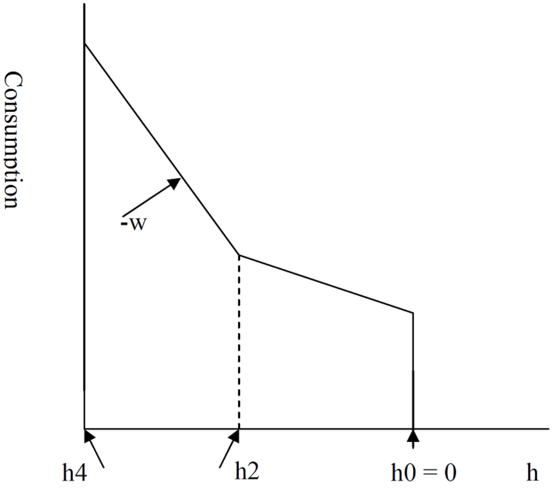
Source: Hoynes

Example 2: EITC



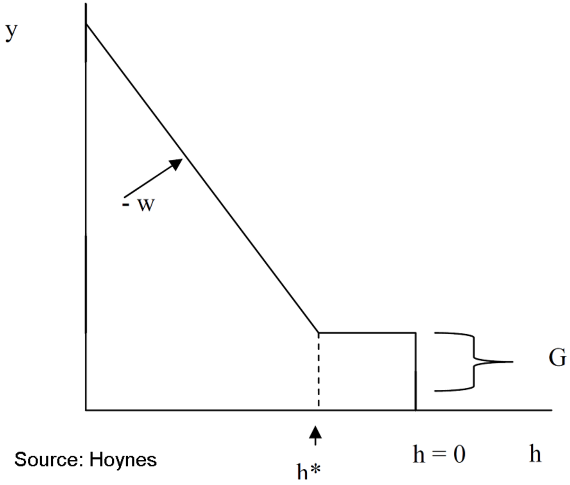
Source: Hoynes

Example 3: Social Security Payroll Tax Cap



Source: Hoynes

Example 4: Negative Income Tax



Source: Hoynes

Progressive Taxes and Labor Supply

- Non-linear budget set creates two problems:

  1. Model mis-specification: OLS regression no longer recovers structural elasticity parameter ε of interest
     - Two reasons: (1) underestimate response because people pile up at kink and (2) mis-estimate income effects
  2. Econometric bias: τ_i depends on income $w_i l_i$ and hence on l_i
     - Tastes for work are positively correlated with $\tau_i \rightarrow$ downward bias in OLS regression of hours worked on net-of-tax rates

- Solution to problem #2: only use reform-based variation in tax rates
- But problem #1 requires fundamentally different estimation method

Hausman: Non-linear Budget Constraints

- Hausman pioneered structural approach to estimating elasticities with non-linear budget sets

- Assume an uncompensated labor supply equation:

$$l = \alpha + \beta w(1 - \tau) + \gamma y + \epsilon$$

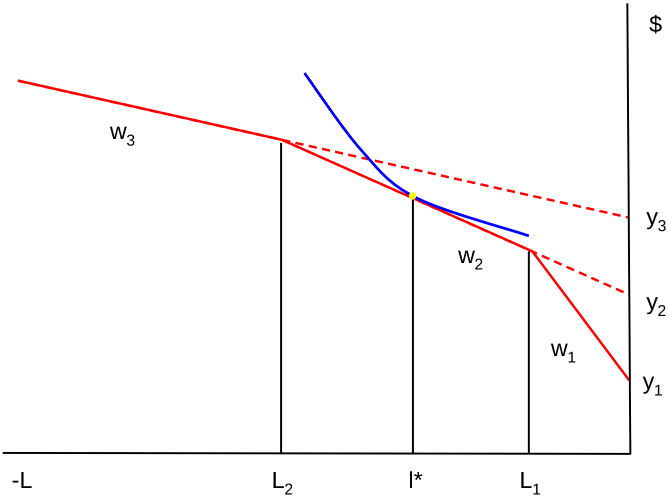
- Error term ϵ is normally distributed with variance σ^2

- 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

NLBS Likelihood Function

- Consider a two-bracket tax system
- Individual can locate on first bracket, on second bracket, or at the kink l_K
- Likelihood = probability that we see individual i at labor supply l_i given a parameter vector
- Decompose likelihood into three components
- Component 1: individual i on first bracket: $0 < l_i < l_K$

$$l_i = \alpha + \beta w_i (1 - \tau^1) + \gamma y^1 + \epsilon_i$$

- Error $\epsilon_i = l_i - (\alpha + \beta w_i (1 - \tau^1) + \gamma y^1)$. Likelihood:

$$L_i = \phi((l_i - (\alpha + \beta w_i (1 - \tau^1) + \gamma y^1)/\sigma)$$

- Component 2: individual i on second bracket: $l_K < l_i$. Likelihood:

$$L_i = \phi((l_i - (\alpha + \beta w_i (1 - \tau^2) + \gamma y^2)/\sigma)$$

Likelihood Function: Located at the Kink

- Now consider individual i located at the kink point
  - If tax rate is τ^1 and virtual income y^1 individual wants to work $l > l_K$
  - If tax is τ^2 and virtual income y^2 individual wants to work $l < l_K$
- These inequalities imply:

$$\begin{aligned} \alpha + \beta w_i(1-\tau^1) + \gamma y^1 + \epsilon_i &> l_K > \alpha + \beta w_i(1-\tau^2) + \gamma y^2 + \epsilon_i \\ l_K - (\alpha + \beta w_i(1-\tau^1) + \gamma y^1) &< \epsilon_i < l_K - (\alpha + \beta w_i(1-\tau^2) + \gamma y^2) \end{aligned}$$

- Contribution to likelihood is probability that error lies in this range:

$$\begin{aligned} L_i &= \Psi[(l_K - (\alpha + \beta w_i(1-\tau^2) + \gamma y^2))/\sigma] \\ &\quad - \Psi[(l_K - (\alpha + \beta w_i(1-\tau^1) + \gamma y^1))/\sigma] \end{aligned}$$

Maximum Likelihood Estimation

- Log likelihood function is $L = \sum_i \log L_i$
- Final step is solving
$$\max L(\alpha, \beta, \gamma, \sigma)$$
- In practice, likelihood function much more complicated because of more kinks, non-convexities, and covariates
- But basic technique remains the same

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, which is a larger issue with structural estimation
  2. No tax reforms, so does not solve fundamental econometric problem that tastes for work may be correlated with w

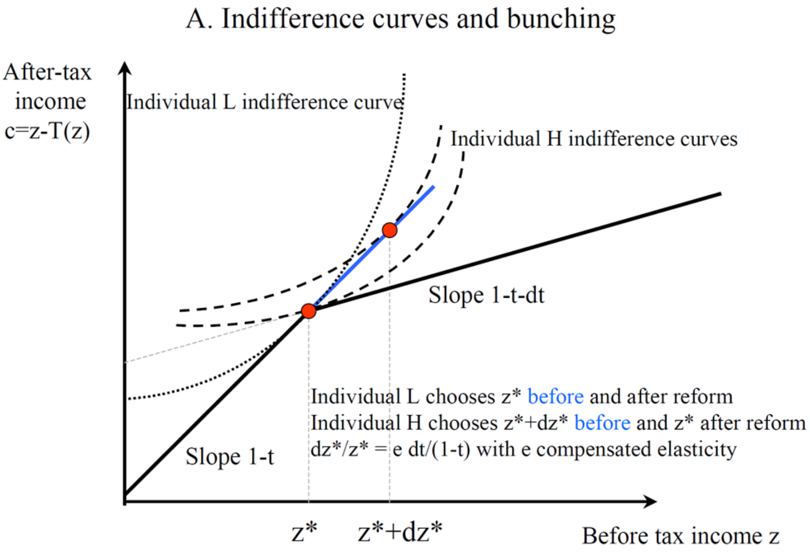
NLBS and Bunching at Kinks

- 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

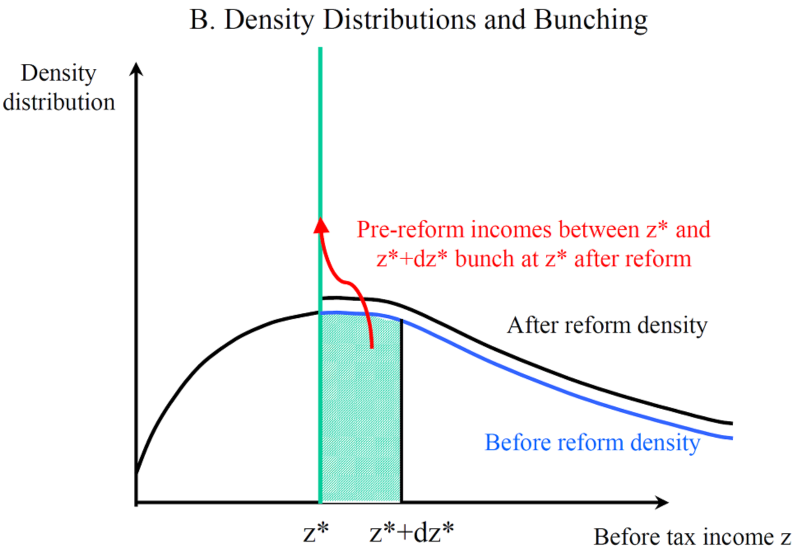
Saez 2009: Bunching at Kinks

- Saez observes that only non-parametric source of identification for elasticity in a cross-section is amount of bunching at kinks
- All other tax variation is contaminated by heterogeneity in tastes
- Develops method of using bunching at kinks to estimate the compensated taxable income elasticity
- Idea: if this simple, non-parametric method does not recover positive compensated elasticities, then little value in additional structure of NLBS models
- Formula for elasticity:

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



Source: Saez (2009)

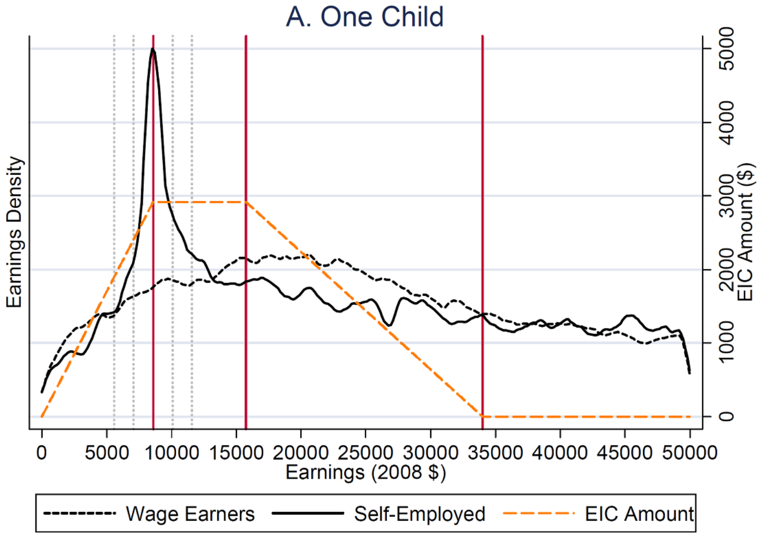


Source: Saez (2009)

Saez 2009: Bunching at Kinks

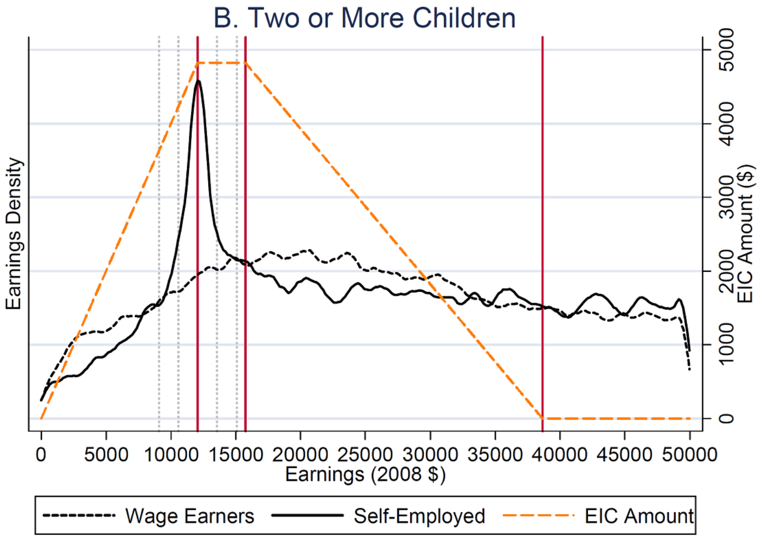
- Saez implements this method using individual tax return micro data (IRS public use files) from 1960 to 2004
- Advantage of dataset over PSID: very little measurement error
- Finds bunching around:
  - First kink point of the Earned Income Tax Credit, especially for self-employed
  - At threshold of the first tax bracket where tax liability starts, especially in the 1960s when this point was very stable
- However, no bunching observed around all other kink points

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



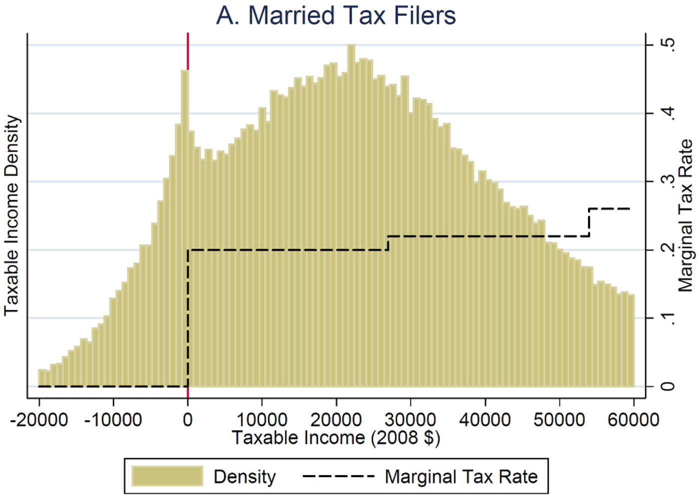
Source: Saez (2009)

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



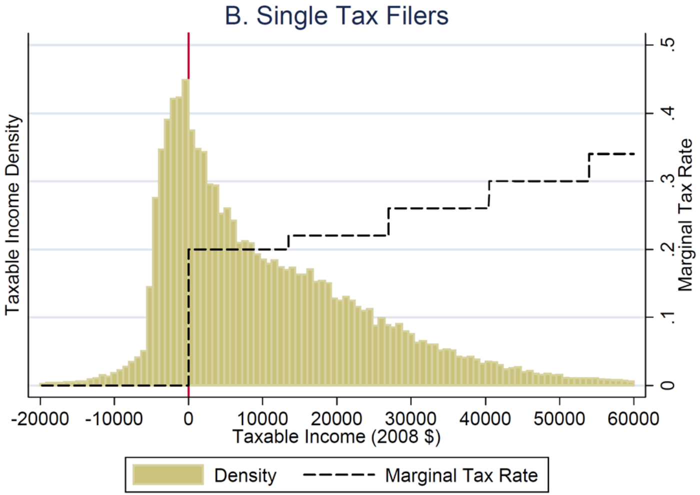
Source: Saez (2009)

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



Source: Saez (2009)

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

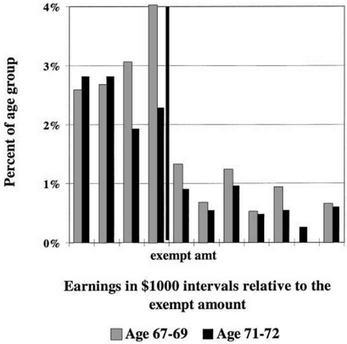


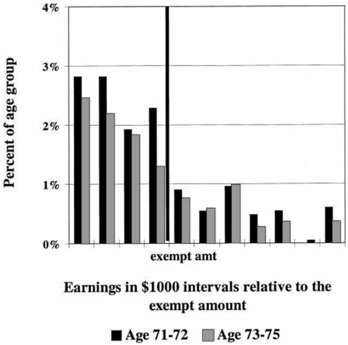
Source: Saez (2009)

Friedberg 2000: Social Security Earnings Test

- Uses CPS data on labor supply of retirees receiving Social Security benefits
- Studies bunching based on responses to Social Security earnings test
- Earnings test: phaseout of SS benefits above an exempt amount
- Phaseout rate varies by age group - 50%, 33%, 0 (lower for older workers)

Earnings distribution 1980-1981





Source: Friedberg (REStat 2000)

Friedberg: Estimates

- Estimates elasticities using Hausman method, finds relatively large compensated and uncompensated elasticities
- Ironically, lost social security benefits are considered delayed retirement with an actuarial adjustment of future benefits

  $\rightarrow$So the one kink where we do find real bunching is actually not real!

Borenstein 2009: Electricity Consumption

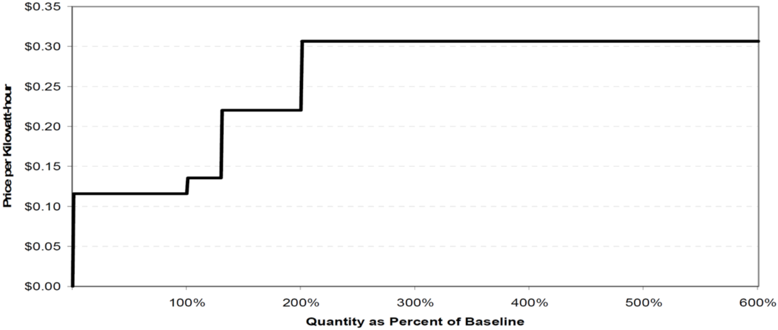


Figure 1: Increasing-Block Residential Electricity Tariff of Southern California Edison, 2006

Source: Borenstein 2009

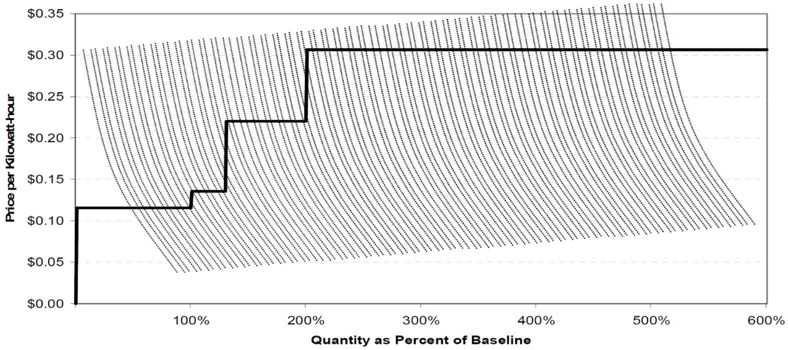


Figure 2: Uniform Distribution of Demand Functions and Bunched Quantities Demanded

Source: Borenstein 2009

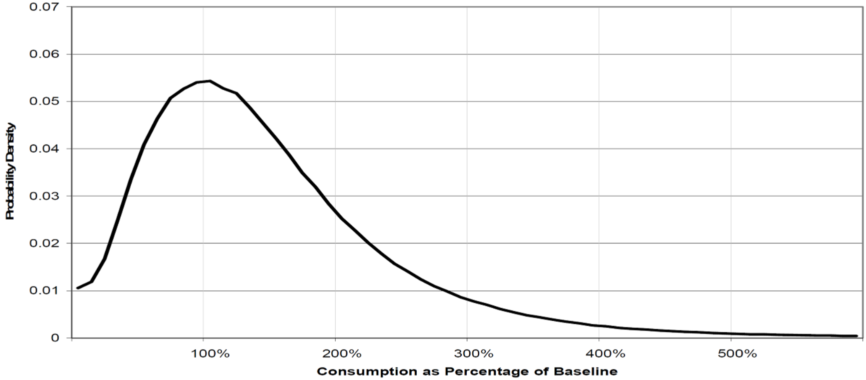


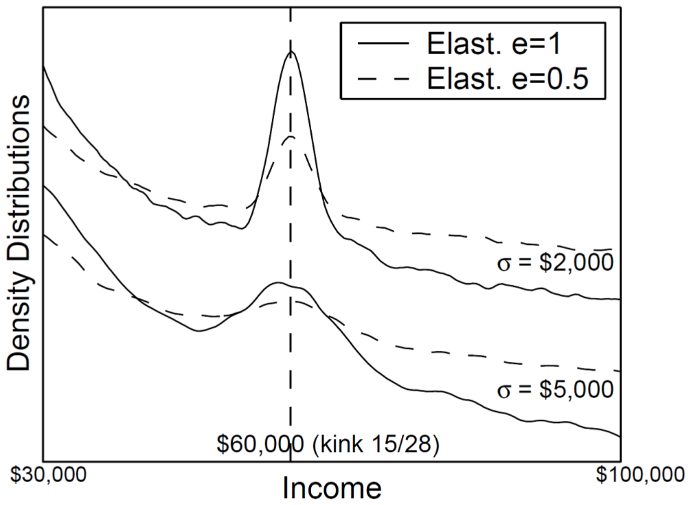
Figure 3: Distribution of Quantity Without Retail Bill Tiering

Source: Borenstein 2009

Why not more bunching at kinks?

1. True elasticity of response may be small
2. Randomness in income generation process
3. Information and salience
    - Liebman and Zeckhauser: "Schmeduling"
    - Chetty and Saez (2009): information significantly affects bunching in EITC field experiment
4. Adjustment costs and institutional constraints (Chetty et al 2009)

A. Simulation kink 15 to 28%



Source: Saez 2002

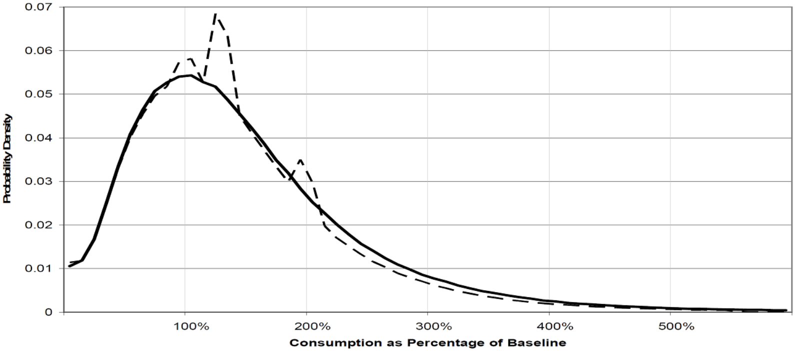


Figure 4: Simulated Distribution of Quantity Demanded Under SCE's 2006 Four-Tier Tariff

Source: Borenstein 2009

Chetty, Friedman, Olsen, and Pistaferri (2009)

- If workers face adjustment costs, may not reoptimize in response to tax changes of small size and scope in short run
  - Search costs, costs of acquiring information about taxes
  - Institutional constraints imposed by firms (e.g. 40 hour week)
- Could explain why macro studies find larger elasticities
- Question: How much are elasticity estimates affected by frictions?

Chetty et al. 2009: Model

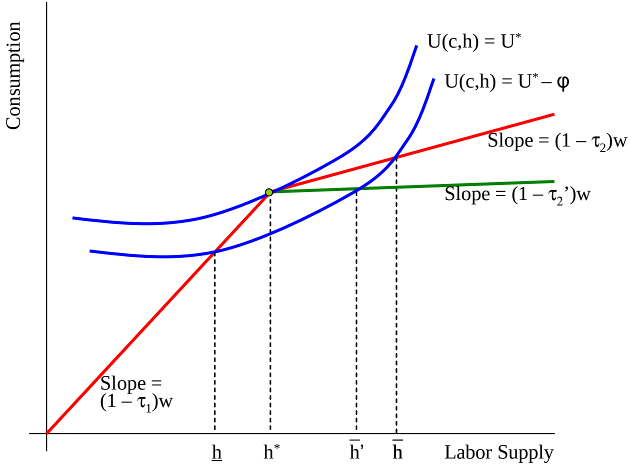
- Firms post jobs with different hours offers
- Workers draw from this distribution and must pay search cost to reoptimize
- Therefore not all workers locate at optimal choice
- Bunching at kink and observed responses to tax reforms attenuated

Chetty et al. 2009: 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 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
3. **[Search Costs]** Workers with lower search costs exhibit larger elasticities from individual bunching

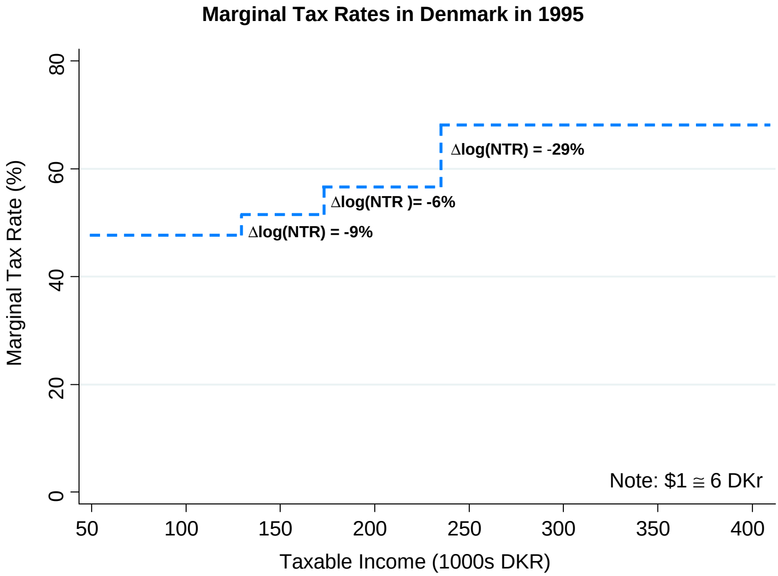
Cost of Bunching at Bracket Cutoff Points in Tax Schedule



Source: Chetty et al. 2009

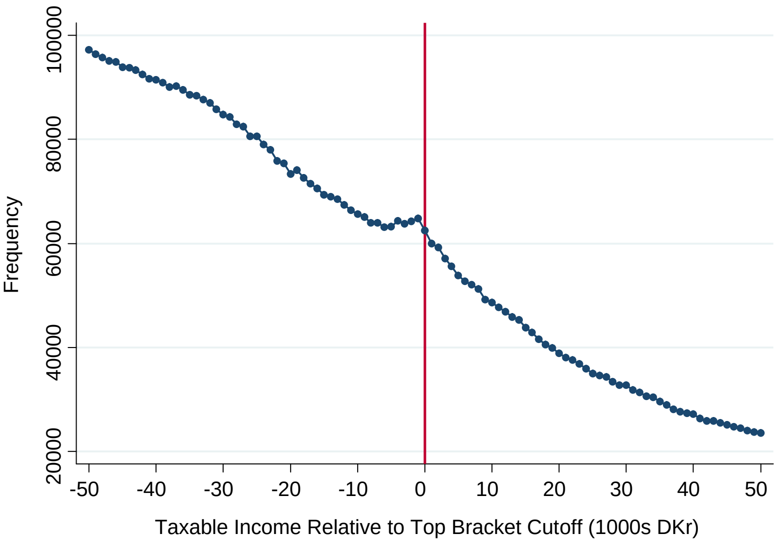
Chetty et al. 2009: Data

- Matched employer-employee panel data with admin tax records for full population of Denmark
    - Income vars: wage earnings, capital and stock income, pension contributions
    - Employer vars: tenure, occupation, employer ID
    - Demographics: education, spouse ID, kids, municipality
- Sample restriction: Wage-earners aged 15-70, 1994-2001
    - Approximately 2.42 million people per year

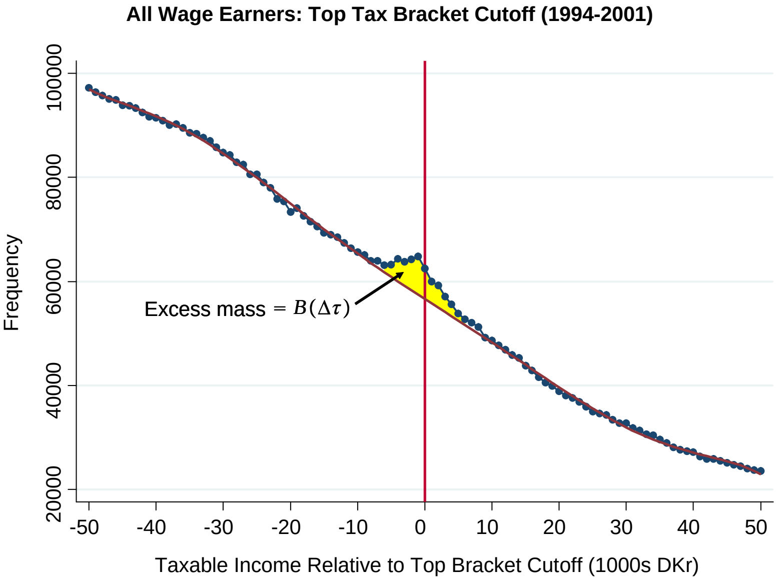


Source: Chetty et al. 2009

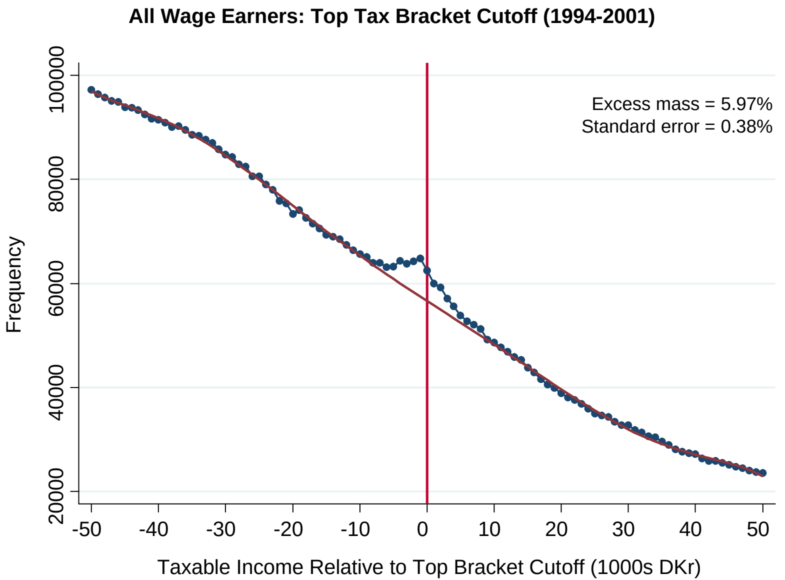
All Wage Earners: Top Tax Bracket Cutoff (1994-2001)



Source: Chetty et al. 2009



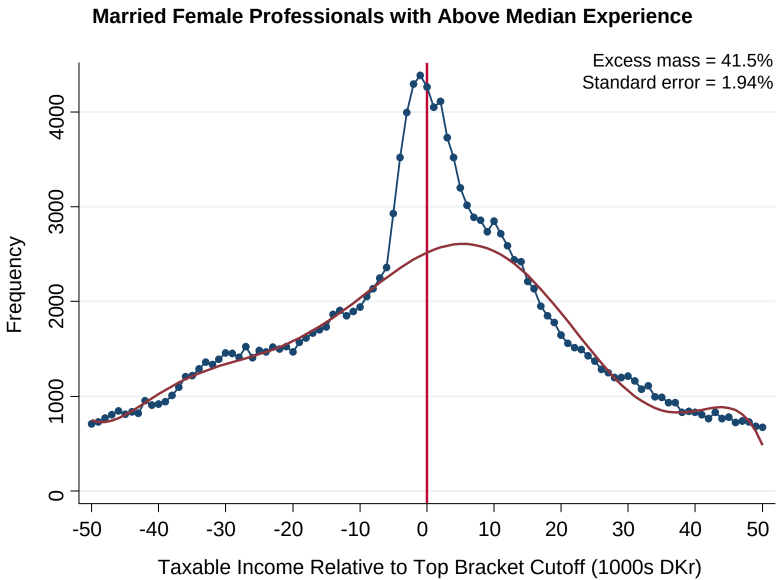
Source: Chetty et al. 2009



Source: Chetty et al. 2009

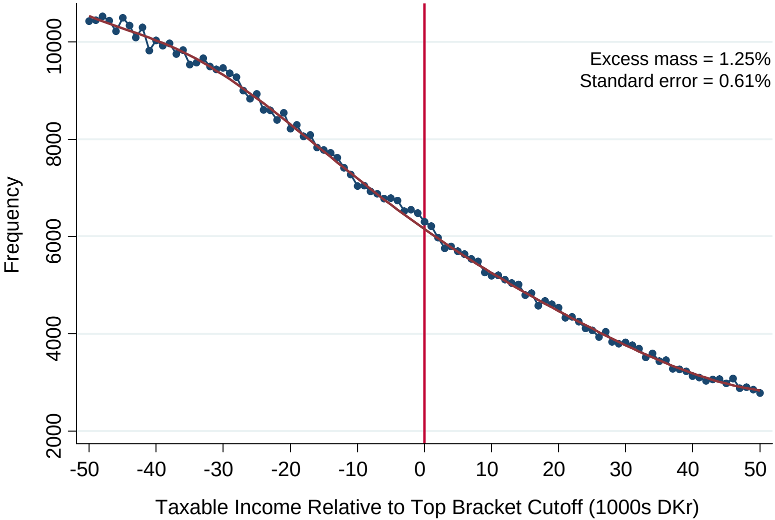
Source: Chetty et al. 2009

Source: Chetty et al. 2009

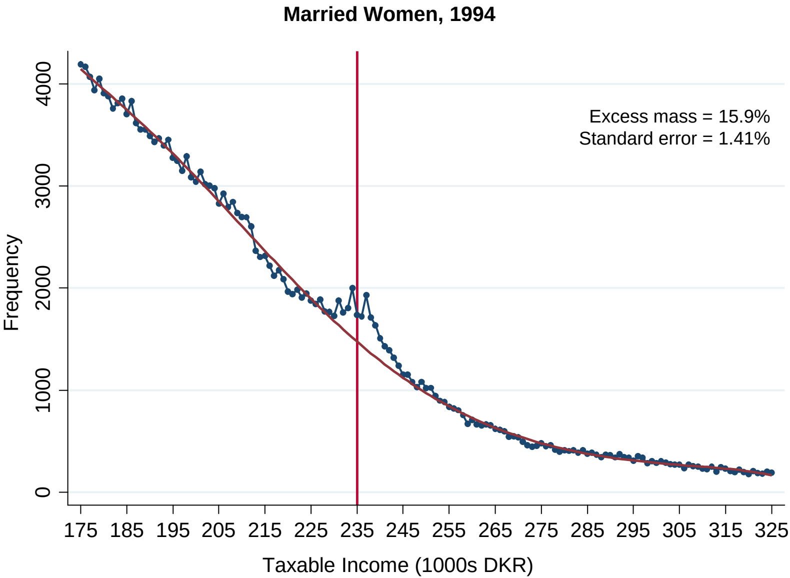


Source: Chetty et al. 2009

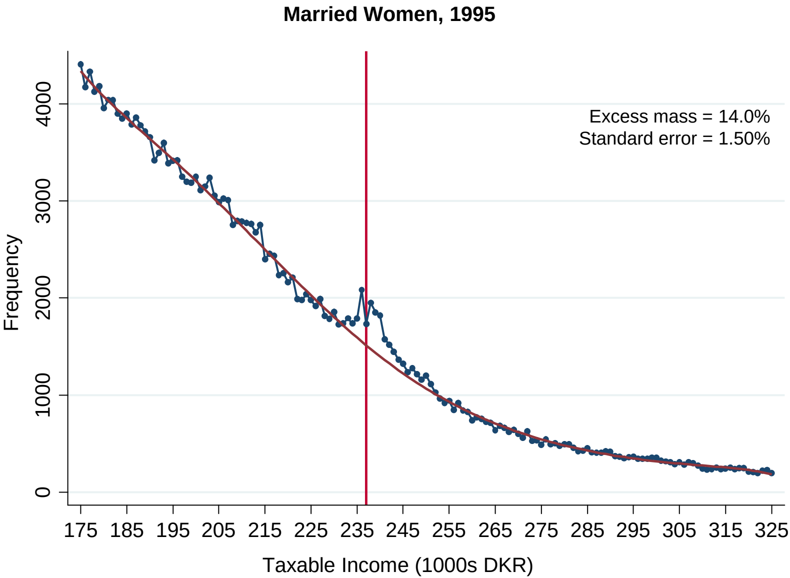
Military



Source: Chetty et al. 2009

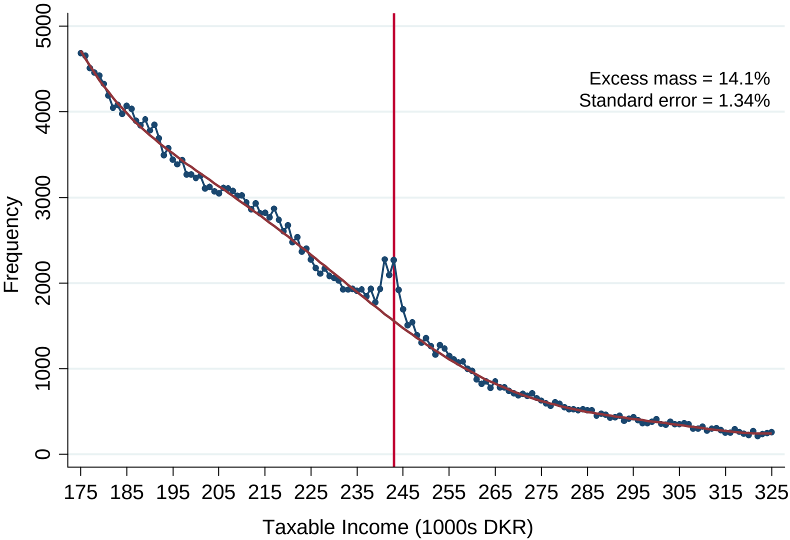


Source: Chetty et al. 2009

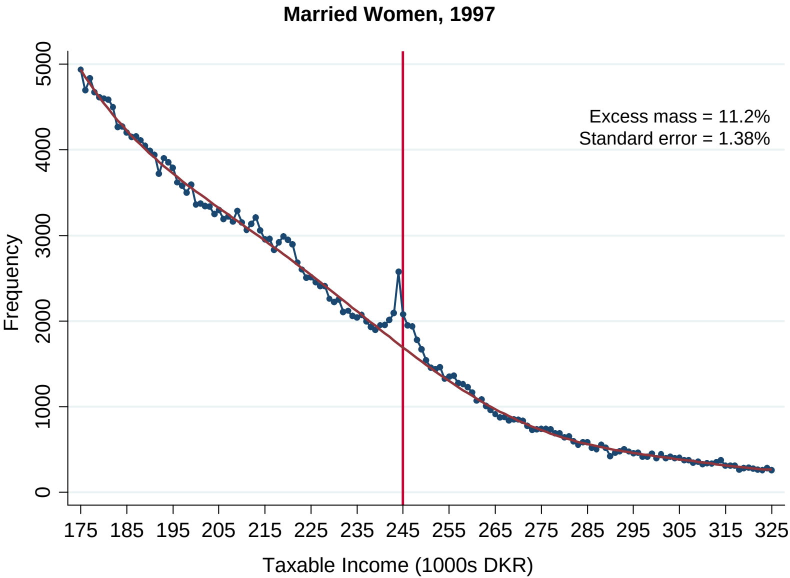


Source: Chetty et al. 2009

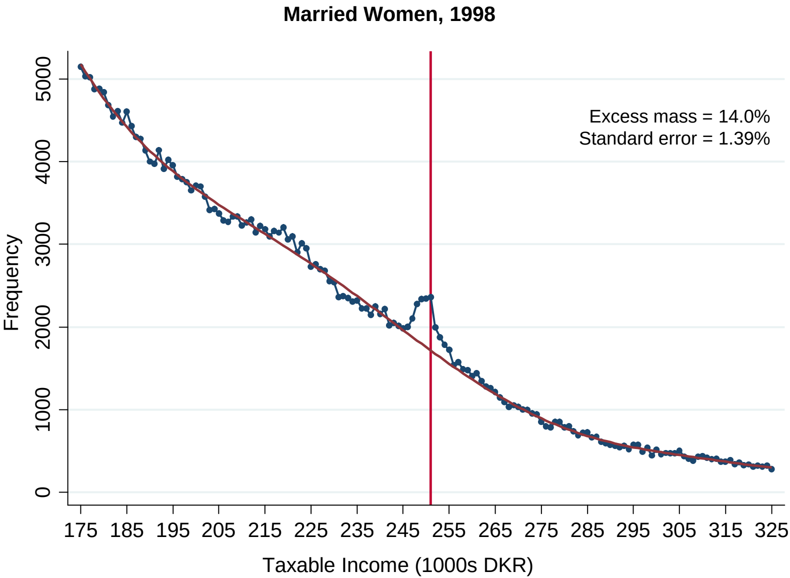
Married Women, 1996



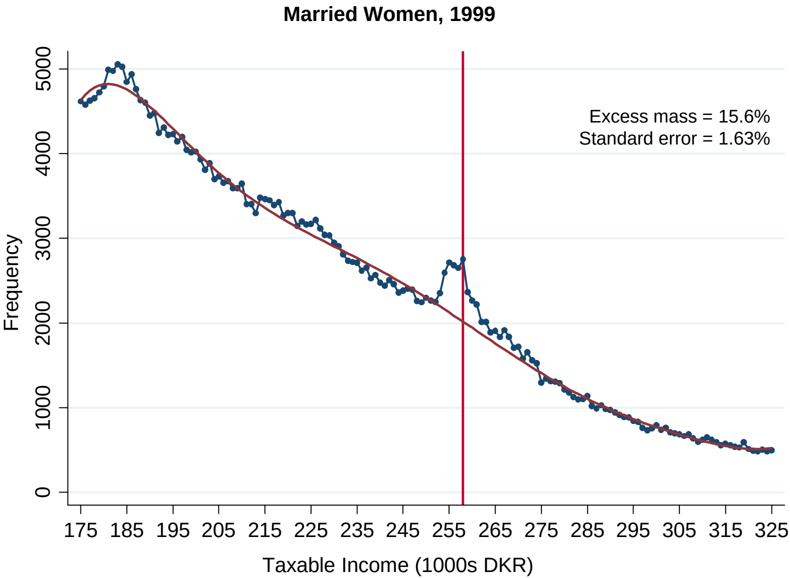
Source: Chetty et al. 2009



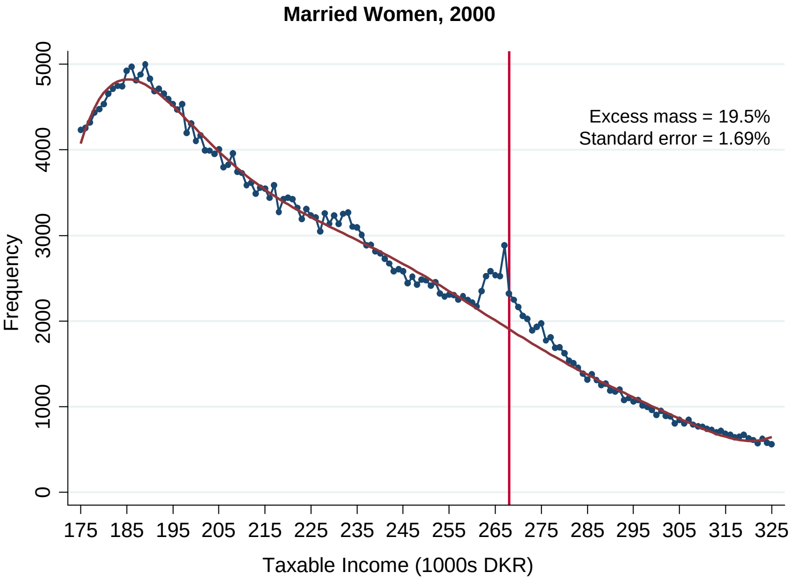
Source: Chetty et al. 2009



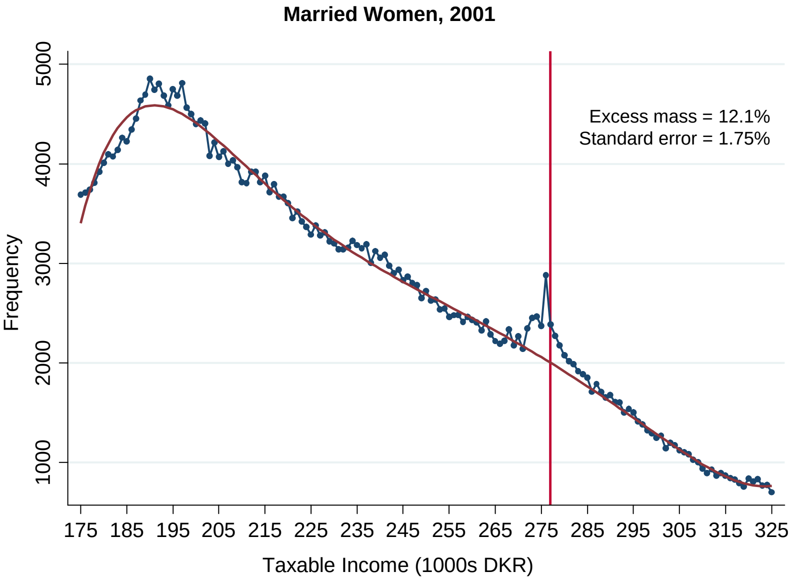
Source: Chetty et al. 2009



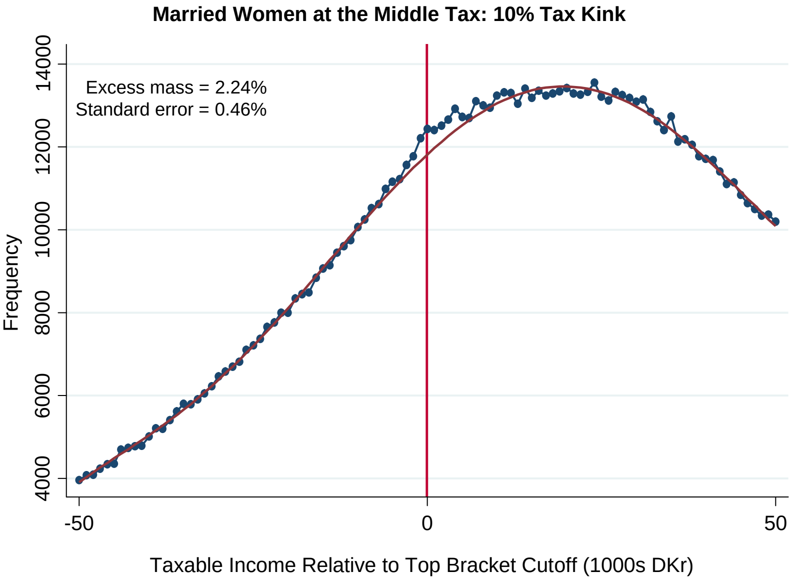
Source: Chetty et al. 2009



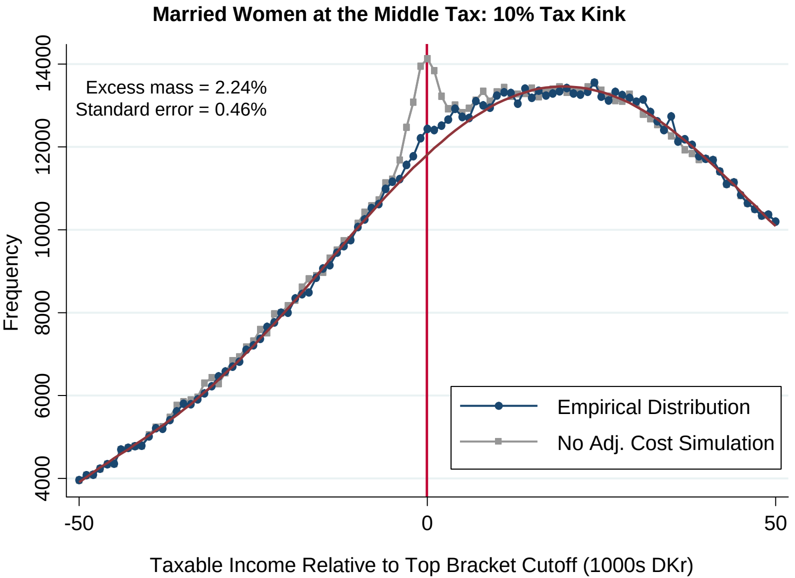
Source: Chetty et al. 2009



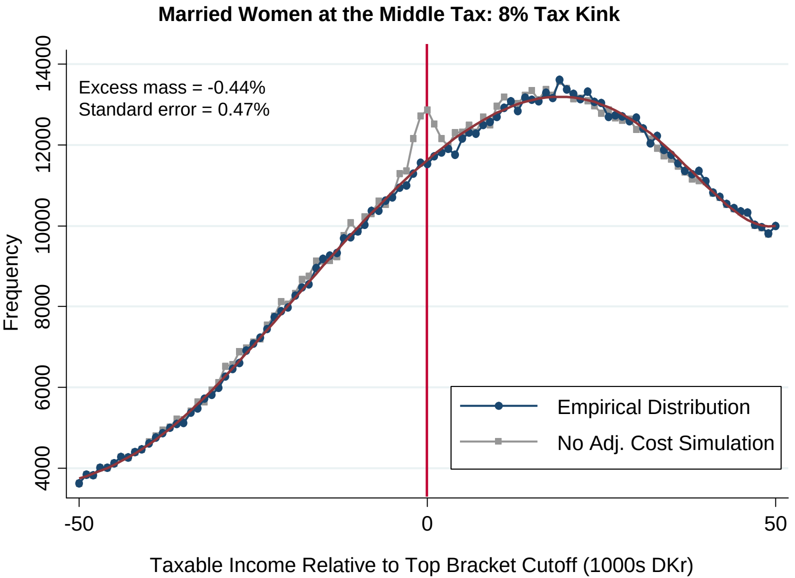
Source: Chetty et al. 2009



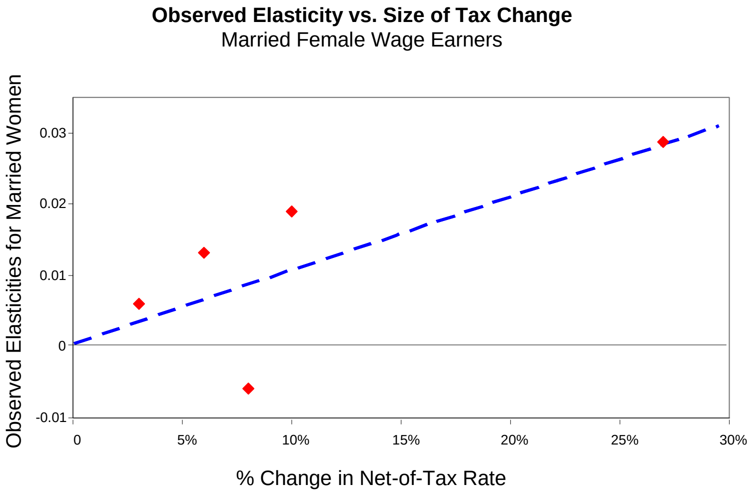
Source: Chetty et al. 2009



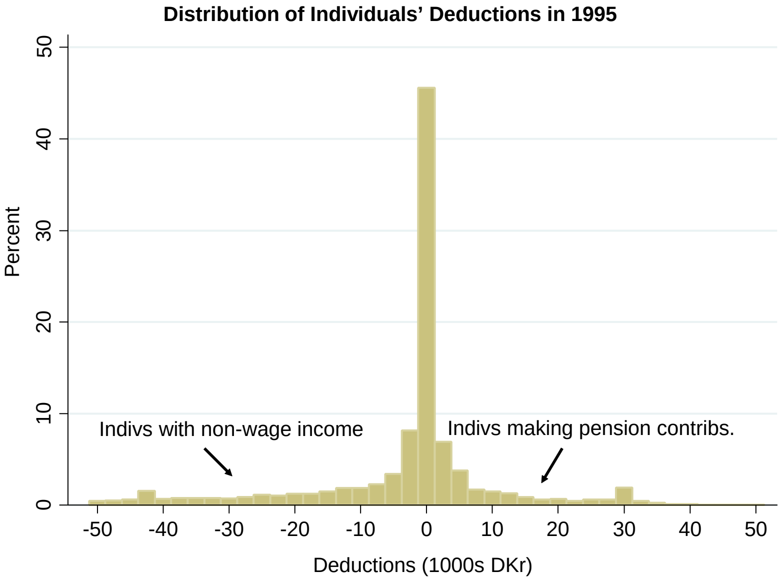
Source: Chetty et al. 2009



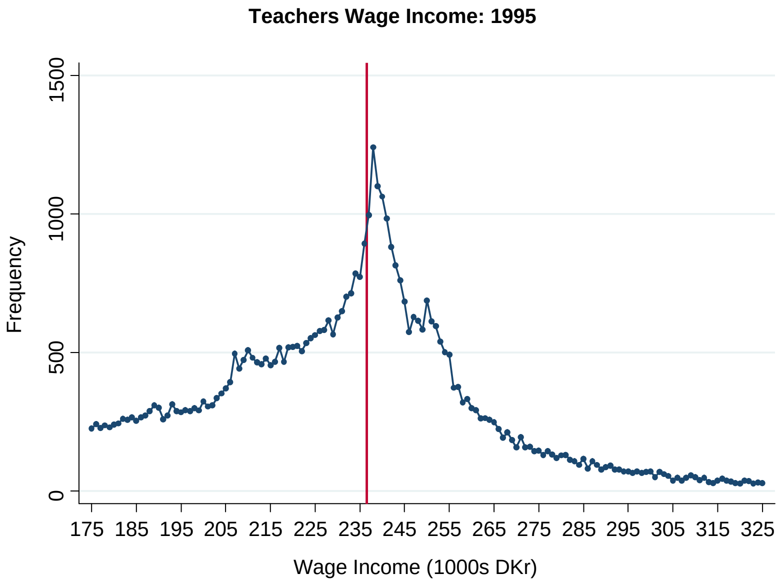
Source: Chetty et al. 2009



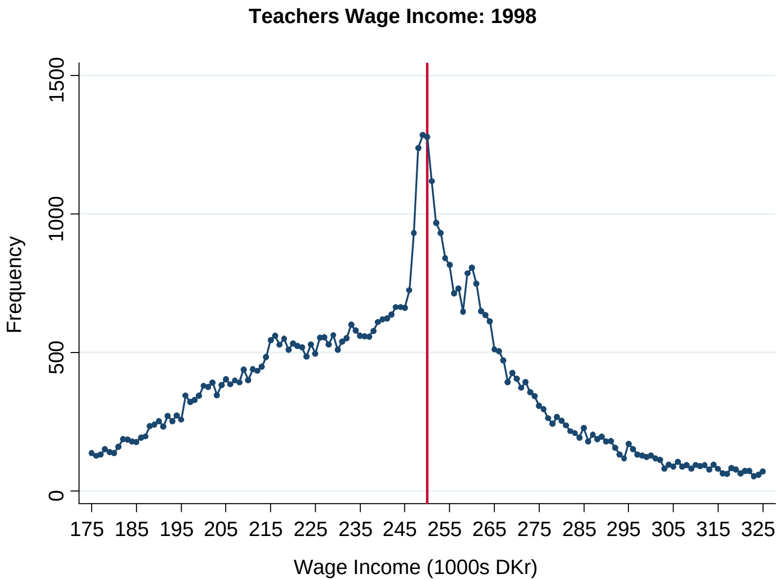
Source: Chetty et al. 2009



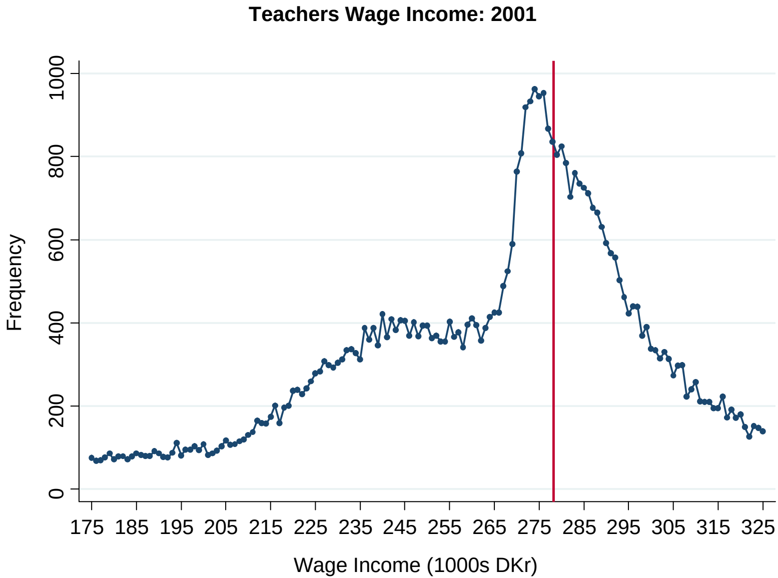
Source: Chetty et al. 2009



Source: Chetty et al. 2009

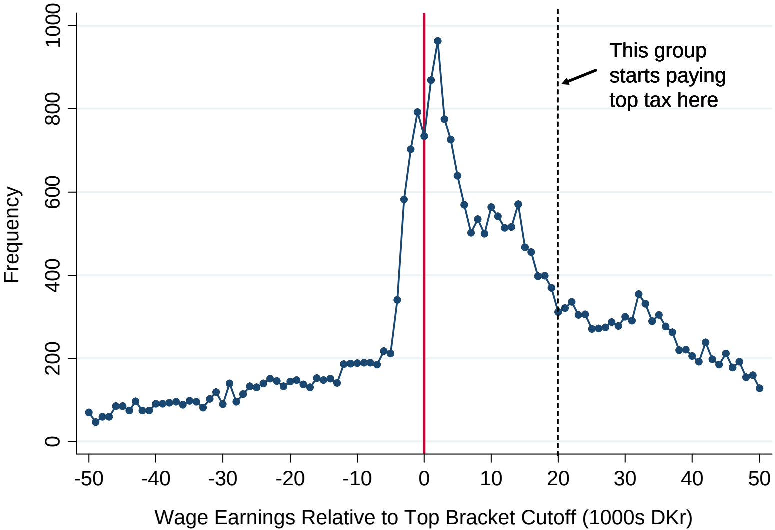


Source: Chetty et al. 2009

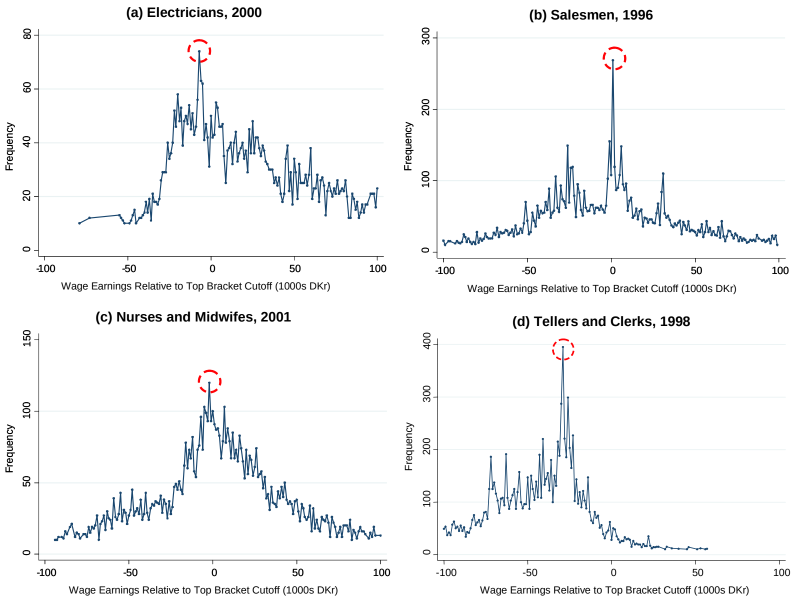


Source: Chetty et al. 2009

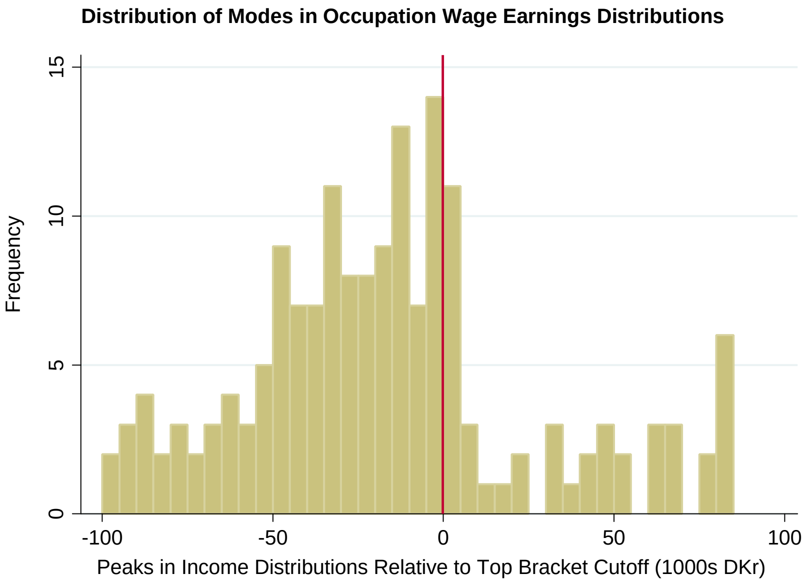
Wage Earnings: Teachers with Deductions > DKr 20,000



Source: Chetty et al. 2009

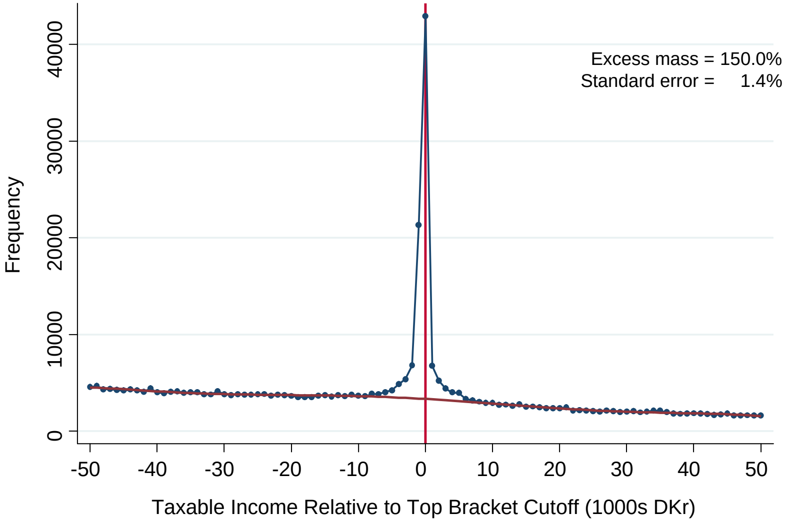


Source: Chetty et al. 2009

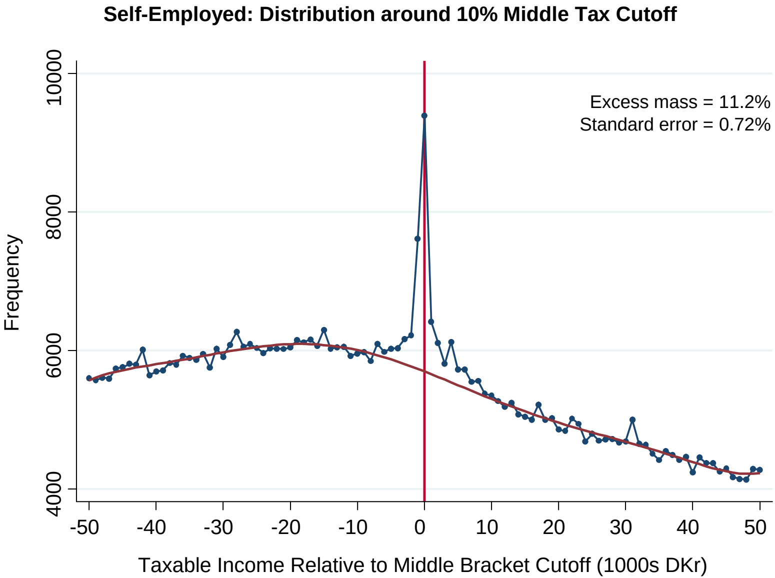


Source: Chetty et al. 2009

Self-Employed: Distribution around Top Tax Cutoff



Source: Chetty et al. 2009



Source: Chetty et al. 2009

Chetty et al. 2009: Results

- Search costs attenuate observed behavioral responses substantially
- Firm responses and coordination critical for understanding behavior: individual and group elasticities may differ significantly
- NLBS models may fit data better if these factors are incorporated
- Standard method of estimating elasticities using small tax reforms on same data yields close-to-zero elasticity estimate

Estimates of Hours and Participation Elasticities

- Return to simple model where we ignore non-linear budget set issues
- Large literature in labor economics estimates effects of taxes and wages on hours worked and participation
- Now discuss some estimates from this older literature

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: Ashenfelter and Plant 1990

- Present non-parametric evidence of labor supply effects
- Compare implied benefit payments to treated vs control households
- Difference in benefit payments aggregates hours and participation responses
- This is the relevant parameter for expenditure calculations and potentially for welfare analysis (revenue method of calculating DWL)
- Shortcoming: approach does not decompose estimates into income and substitution effects
- 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)

				Payments for Year of Experiment ($)			
G ($)	τ	Declining Tax Rate	Preexperimental Payment ($)	1	2	3	Postexperimental Payment ($)
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)

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

Source: Ashenfelter and Plant 1990

NIT Experiments: Findings

- Significant labor supply response but small overall
- Implied earnings elasticity for males around 0.1
- Implied earnings elasticity for women around 0.5
- Academic literature not careful to decompose response along intensive and extensive margin
- Response of women is concentrated along the extensive margin (can only be seen in official govt. report)
- Earnings of treated women who were working before the experiment did not change much

Problems with Experimental Design

Estimates from NIT not considered credible due to several shortcomings:

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. Response might be smaller than real reform b/c of GE effects

Social Experiments: Costs/Benefits

- Cost of NIT experiments: around \$1 billion (in today's dollars)
- Huge cost for a social experiment but trivial relative to budget of the US federal government (\$2 trillion)
- Should the government do more experimentation? Potential benefits:
  - Narrow the standard error around estimates
  - Allow implementation of better tax and redistribution policy

Instrumental Variable Methods

- Another strategy to overcome endogeneity is instrumenting for wage rate
- Mroz (1987): often-cited survey/meta-analysis of earlier studies
- Uses PSID to test widely-used IV's for married women's wage

$$\begin{aligned} l_i &= \alpha + \beta w + \gamma X + \varepsilon \\ w &= \theta Z + \mu \end{aligned}$$

- Uses Hausman specification/overidentification test to show that many instruments violate $EZ\varepsilon = 0$

Hausman Test

- Suppose you can divide instrument set into those that are credibly exogenous (Z) and those that are questionable (Z^*)
- Null hypothesis: both are exogenous
- Alternative hypothesis: Z^* is endogenous
- Compute IV estimate of β with small and large instrument set and test for equality of the coefficients
- Note that is often a very lower power test (accept validity if instruments are weak)

Mroz 1987: Setup and Results

- Uses background variables as "credibly exogenous "instruments
  - Parents' education, wife' age, education polynomials
- Tests validity of labor market experience, average hourly earnings, and previous reported wages
- Rejects validity of all three
- Shows that earlier estimates are highly fragile and unreliable
- Contributed to emerging view that policy variation (e.g., taxes) was necessary to really identify these elasticities properly

Tax Reform Variation (Eissa 1995)

- 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

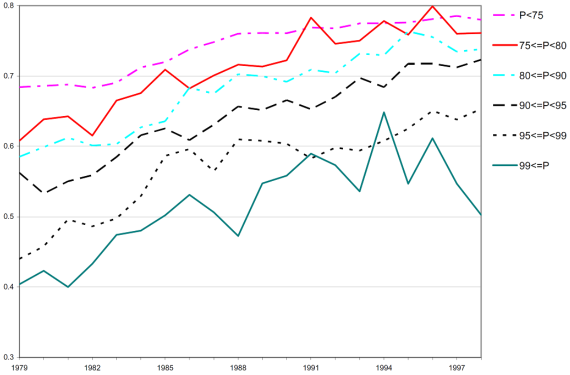
Eissa 1995: Results

- 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)

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.

Source: Liebman and Saez (2006)

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

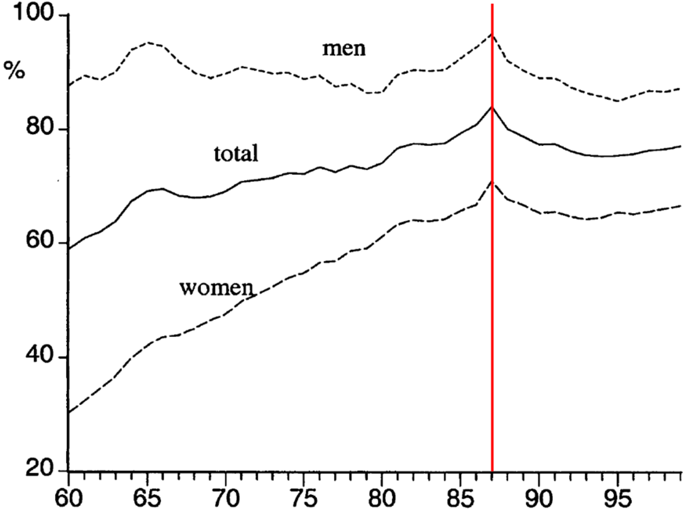


FIGURE 1. THE EMPLOYMENT RATE IN ICELAND, 1960–1996

Source: Bianchi, Gudmundsson, and Zoega 2001

Table 3—Changes in Labor Supply (Percent) Relative to the Average of 1986 and 1988

	Weeks		Earnings	
	Male	Female	Male	Female
Entry and exit in 1987	−1.4	−0.6	0.0	0.2
Δweeks (Δearnings)	6.6	2.0	8.9	2.9
Entry in 1988	−0.2	−0.8	0.0	0.0
Sum	5.0	0.6	8.9	3.1
Both sexes	5.6		12.0	

Source: Bianchi, Gudmundsson, and Zoega 2001

Table 5—Elasticity of Labor Supply for Workers Employed in 1986 (Calculated from Sample Averages)

	Number of observations		$\Delta L/L_A$		T/E		Elasticity	
	Male	Female	Male	Female	Male	Female	Male	Female
All workers	4,180	3,346	0.143	0.103	0.157	0.099	0.53	0.04
			(0.016)	(0.019)	(0.002)	(0.002)	(0.08)	(0.13)
Self-employed	812	264	0.241	−0.070	0.182	0.122	0.71	−0.34
			(0.051)	(0.043)	(0.006)	(0.007)	(0.21)	(0.26)
Employed	3,368	3,082	0.119	0.118	0.151	0.097	0.46	0.08
			(0.015)	(0.021)	(0.002)	(0.002)	(0.08)	(0.14)
Married	2,535	1,964	0.108	0.092	0.194	0.104	0.54	0.05
			(0.015)	(0.025)	(0.002)	(0.002)	(0.07)	(0.18)
Single	1,645	1,382	0.197	0.119	0.099	0.092	0.48	0.02
			(0.033)	(0.030)	(0.003)	(0.003)	(0.22)	(0.19)
With children	1,701	1,321	0.104	0.107	0.194	0.106	0.44	0.10
			(0.013)	(0.031)	(0.003)	(0.003)	(0.06)	(0.23)
Without children	2,479	2,025	0.169	0.101	0.135	0.093	0.61	0.00
			(0.025)	(0.025)	(0.003)	(0.003)	(0.13)	(0.15)
Self-employed and married	613	215	0.237	−0.126	0.195	0.120	0.75	−0.60
			(0.021)	(0.049)	(0.006)	(0.007)	(0.09)	(0.31)
Self-employed and single	199	49	0.253	0.174	0.143	0.129	0.58	0.35
			(0.199)	(0.082)	(0.014)	(0.023)	(0.84)	(0.42)

Note: Standard errors are in parentheses.

Source: Bianchi, Gudmundsson, and Zoega 2001

Bianchi, Gudmundsson, and Zoega 2001

- Large, salient change: $\Delta \log(1 - MTR) \approx 43\%$, much bigger than most studies
- Note that elasticities reported in paper are w.r.t. *average* tax rates:

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

$$\varepsilon_{E,T/E} = \frac{\sum (E_{87} - E_A)/E_A}{\sum T_{86}/E_{86}}$$

- Estimates imply hours elasticity w.r.t. marginal tax rate of roughly 0.29
    - Is this a Frisch or Hicksian elasticity?

Responses to Low-Income Transfer Programs

- Particular interest in treatment of low incomes in a progressive tax system: are they responsive to incentives?
- Complicated set of transfer programs in US
  - In-kind: food stamps, Medicaid, public housing, job training, education subsidies
  - Cash: TANF, EITC, SSI

Overall Costs of Anti Poverty Programs

- US government (fed+state and local) spent \$520bn in 2002 on income-tested programs
  - About 5% of GDP but 15% of \$3.5 Trillion govt budget (fed+state+local).
  - About 50% is health care (Medicaid)
  - Only \$100 billion in cash (1% of GDP, or 20% of transfer spending)

1996 Welfare Reform

- Largest change in welfare policy
- Reform modified AFDC cash welfare program to provide more incentives to work by
  1. Requiring recipients to go to job trainings
  2. Limiting the duration for which families able to receive welfare
  3. Reducing phase out to 66 cents of benefits per $1 earnings instead of 100% cliff
- Variation across states because Fed govt. gave block grants with guidelines
- EITC also expanded during this period: general shift from welfare to "workfare"

Monthly Welfare Case Loads: 1963-2000

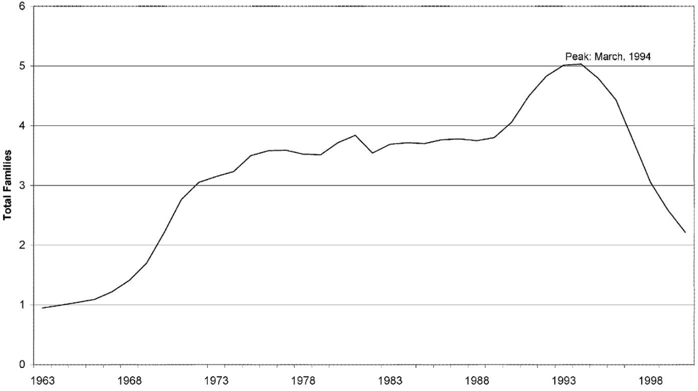


Fig. 1. Average monthly AFDC/TANF caseloads (1963 – 2000) (in millions).

Source: Meyer and Sullivan 2004

Welfare Reform: Two Empirical Questions

1. Incentives: did welfare reform actually increase labor supply
   - Test whether EITC expansions affect labor supply
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

Behavioral Responses to the EITC

1. Phase in:
   - Substitution effect: work more due to 40% inc. in net wage
   - Income effect: work less

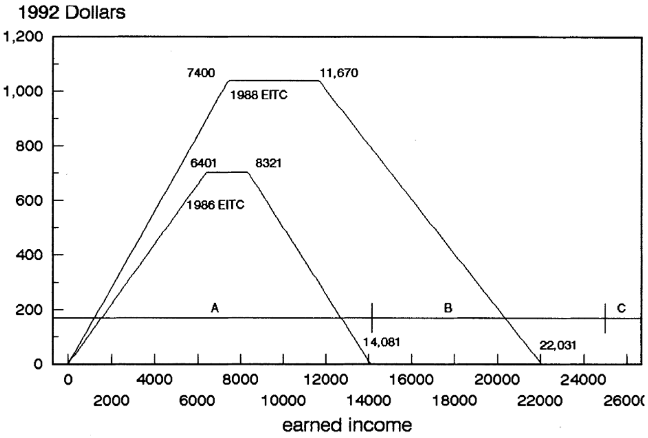
     →Net effect: ambiguous; probably work more

2. Plateau:
   - Pure income effect (no change in net wage)

     →Net effect: work less

3. Phase out:
   - Substitution effect: work less because reduces net wage to $0.80/hr
   - Income effect: also make you work less

     →Net effect: work less



Source: Eissa and Liebman 1996

Eissa and Liebman 1996

- Study labor force participation of single mothers before/after 1986-7 EITC expansion
- Limitation: this expansion was relatively small
- Diff-in-Diff strategy:
    - Treatment group: women with kids
    - Control group: women without kids

TABLE II

LABOR FORCE PARTICIPATION RATES OF UNMARRIED WOMEN

	Pre-TRA86 (1)	Post-TRA86 (2)	Difference (3)	Difference-in-differences (4)
A. *Treatment group:*				
With children [20,810]	0.729 (0.004)	0.753 (0.004)	0.024 (0.006)	
Control group:				
Without children [46,287]	0.952 (0.001)	0.952 (0.001)	0.000 (0.002)	*0.024 (0.006)*

Source: Eissa and Liebman 1996

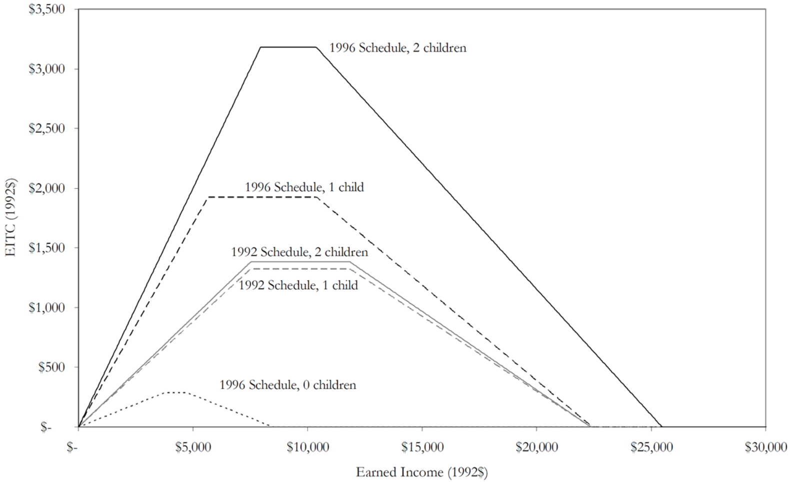
Eissa and Liebman: Results

- Find a small but significant DD effect: 2.4%
- Note: the labor force participation for women with/without children are not great comparison groups (70% LFP vs. +90%)
- Subsequent studies have used much bigger EITC expansions of the mid 1990s
    - Also find positive effects on labor force participation of single women/single mothers

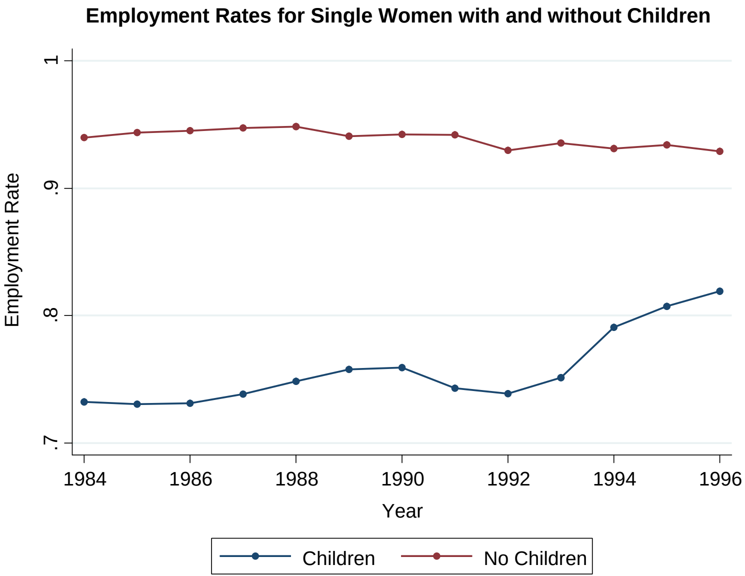
Meyer and Rosenbaum 2001

- Exploit the much bigger 1990s expansion in EITC
- Document dramatic (6 pp, 10%) increase in LFP for single women with children around EITC expansion
- No change for women without children
- Problem: expansion took place at same time as welfare reform
- Try to disentangle effects of welfare waivers, changes in AFDC and state taxes, etc. using state-level variation
- Bottom line: elasticity of participation w.r.t. tax/transfer incentives is significant
- But no clear elasticity estimate to use as an input for optimal policy

Figure 1. EITC Schedule, 1992 and 1996 by number of children



Source: Rothstein 2005



Source: Meyer and Rosenbaum 2001

Meyer and Rosenbaum 2001

- Analyze the introduction of EITC and Welfare waivers for the period 1984-1996 using CPS data
- Identification strategy: compare single mothers to single women without kids
- Key covariates in regression model:
  - EITC
  - AFDC benefits
  - Medicaid
  - Waivers
  - Training
  - Child Care

Meyer and Rosenbaum 2001

- From 1984-1996, the extra increase in single mom's relative to single women without kids is explained by:

  1. EITC expansion (60%)
  2. Welfare max benefit reduction (AFDC and food stamps) (25%)
  3. Medicaid if work (-10%) (insignificant and wrong sign)
  4. Welfare waivers (time limits) 15%
  5. Child care and training: 15%

Eissa and Hoynes 2004

- EITC based on family rather than individual income
- Study married couples with low earnings, recognizing that EITC *reduces* their incentive to work
- Married women with husband earning $10-15K are in the phase-out range and face high MTR's
  - Payroll tax 15%
  - EITC phase-out 20%
  - State and federal income tax 0-20%
- Similar identification strategy: compare those with and without kids

Eissa and Hoynes: Results

- Conclude that EITC expansions between 1984 and 1996:
  - Increased married men's labor force participation by 0.2%
  - Reduced married women's labor force participation by $>$1%
- Implies that the EITC is effectively subsidizing married mothers to stay at home and *reducing* total labor supply for married households

Meyer and Sullivan 2004

- Examine the consumption patterns of single mothers and their families from 1984–2000 using CEX data
- Question: did single mothers' consumption fall because they lost welfare benefits and were forced to work?

Total Consumption: Single Mothers 1984-2000

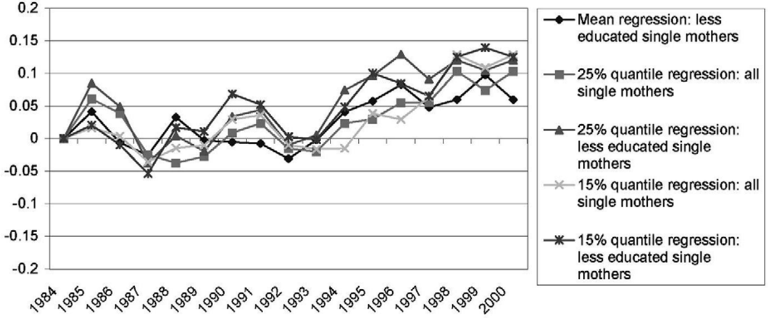


Fig. 2. Total consumption: single mothers, 1984–2000.

Source: Meyer and Sullivan 2004

Relative Consumption: single women with/without children

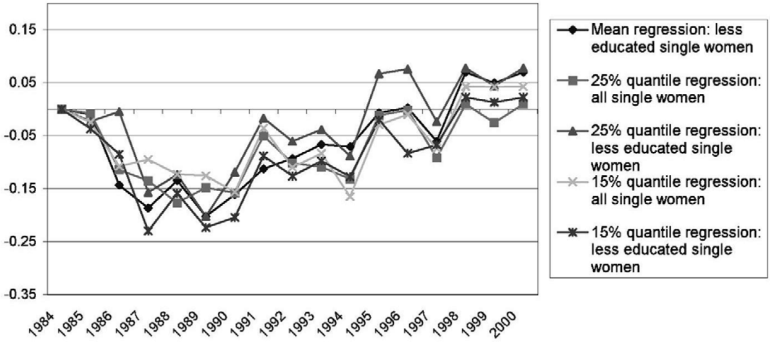


Fig. 3. Relative total consumption: single mothers vs. single women without children, 1984–2000.

Source: Meyer and Sullivan 2004

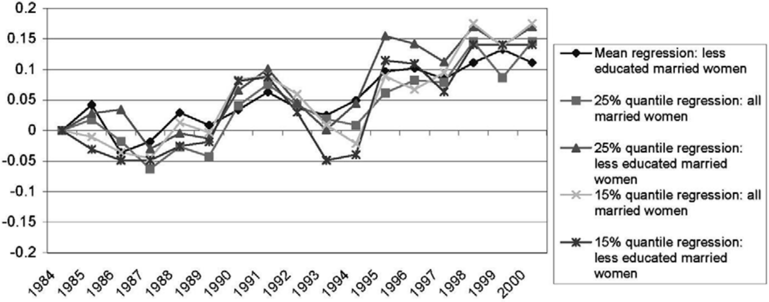


Fig. 4. Relative total consumption: single mothers vs. married mothers, 1984–2000.

Source: Meyer and Sullivan 2004

Meyer and Sullivan: Results

- Material conditions of single mothers did not decline in recent years, either in absolute terms or relative to single childless women or married mothers
- In most cases, evidence suggests that the material conditions of single mothers have improved slightly
- Question: is this because economy was booming in 1990s?
- Is workfare approach more problematic in current economy?

Other Behavioral Responses to Transfer Programs

- Bitler, Gelbach, and Hoynes (2005): distributional effects are very important in understanding welfare programs because of nonlinearities in bc $\rightarrow$ cannot just look at means
- Other studies have examined effects of low-income assistance programs on other margins such as family structure (divorce rate, number of kids) and find limited effects
- Empirical work on tagging and in-kind programs is more limited and is an important area for further research

Changing Elasticities: Blau and Kahn 2007

- Identify elasticities from 1980-2000 using grouping instrument

  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

Summary of Static Labor Supply Literature

1. Small elasticities for prime-age males
   - Probably institutional restrictions, need for one income, etc. prevent a short-run response
2. Larger responses for workers who are less attached to labor force
   - Married women, low incomes, retirees
3. Responses driven by extensive margin
   - Ext margin (participation) elasticity around 0.2
   - Int margin (hours) elasticity close 0

Intertemporal Models and the MaCurdy Critique

- What parameter do reduced-form regressions of labor supply on wages or taxes identify?
- MaCurdy critique: reduced-form studies did not identify any parameter of interest in a dynamic model
- Instead, estimate a mix of income effects, intertemporal substitution effects, and compensated wage elasticities
- MaCurdy (1981) develops a structural estimation method (two stage budgeting) to identify preference parameters in a life-cycle model of labor supply
- Chetty (2006) presents a simple exposition of two-stage budgeting

Life Cycle Model of Labor Supply

- General model is of the form:

$$\begin{aligned} & U(c_0, .., c_T, l_0, .., l_T) \\ \text{s.t. } & A_0 + \sum w_t l_t/(1+r)^t \geq \sum c_t/(1+r)^t (\lambda) \end{aligned}$$

- First order conditions:

$$\begin{aligned} U_{l_t}(c_0, .., c_T, l_0, .., l_T) + \lambda w_t/(1+r)^t &= 0 \\ U_{c_t}(c_0, .., c_T, l_0, .., l_T) + \lambda/(1+r)^t &= 0 \end{aligned}$$

- In the general case, $l_t(A_0, w_0, .., w_T)$ same as the multi-good choice – no generic results

Life Cycle Model: Time Separability

- By assuming time separability can rewrite the problem as:

$$U = \sum_{t=0}^{T} \beta^t u(c_t, l_t)$$

- Leads to simpler first order conditions

$$\begin{aligned} l_t &: \quad \beta^t u_{l_t} + \lambda w_t / (1+r)^t = 0 \\ c_t &: \quad \beta^t u_{c_t} + \lambda / (1+r)^t = 0 \end{aligned}$$

- Combining yields: $-u_l(l_t) = w_t u_c$
- Intratemporal f.o.c. same as in static model
- Intertemporal f.o.c.: $u_{c_t} / u_{c_{t+1}} = \beta(1+r)$

Dynamic Life Cycle Model: Policy Rules

- $\lambda = u_{c_0}$ is the marginal utility of initial consumption
- The two first order conditions imply that

$$\begin{aligned} l_t &= l(w_t, \lambda/(\beta(1+r))^t) \\ c_t &= c(w_t, \lambda/(\beta(1+r))^t) \end{aligned}$$

- Current labor and consumption choice depends on current w_t
- All other wage rates and initial wealth enter only through the budget constraint multiplier λ (MaCurdy 1981)
- Easy to see for separable utility:

$$\begin{aligned} u(c,l) &= u(c) - v(l) \\ \Rightarrow v'(l_t) &= \lambda w_t/[\beta(1+r)]^t \\ \Rightarrow l_t &= v'^{-1}(\lambda w_t/[\beta(1+r)]^t) \end{aligned}$$

- Sufficiency of λ greatly simplifies solution to ITLS model

Dynamic Life Cycle Model: Frisch Elasticity

- Frisch intertemporal labor supply elasticity defined as:

$$\delta = (\frac{w_t}{l_t})\frac{\partial l}{\partial w_t}|_\lambda$$

- Experiment: change wage rate in one period only, holding all other wages, and consumption profile constant

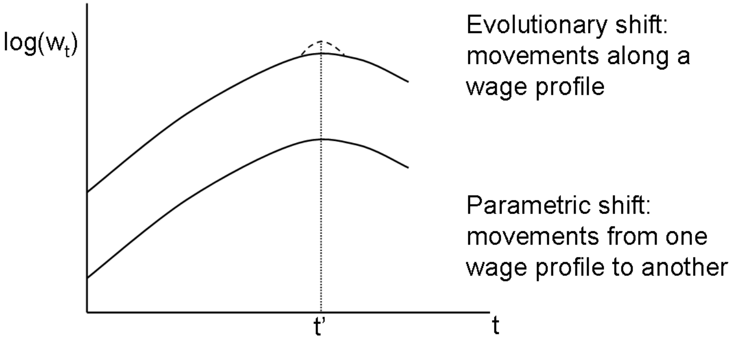
- Can show that $\delta > 0$: work more today to take advantage of temporarily higher wage

- In separable case:

$$\begin{aligned} l_t &= v'^{-1}(\lambda w_t / [\beta(1+r)]^t) \\ \Rightarrow & \frac{\partial l}{\partial w_t}|_\lambda = \frac{\lambda}{\beta(1+r)^t v''(l_t)} > 0 \end{aligned}$$

Dynamic Life Cycle Model: Three Types of Wage Changes

1. Evolutionary wage change: movements along profile
2. Parametric change: temporary tax cut
3. Profile shift: changing the wage rate in all periods
    - Equivalent to a permanent parametric change
    - Implicitly the elasticity that static studies estimate with unanticipated tax changes



Source: MaCurdy 1981

Frisch vs. Compensated vs. Uncompensated Elasticities

$$\begin{aligned} \text{Frisch elasticity} &\geq \text{Compensated static elasticity} \\ \text{Compensated static elasticity} &\geq \text{Uncompensated static elasticity} \end{aligned}$$

- Compensated static elasticity: changing wages in all periods but keeping utility constant
- Uncompensated static elasticity: changing wages in each period with no compensation
- First inequality is due to inter-temporal substitution:
  - When wage increases only in 1 period, substitute labor from other periods toward this period
  - When it increases in all periods, do not have this motive
- Second inequality is due to income effects (as in static model)

Frisch vs. Compensated vs. Uncompensated Elasticities

$$\begin{aligned} \text{Frisch elasticity} &\geq \text{Compensated static elasticity} \\ \text{Compensated static elasticity} &\geq \text{Uncompensated static elasticity} \end{aligned}$$

- Without income effects, all three elasticities are equal
- Otherwise inequalities are strict
- Difference in elasticities related to anticipated vs. unanticipated changes
  - Looney and Singhal (2007) exploit this logic to identify Frisch elasticity
- Frisch elasticity is of central interest for calibration of macro business cycle models

Structural Estimates: MaCurdy 1983 and Pencavel 2002

- MaCurdy (1983)
  - Structural estimate using panel data for men and within-person wage variation
  - Find both Frisch and compensated wage elasticity of around 0.15
  - But wage variation is not exogenous
- Pencavel (2002)
  - Instruments with trade balance interacted with schooling and age
  - Frisch elasticity: 0.2
  - Uncompensated wage elasticity: 0-0.2

Card Critique of ITLS models

- Critiques value of ITLS model
  - Fails to explain most variation in hours over lifecycle
  - Sheds little light on profile-shift elasticities that we care about
  - Difficult to identify key parameters
- Exemplifies structural vs. reduced-form divide in applied microeconomics
  - Tradeoff between credible identification and identification of structural parameters

Blundell, Duncan, and Meghir 1998

- Good combination of structural and reduced form methods on labor supply
- Argue against standard DD approach, where treatment/control groups are endogenously defined based on income
  - E.g., reduced tax rate may pull households into that tax group
  - Need group definitions that are stable over time
- Use birth cohort (decade) interacted with education (e.g. high school or more)

Blundell, Duncan, and Meghir 1998

- Construct group-level labor supply measures for women
- Measure how labor supply co-varies with wages rates net of taxes in the UK in 1980s
- Importantly, tax reforms during this period affected groups very differently
- Use consumption data as a control for permanent income
- Can therefore obtain a structurally interpretable (λ constant) estimate

TABLE IV

ELASTICITIES: GROUPING INSTRUMENTS: COHORT AND EDUCATION

	Wage	Compensated Wage	Other Income	Group Means: Hours	Group Means: Wage	Group Means: Income
No Children	0.140	0.140	0.000	32	2.97	88.63
	(0.075)	(0.088)	(0.041)			
Youngest Child 0–2	0.205	0.301	−0.185	20	3.36	129.69
	(0.128)	(0.144)	(0.104)			
Youngest Child 3–4	0.371	0.439	−0.173	18	3.10	143.64
	(0.150)	(0.159)	(0.139)			
Youngest Child 5–10	0.132	0.173	−0.102	21	2.86	151.13
	(0.117)	(0.127)	(0.109)			
Youngest Child 11 +	0.130	0.160	−0.063	25	2.83	147.31
	(0.107)	(0.117)	(0.084)			

Note: Asymptotic standard errors in parentheses.

Source: Blundell, Duncan, and Meghir 1998

Blundell, Duncan, and Meghir: Results

- Compensated wage elasticities: 0.15-0.3, depending on number of kids
- Virtually no income effects
- Identification assumption is common trends across cohort/ed groups
- However, reforms in 80s went in opposite directions at different times

  $\rightarrow$Secular trends cannot explain everything
- See Pencavel (1986) and Blundell and MaCurdy (1999) for additional ITLS estimates

Intertemporal Substitution: High Frequency Studies

- Recent literature focuses on groups such as cab drivers with highly flexible and well measured labor supply
- Camerer et al. 1997: examine how variation across days in wage rate for cab drivers (arising from variation in waiting times) correlates with hours worked
- Striking finding: strong negative effect
- Interpret this as "target earning" – strongly contradicts standard intertemporal labor supply model
- Would imply counterintuitive effects for temporary tax changes

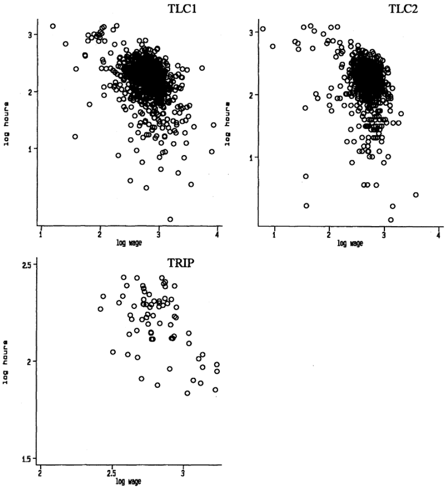


FIGURE I
Hours-Wage Relationships

Source: Camerer et al. 1997

TABLE II
OLS Log Hours Worked Equations

Sample	TRIP		TLC1		TLC2
Log hourly wage	−.411	−.186	−.501	−.618	−.355
	(.169)	(.129)	(.063)	(.051)	(.051)
High temperature	.000	−.000	.001	.002	−.021
	(.002)	(.002)	(.002)	(.002)	(.007)
Shift during week	−.057	−.047	−.004	.030	—
	(.019)	(.033)	(.035)	(.042)	
Rain	.002	.015	—	—	−.150
	(.035)	(.035)			(.062)
Night shift dummy	.048	−.049	−.127	−.294	−.253
	(.053)	(.049)	(.034)	(.047)	(.038)
Day shift dummy	—	—	.000	.053	—
			(.028)	(.045)	
Fixed effects	No	Yes	No	Yes	No
Adjusted R^2	.243	.484	.175	.318	.146
Sample size	70	65	1044	794	712
Number of drivers	13	8	484	234	712

Dependent variable is the log of hours worked. Standard errors are in parentheses and are corrected for the nonfixed effects estimates in coulmns 1 and 3 to account for the panel structure of the data. Explanatory variables are described in Appendix 1.

Source: Camerer et al. 1997

Farber 2005: Division Bias

- Argues that Camerer et al. evidence of target earning behavior is driven by econometric problems

- Camerer et al. regression specification:

$$h_{it} = \alpha + \beta e_{it} / h_{it} + \varepsilon_{it}$$

- Camerer et al. recognize this and try to instrument with average daily wage for each individual's wage

- But there may be a random component to hours at the group level (e.g., some days people just randomly report many hours on the job)

  $\rightarrow$ Spuriously find a negative association between average daily wage and average hours

Farber: Alternative Test

- Farber's alternative test for target earnings: hazard model

$$Quit = f(cum\_hours, cum\_inc)$$

- Result: main determinant of quitting is hours worked, NOT cumulative income

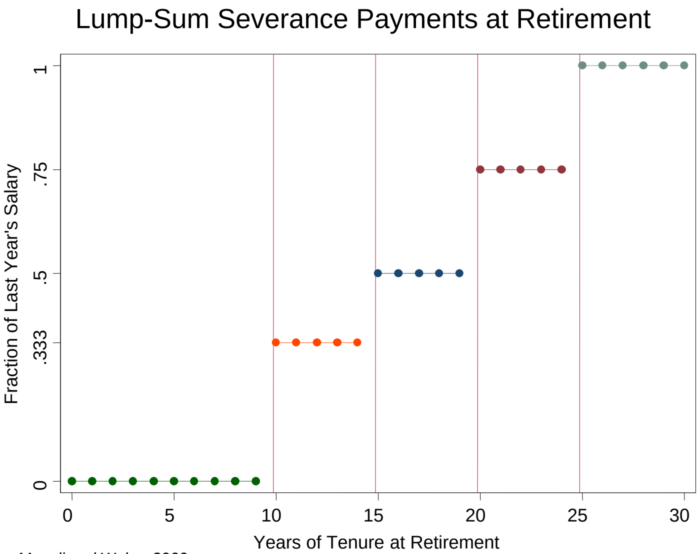
- Rejects target earning, but does not yield ITLS estimate

- Other studies find positive ITLS
  - Bicycle messengers (Fehr and Goette 2007 randomized experiment)
  - Stadium vendors (Oettinger 1999: vendors show up more to high attendance games)

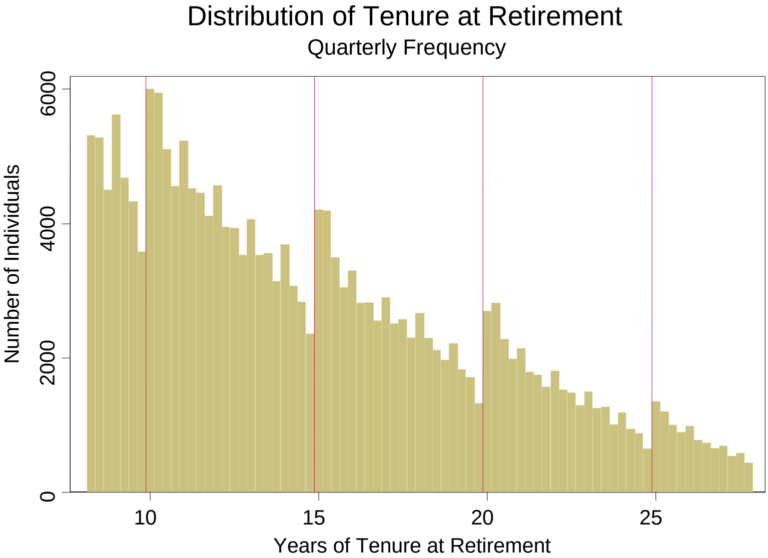
- But structural parameters estimated in these studies are not of direct interest to macro models or public finance because they are too high frequency

Manoli and Weber 2009

- Use variation in retirement benefits as a function of job tenure in Austria to estimate Frisch elasticity
- Question: how much do people delay retirement in order to get higher (anticipated) benefits?
- Dataset: administrative panel for full population of Austria, 1980-2005
- Rough estimate of Frisch elasticity: 0.2 at annual level



Source: Manoli and Weber 2009



Source: Manoli and Weber 2009

Taxable Income Elasticities

- Modern public finance literature focuses on taxable income elasticities instead of hours/participation elasticities
- Two main reasons
  1. Convenient sufficient statistic for all distortions created by income tax system (Feldstein 1999)
  2. Data availability: taxable income is precisely measured in tax return data
- Good overview of this literature: Saez et al. 2009

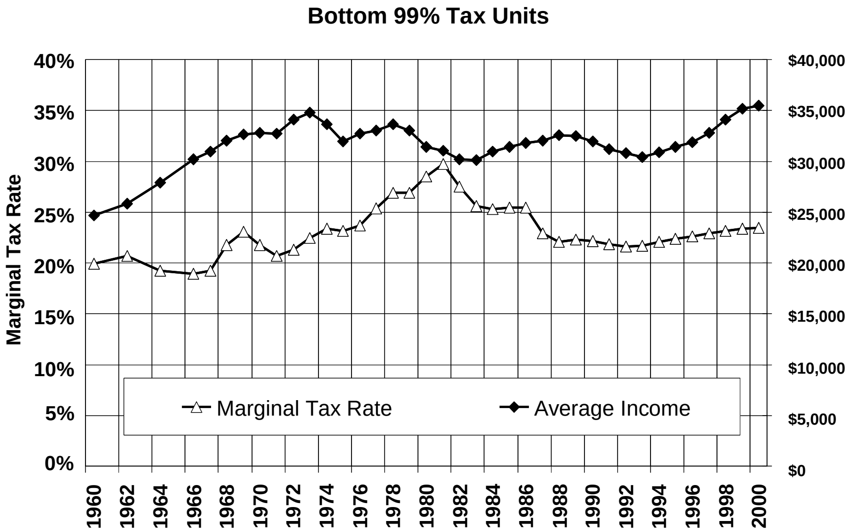
US Income Taxation: Trends

- The biggest changes in MTRs are at the top

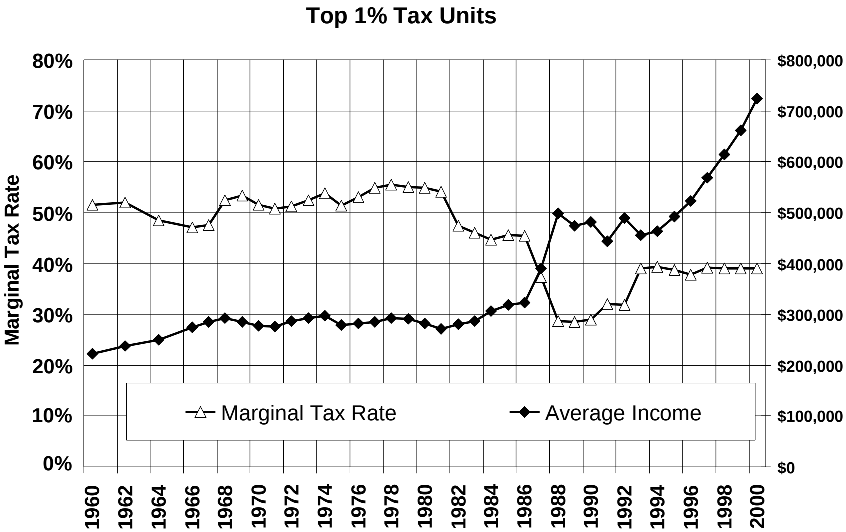
1. [Kennedy tax cuts]: 91% to 70% in '63-65
2. [Reagan I, ERTA 81]: 70% to 50% in '81-82
3. [Reagan II, TRA 86]: 50% to 28% in '86-88
4. [Bush I tax increase]: 28% to 31% in '91
5. [Clinton tax increase]: 31% to 39.6% in '93
6. [Bush Tax cuts]: 39.6% to 35% in '01-03

Saez 2004: Long-Run Evidence

- Compares top 1% relative to the bottom 99%
- Bottom 99% real income increases up to early 1970s and stagnates since then
- Top 1% increases slowly up to the early 1980s and then increases dramatically up to year 2000.
    - Corresponds to the decrease in MTRs
- Pattern exemplifies general theme of this literature: large responses for top earners, no response for rest of the population



Source: Saez 2004



Source: Saez 2004

Feldstein 1995

- First study of taxable income: Lindsey (1987) using cross-sections around 1981 reform
  - Limited data and serious econometric problems
- Feldstein (1995) estimates the effect of TRA86 on taxable income for top earners
- Constructs three income groups based on income in 1985
- Looks at how incomes and MTR evolve from 1985 to 1988 for individuals in each group using panel

TABLE 2

ESTIMATED ELASTICITIES OF TAXABLE INCOME WITH RESPECT TO NET-OF-TAX RATES

Taxpayer Groups Classified by 1985 Marginal Rate	Net of Tax Rate (1)	Adjusted Taxable Income (2)	Adjusted Taxable Income Plus Gross Loss (3)
	Percentage Changes, 1985–88		
1. Medium (22–38)	12.2	6.2	6.4
2. High (42–45)	25.6	21.0	20.3
3. Highest (49–50)	42.2	71.6	44.8
	Differences of Differences		
4. High minus medium	13.4	14.8	13.9
5. Highest minus high	16.6	50.6	24.5
6. Highest minus medium	30.0	65.4	38.4
	Implied Elasticity Estimates		
7. High minus medium		1.10	1.04
8. Highest minus high		3.05	1.48
9. Highest minus medium		2.14	1.25

NOTE.—The calculations in this table are based on observations for married taxpayers under age 65 who filed joint tax returns for 1985 and 1988 with no age exemption in 1988. Taxpayers who created a subchapter S corporation between 1985 and 1988 are eliminated from the sample.

Source: Feldstein 1995

Feldstein: Results

- Feldstein obtains very high elasticities (above 1) for top earners
- Implication: we were on the wrong side of the Laffer curve for the rich
  - Cutting tax rates would raise revenue

Feldstein: Econometric Criticisms

- DD can give very biased results when elasticity differ by groups
- Suppose that the middle class has a zero elasticity so that

$$\Delta \log(z^M) = 0$$

- Suppose high income individuals have an elasticity of e so that

$$\Delta \log(z^H) = e\Delta \log(1 - \tau^H)$$

- Suppose tax change for high is twice as large:

$$\Delta \log(1 - \tau^M) = 10\% \text{ and } \Delta \log(1 - \tau^H) = 20\%$$

- Estimated elasticity $\hat{e} = \frac{e \cdot 20\% - 0}{20\% - 10\%} = 2e$

Feldstein: Econometric Criticisms

- Sample size: results driven by very few observations (Slemrod 1996)
  - Auten-Carroll (1999) replicate results on larger Treasury dataset
  - Find a smaller elasticity: 0.65
- Different trends across income groups (Goolsbee 1998)
  - Triple difference that nets out differential prior trends yields elasticity <0.4 for top earners

Slemrod: Shifting vs. "Real" Responses

- Slemrod (1996) studies "anatomy" of the behavioral response underlying change in taxable income
- Shows that large part of increase is driven by shift between C corp income to S corp income
  - Looks like a supply side story but government is actually losing revenue at the corporate tax level
  - Shifting across tax bases not taken into account in Feldstein efficiency cost calculations

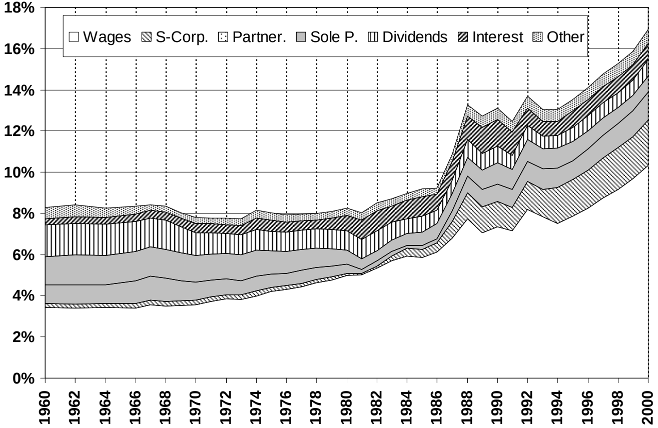


FIGURE 7

The Top 1% Income Share and Composition, 1960-2000

Source: Saez 2004

Goolsbee: Intertemporal Shifting

- Goolsbee (2000) hypothesizes that top earners' ability to retime income drives much of observed responses
- Analogous to identification of Frisch elasticity instead of compensated elasticitiy
- Regression specification:

$$TLI = \alpha + \beta_1 \log(1 - tax_t) + \beta_2 \log(1 - tax_{t+1})$$

- Long run effect is $\beta_1 + \beta_2$
- Uses ExecuComp data to study effects of the 1993 Clinton tax increase on executive pay

TABLE 4

RESPONSE OF TAXABLE INCOME FOR VARIOUS GROUPS

	PERMANENT INCOME GROUP				OPTIONS		SALARY AND BONUS
	All Executives >$275,000 (1)	$275,000–500,000 (2)	$500,000–1,000,000 (3)	>$1,000,000 (4)	No (5)	Yes (6)	(7)
$\ln(1 - \text{tax}_t)$	1.159	.394	.810	2.218	.290	1.289	.150
	(.119)	(.139)	(.178)	(.281)	(.311)	(.128)	(.073)
$\ln(1 - \text{tax}_{t+1})$	−.763	−.051	−.433	−1.663	−.181	−.853	−.060
	(.106)	(.132)	(.158)	(.240)	(.279)	(.115)	(.065)
$\ln(1 - \text{tax}_c) \times [I > 0]$	.282	...	.851	.140	.943	.175	.187
	(.140)		(.639)	(.189)	(.344)	(.153)	(.094)
ln(market value)	.610	.337	.559	.999	.518	.619	.289
	(.014)	(.015)	(.021)	(.033)	(.041)	(.015)	(.008)
Earnings/assets	.510	.311	.681	.823	.344	.542	.423
	(.056)	(.059)	(.089)	(.144)	(.129)	(.062)	(.035)
Time	.077	.068	.073	.061	.060	.079	.082
	(.008)	(.009)	(.012)	(.019)	(.020)	(.008)	(.005)
Year dummies	no	no	no	no	no	yes	no
Observations	16,477	5,918	5,680	4,879	2,122	14,330	18,628
R^2	.77	.41	.41	.58	.76	.77	.85

NOTE.—The sample in each regression pertains to 1991–95. The dependent variable is the log of taxable income. Cols. 1–4 look at executives with permanent income in the listed ranges. Cols. 5 and 6 look at executives divided by whether or not they received any options from 1992 to 1995 . Col. 7 looks at taxable income without options exercised. All regressions include individual fixed effects. Standard errors are in parentheses.

Source: Goolsbee 2000

Goolsbee 2000

- Most affected groups (income>$250K) had a surge in income in 1992 (when reform was announced) relative to 1991 followed by a sharp drop in 1993
- Simple DD estimate would find a large effect here, but it would be picking up pure re-timing
- Concludes that long run effect is 20x smaller than substitution effect
- Effects driven almost entirely by retiming exercise of options
- Long run elasticity <0.4 and likely close to 0

Gruber and Saez 2002

- First study to examine taxable income responses for general population, not just top earners
- Use data from 1979-1991 on all tax changes available rather than a single reform
- Simulated instruments methodology
  - Step 1: Simulate tax rates based on period t income and characteristics
$$\begin{aligned} MTR^P_{t+3} &= f_{t+3}(y_t, X_t) \\ MTR_{t+3} &= f_{t+3}(y_{t+3}, X_{t+3}) \end{aligned}$$
  - Step 2 [first stage]: Regress $\log(1 - MTR_{t+3}) - \log(1 - MTR_t)$ on $\log(1 - MTR^P_{t+3}) - \log(1 - MTR_t)$
  - Step 3 [second stage]: Regress $\Delta \log TI$ on predicted value from first stage
- Isolates changes in laws (f_t) as the only source of variation in tax rates

Table 4
Basic elasticity results[a]

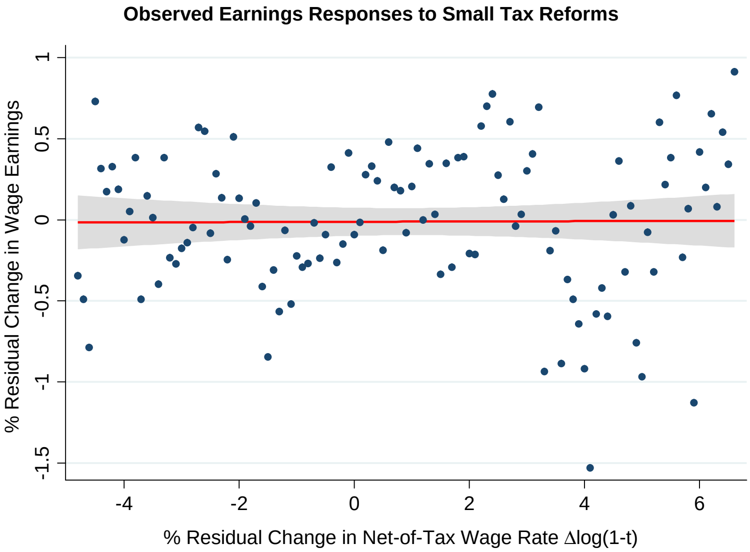
Income controls	None		Log income		Log income 10-piece spline	
	Broad income (1)	Taxable income (2)	Broad income (3)	Taxable income (4)	Broad income (5)	Taxable income (6)
Elasticity	−0.300	−0.462	0.170	0.611	0.120	0.400
	(0.120)	(0.194)	(0.106)	(0.144)	(0.106)	(0.144)
Dummy for marrieds	−0.008	−0.062	0.045	0.049	0.050	0.055
	(0.010)	(0.018)	(0.014)	(0.023)	(0.012)	(0.021)
Dummy for singles	−0.037	−0.053	−0.034	−0.032	−0.036	−0.027
	(0.012)	(0.019)	(0.013)	(0.022)	(0.013)	(0.021)
Log(income) control			−0.083	−0.167		
			(0.015)	(0.021)		

Source: Gruber and Saez 2002

Gruber and Saez: Results

- Find an elasticity of roughly 0.3-0.4 with splines
  - But this is very fragile (Giertz 2008)
  - Sensitive to exclusion of low incomes (min income threshold)
  - Sensitive to controls for mean reversion
- Subsequent studies find smaller elasticities using data from other countries

Evidence from Danish Tax Reforms



Source: Chetty et al. 2009

Imbens et al. 2001: Income Effects

- Estimate income effects using lottery winnings
- Survey responses matched with administrative data on earnings from Social Security Administration
- Divide sample into three subgroups:
  1. Losers [$N = 259$]: "season ticket holders" who won \$100-\$5K
  2. Winners [$N = 237$]: anyone who won prizes of \$22K to \$9.7 mil
  3. Big Winners [$N = 43$]: winners of prizes >\$2 mil total (\$100K/yr)
- Estimate marginal propensity to earn out of unearned income of $d[wl]/dy = -0.1$

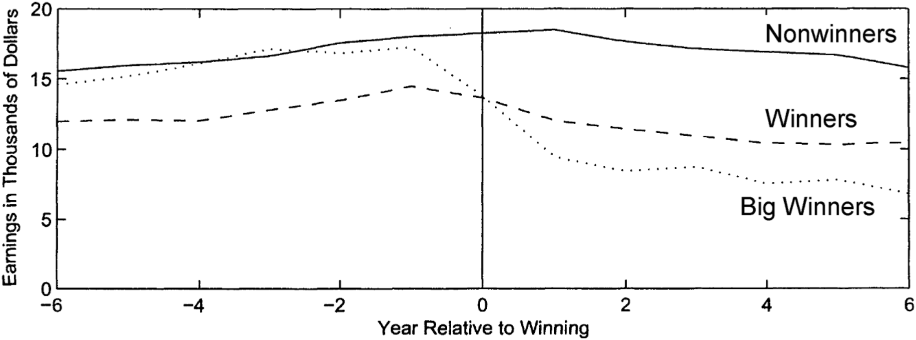


FIGURE 1. AVERAGE EARNINGS FOR NONWINNERS, WINNERS, AND BIG WINNERS

Source: Imbens, Rubin, and Sacerdote 2001

Taxable Income Literature: Summary

- Large responses for the rich, mostly intertemporal substitution and shifting
- Responses among lower incomes small at least in short run
  - Perhaps not surprising if they have little flexibility to change earnings
- Pattern confirmed in many settings (e.g. Kopczuk 2009 - Polish flat tax reform)
- But many methdological problems remain to be resolved
  - Econometric issues: mean reversion, appropriate counterfactuals
  - Which elasticity is being identified?

Macro Evidence

- Macroeconomists estimate/calibrate elasticities by examining long-term trends/cross-country comparisons
- Identification more questionable but estimates perhaps more relevant to long-run policy questions of interest
- Use aggregate hours data and aggregate measures of taxes (average tax rates)
- But highly influential in calibration of macroeconomic models
  - Macro models require high elasticities to fit both business cycle and cross-country data

Prescott 2004

- Uses data on hours worked by country in 1970 and 1995 for 7 OECD countries
- Technique to identify elasticity: calibration of GE model
- Rough intuition: posit a labor supply model, e.g.

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

- Finds that elasticity of $\varepsilon = 1.2$ best matches time series and cross-sectional patterns
- Note that this is analogous to a regression without controls for other variables
- Results verified in subsequent calibrations by Rogerson and others using more data

Table 2

Actual and Predicted Labor Supply

In Selected Countries in 1993–96 and 1970–74

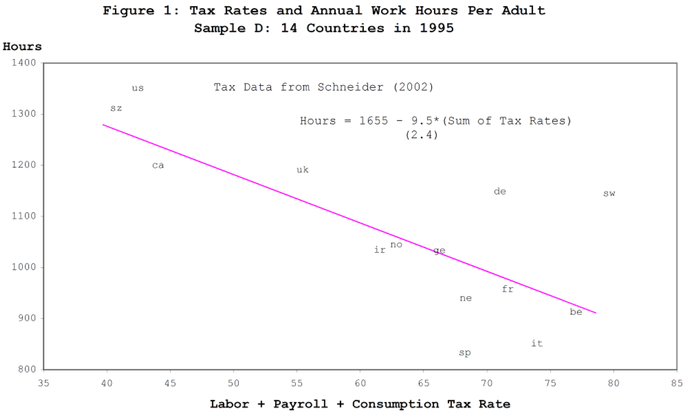
Period	Country	Labor Supply*		Differences (Predicted Less Actual)	Prediction Factors	
		Actual	Predicted		Tax Rate τ	Consumption/ Output (c/y)
1993–96	Germany	19.3	19.5	.2	.59	.74
	France	17.5	19.5	2.0	.59	.74
	Italy	16.5	18.8	2.3	.64	.69
	Canada	22.9	21.3	–1.6	.52	.77
	United Kingdom	22.8	22.8	0	.44	.83
	Japan	27.0	29.0	2.0	.37	.68
	United States	25.9	24.6	–1.3	.40	.81
1970–74	Germany	24.6	24.6	0	.52	.66
	France	24.4	25.4	1.0	.49	.66
	Italy	19.2	28.3	9.1	.41	.66
	Canada	22.2	25.6	3.4	.44	.72
	United Kingdom	25.9	24.0	–1.9	.45	.77
	Japan	29.8	35.8	6.0	.25	.60
	United States	23.5	26.4	2.9	.40	.74

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

Source: Prescott (2004)

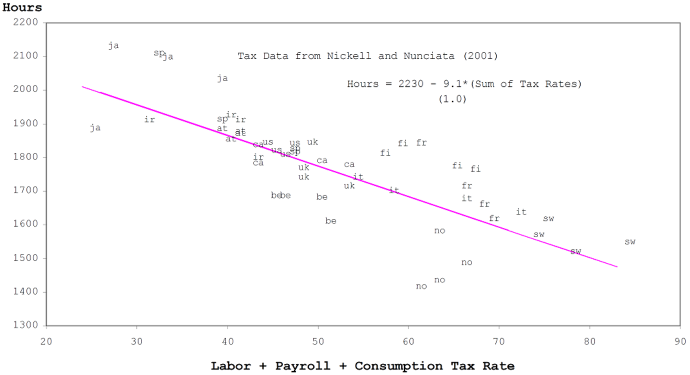
Davis and Henrekson 2005

- Run regressions of hours worked on tax variables with various controls
- Some panel evidence, but primarily cross-sectional
- Separate intensive and extensive margin responses



Source: Davis and Henrekson 2005

Figure 2: Tax Rates and Annual Hours Per Employed Person
Sample A: 13 Countries with Data for 1977, 1983, 1990 and 1995



Source: Davis and Henrekson 2005

Reconciling Micro and Macro Estimates

- Recent interest in reconciling micro and macro elasticity estimates
- Three potential explanations

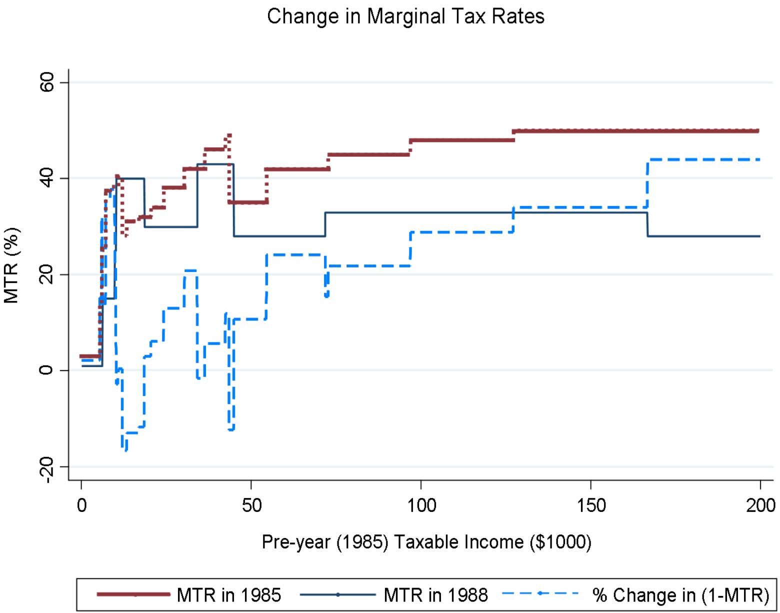
1. Statistical Bias: regulations, culture differs in countries with higher tax rates [Alesina, Glaeser, Sacerdote 2005]
2. Extensive vs Intensive margin [Rogerson and Wallenius 2008]

$$\begin{aligned} L &= Nh \\ \frac{d\log L}{d(1-\tau)} &= \frac{d\log N}{d(1-\tau)} + \frac{d\log h}{d(1-\tau)} > \frac{d\log h}{d(1-\tau)} \end{aligned}$$

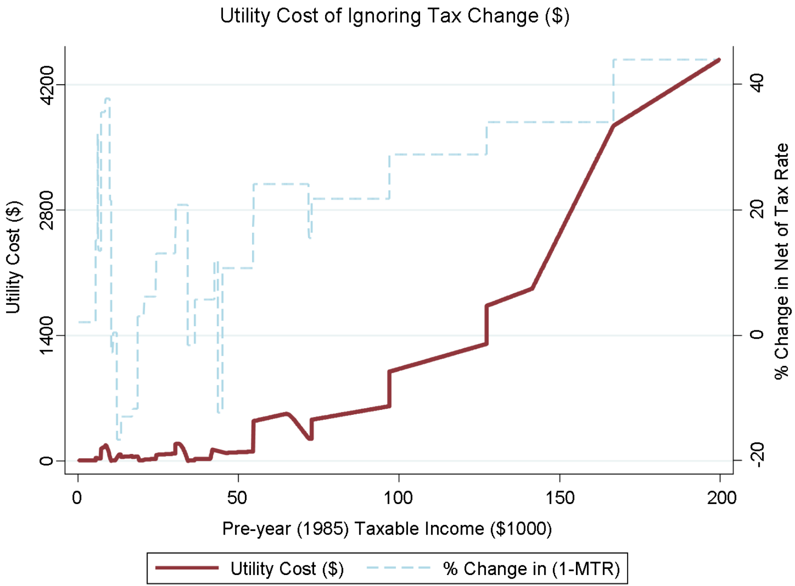
3. Optimization frictions: short run vs. long run [Chetty 2009]

Optimization Frictions

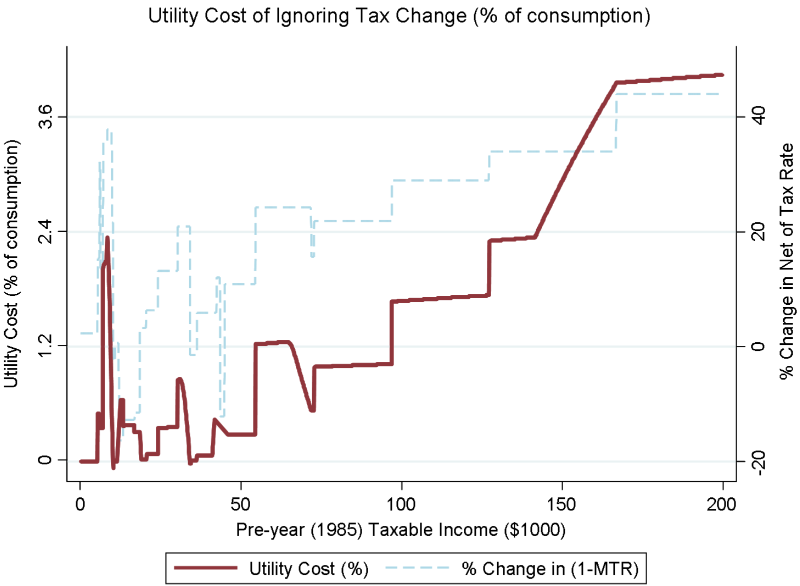
- Many frictions may cause agents to deviate from unconstrained optimum, e.g. adjustment costs or inattention
- These frictions may attenuate short-run responses to tax reforms
- Chetty (2009) asks two questions
  1. Can frictions quantitatively explain micro-macro puzzle and other puzzles in labor supply literature?
  2. Given frictions, what can we say about the "structural" elasticity?
     - Structural elasticity controls long run responses (e.g. Europe vs US)
- Example: calculate utility loss from **ignoring** tax changes under neoclassical model with $\varepsilon = 0.5$



Source: Chetty 2009



Source: Chetty 2009



Source: Chetty 2009

Setup

- Consider a static demand model; results hold in dynamic model
- N individuals with quasilinear utility over two goods:

$$u_i(x, y) = y + a_i \frac{x^{1-1/\varepsilon}}{1 - 1/\varepsilon}$$

- Agent i's optimal demand for good x:

$$\begin{aligned} x_i^*(p) &= (\frac{a_i}{p})^{\varepsilon} \\ \Rightarrow \log x_i^*(p) &= \alpha - \varepsilon \log p + v_i \end{aligned}$$

where $v_i = \alpha_i - \alpha$ denotes i's deviation from mean demand

- Under orthogonality condition $\mathbb{E} v_i | p = 0$,

$$\varepsilon = \frac{\mathbb{E} \log x_1^* - \mathbb{E} \log x_0^*}{\log p_1 - \log p_0}$$

$\rightarrow$Observed response to price increase (p_0 to p_1) identifies ε.

Optimization Frictions: Examples

- Agent pays adjustment cost k_i to change consumption
- Demand set optimally at initial price p_0
- Let $x(p)$ denote observed demand at price p
- Define **observed** elasticity estimated from price increase as

$$\widehat{\varepsilon} = \frac{\mathbb{E}\log x_1 - \mathbb{E}\log x_0}{\log p_1 - \log p_0}$$

- Observed elasticity confounds **structural** elasticity ε with adjustment cost distribution:

$$\widehat{\varepsilon} = P(\Delta u_i > k_i)\varepsilon$$

- Behavioral example: price misperception $\widetilde{p}(p)$

$$\widehat{\varepsilon} = \varepsilon \frac{\mathbb{E}\log \widetilde{p}(p_1) - \mathbb{E}\log \widetilde{p}(p_0)}{\log p_1 - \log p_0}$$

Optimization Frictions

- Define agent i's "optimization error" as difference between observed and optimal demand:

$$\phi_i = \log x_i(p) - \log x_i^*(p)$$

- Observed demand can be written as

$$\log x_i(p) = \alpha - \varepsilon \log p + \nu_i + \phi_i$$

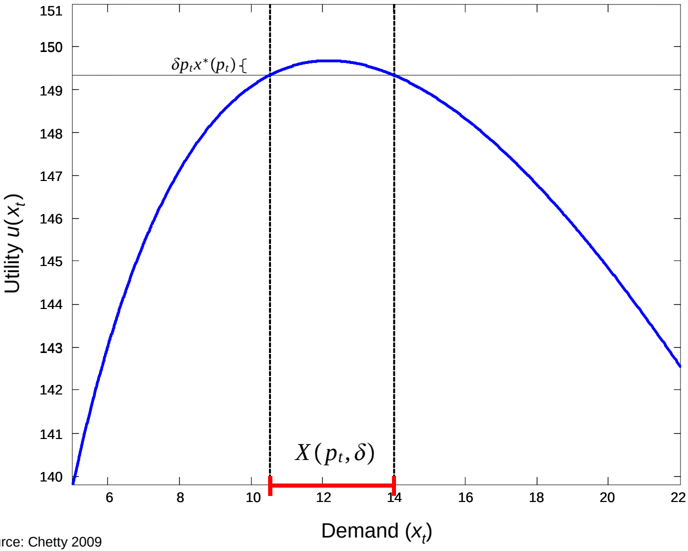
- Difference between optimization error (ϕ_i) and preference heterogeneity error (v_i): do not know distribution of ϕ_i
  - Want to remain agnostic about $\mathbb{E}\phi_i|p$

- Without restrictions on ϕ_i, observed response does not identify ε

Optimization Frictions

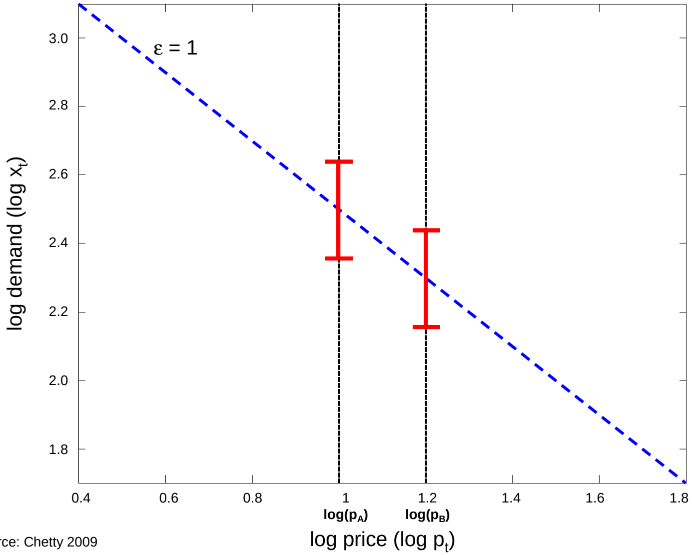
- Restriction on ϕ: Utility loss from deviations is less than exogenous threshold δ:
$$U(x_i^*) - U(x_i^* + \phi_i) < \delta p x_i^*$$
- Frictions make agents deviate from optimal behavior as long as utility costs are not too large
- This restriction generates a class of models around nominal model
  - Includes adjustment cost models, inattention, etc.
- A δ class of models maps price to a choice **set** $X(p, \delta)$ instead of a single point $x^*(p)$

Construction of Choice Set



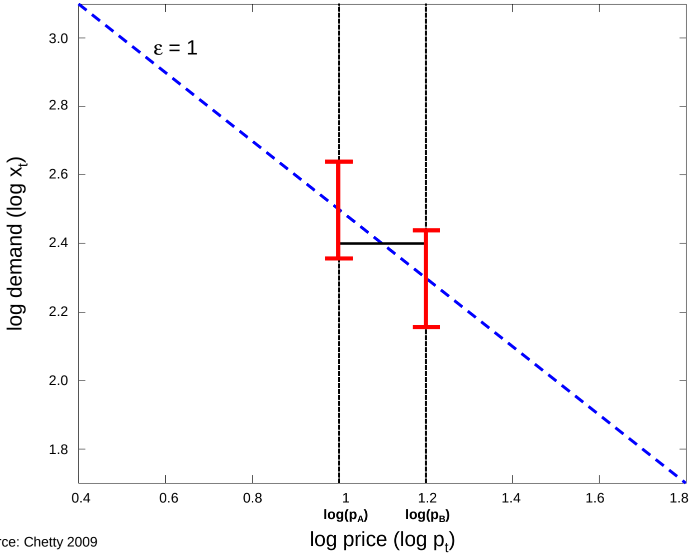
Source: Chetty 2009

Identification with Optimization Frictions



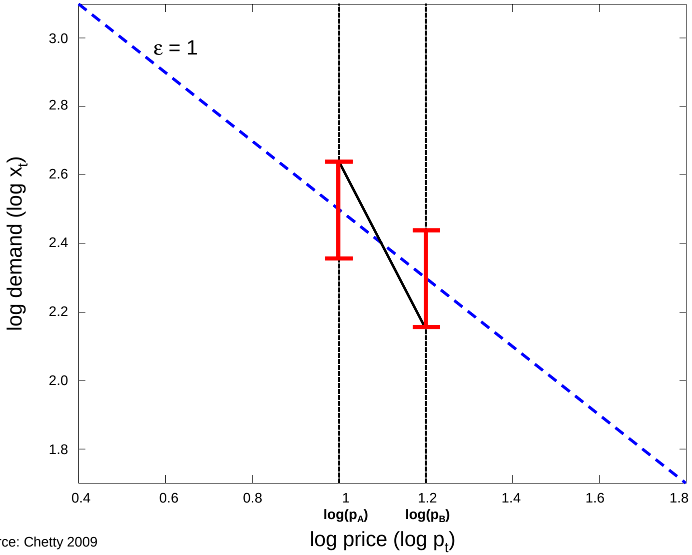
Source: Chetty 2009

Identification with Optimization Frictions



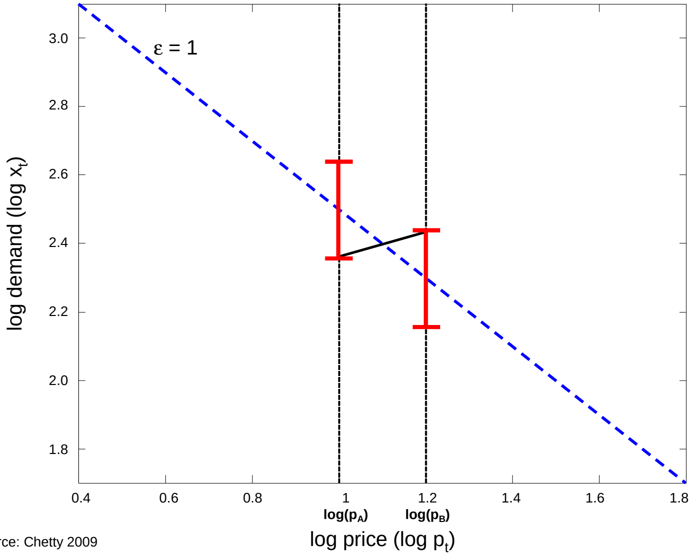
Source: Chetty 2009

Identification with Optimization Frictions



Source: Chetty 2009

Identification with Optimization Frictions



Source: Chetty 2009

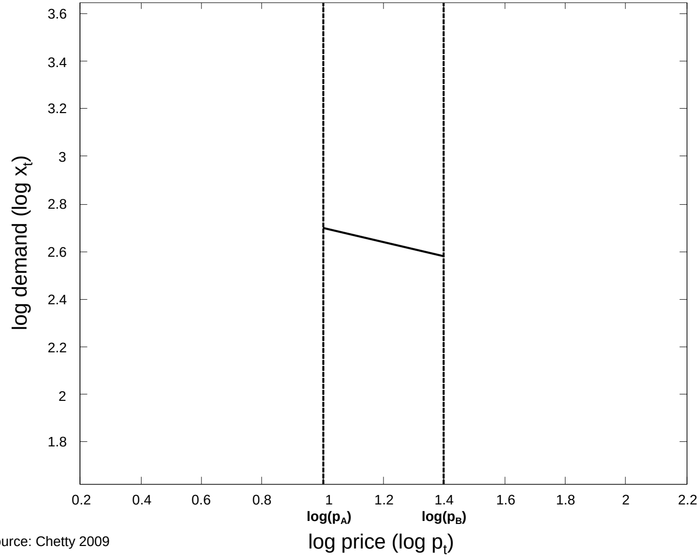
Bounds on Elasticities

- Multiple observed elasticities $\widehat{\varepsilon}$ can be generated by a model with a given structural elasticity when $\delta > 0$

- Conversely, multiple structural elasticities consistent with an observed $\widehat{\varepsilon}$

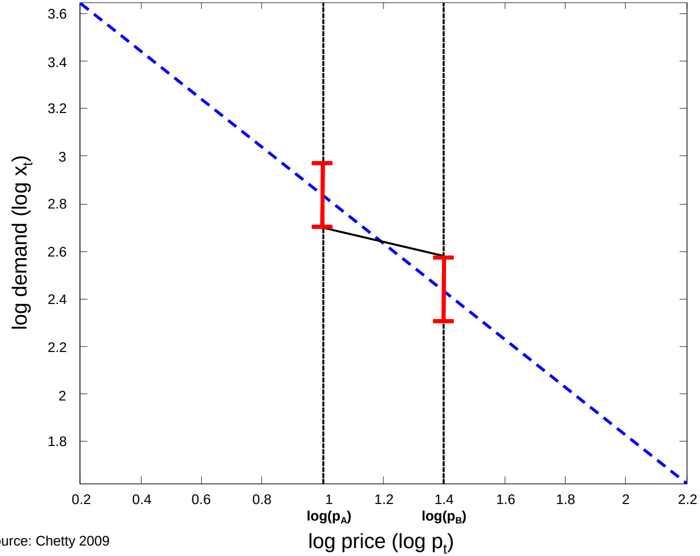
- Objective: derive bounds $(\varepsilon_L, \varepsilon_U)$ on smallest and largest structural elasticities consistent with $\widehat{\varepsilon}$

Calculation of Bounds on Structural Elasticity

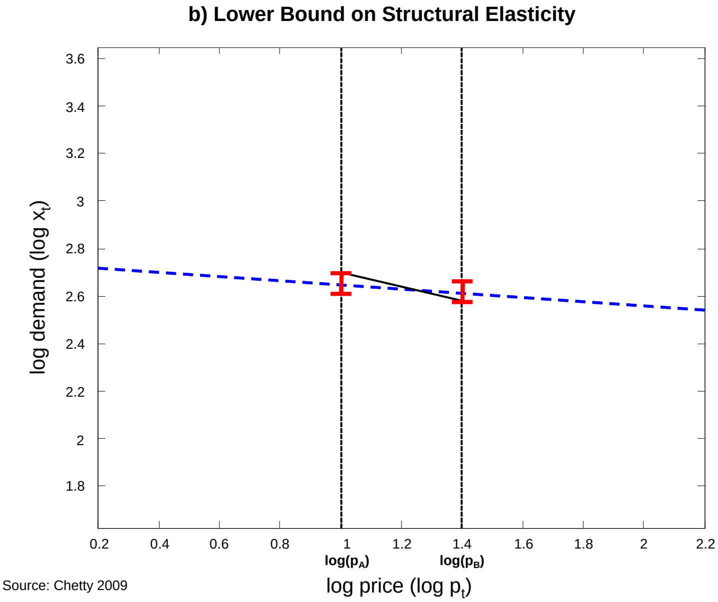


Source: Chetty 2009

a) Upper Bound on Structural Elasticity



Source: Chetty 2009



b) Lower Bound on Structural Elasticity
log demand (log x_t)
3.6
3.4
3.2
3
2.8
2.6
2.4
2.2
2
1.8
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
2.2
log(p_A)
log(p_B)
log price (log p_t)
Source: Chetty 2009

Bounds on Elasticity with Optimization Frictions

- For small δ, the range of structural elasticities consistent with an observed elasticity $\widehat{\varepsilon}$ in a δ class of models is approximately

$$[\widehat{\varepsilon} + \frac{4\delta}{(\Delta \log p)^2}(1-\rho), \widehat{\varepsilon} + \frac{4\delta}{(\Delta \log p)^2}(1+\rho)]$$

$$\text{where } \rho = (1 + \frac{1}{2}\frac{\widehat{\varepsilon}}{\delta}(\Delta \log p)^2)^{1/2}$$

- Maps an observed elasticity $\hat{\varepsilon}$, size of price change $\Delta \log p$, and degree of optimization frictions δ to bounds on ε.

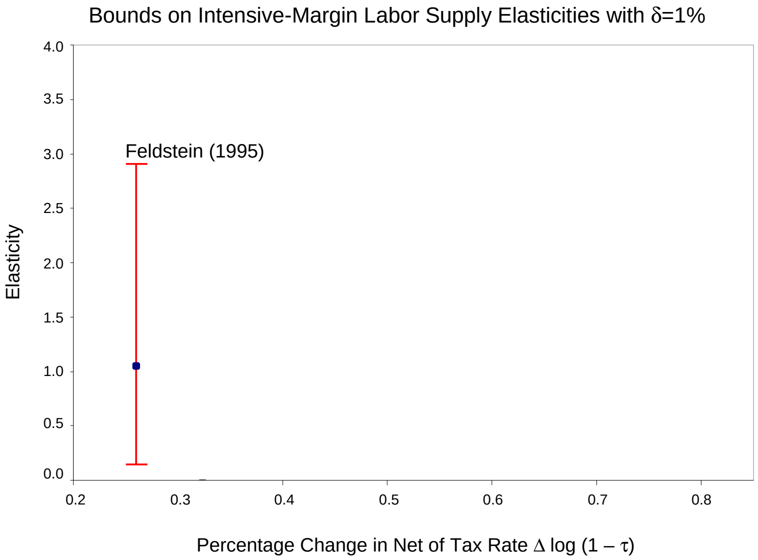
- Bounds shrink with $(\Delta \log p)^2$

Application to Taxation and Labor Supply

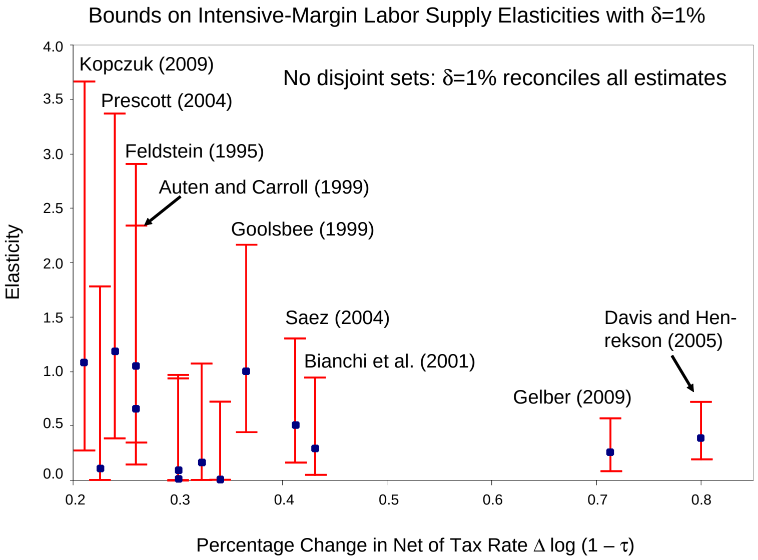
- Calculate bounds on taxable income elasticity using intensive-margin estimates from 20 recent studies
- Assume $\delta = 1\%$
- Ignore statistical imprecision for simplicity here
  - Text shows bounds for 95% confidence interval

Bounds on Intensive-Margin Labor Supply Elasticities with δ=1%

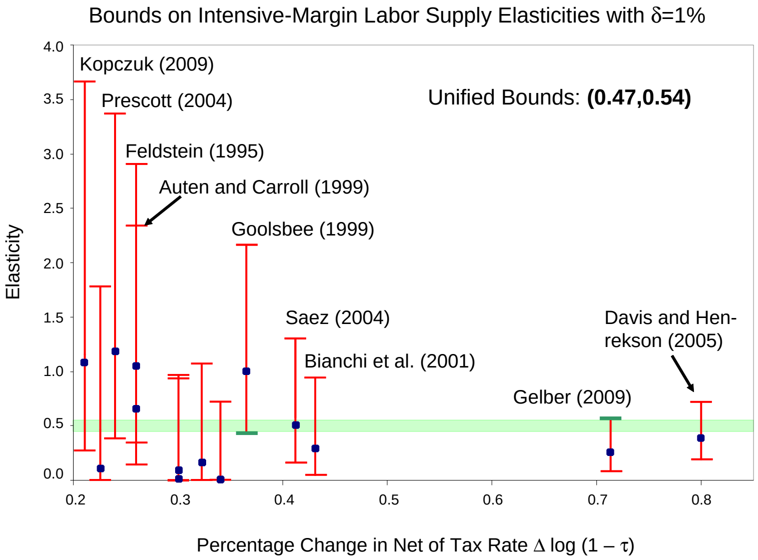
	Study (1)	$\hat{\varepsilon}$ (2)	$\Delta\log(1-\tau)$ (3)	ε_U (4)	ε_L (5)
Hours Elasticities	1. MaCurdy (1981)	0.15	0.12	0.00	5.63
	2. Eissa and Hoynes (1998)	0.14	0.11	0.00	6.56
	3. Blundell, Duncan, and Meghir (1998)	0.14	0.19	0.01	2.54
	4. Ziliak and Kniesner (1999)	0.15	0.32	0.02	1.05
	5. Bianchi, Gudmundson, and Zoega (2001)	0.29	0.43	0.09	0.91
Taxable Income Elasticities	6. Gruber and Saez (2002)	0.14	0.13	0.00	5.02
	7. Saez (2004)	0.09	0.23	0.00	1.75
	8. Chetty et al. (2009)	0.03	0.30	0.00	0.95
	9. Chetty et al. (2009)	0.00	0.10	0.00	8.00
	10. Gelber (2009)	0.25	0.71	0.12	0.54
	11. Kleven and Schultz (2009)	0.01	0.30	0.00	0.91
	12. Saez (2009)	0.00	0.34	0.00	0.70
Top Income Elasticities	13. Feldstein (1995)	1.04	0.26	0.37	2.89
	14. Auten and Carroll (1999)	0.66	0.26	0.19	2.32
	15. Goolsbee (1999)	1.00	0.37	0.47	2.14
	16. Saez (2004)	0.50	0.41	0.20	1.28
	17. Kopczuk (2009)	1.07	0.21	0.31	3.64
Macro/Trend -Based	18. Prescott (2004)	1.18	0.24	0.42	3.34
	19. Davis and Henrekson (2005)	0.39	0.80	0.23	0.69
	20. Blau and Kahn (2007)	0.56	0.16	0.07	4.33



Source: Chetty 2009



Source: Chetty 2009



Source: Chetty 2009

Information and Salience in Income Taxation

- Recent evidence indicates that one important “friction” is information/salience.
- Confusion between average and marginal tax rates: de Bartolome (1996), Liebman and Zeckhauser (2004)
- Evidence that information affects behavioral responses to income taxes: Chetty and Saez (2008)

Chetty and Saez 2009: Experimental Design

- 119 H&R Block offices in Chicago metro area; 43,000 EITC clients
- 1,461 tax professionals implemented experiment
- Tax Season 2007: Jan. 1 to April 15, 2007
- EITC clients randomly assigned to control or treatment group
- Control group: standard tax preparation procedure
  - Only mentions the EITC amount, with no info on EITC structure

Explaining EIC: 4 steps

Single With Two or More Children

The EIC (Earned Income Credit) is a tax refund that gives families as much as $4,500 per year.

We want to explain how the **EIC** works to help you decide how much to work and earn this year.

In 2006, you made $ 10,000 → you are getting an **EIC** of $ 4,000 in your refund.

- Your earnings this year (in 2007) will determine the size of your **EIC** refund next year
- The **EIC** has 3 ranges: 1) Increasing, 2) Peak, 3) Decreasing

1. Fill in earnings, EIC amount

2. Explain and dot graph

4. Take-home Message

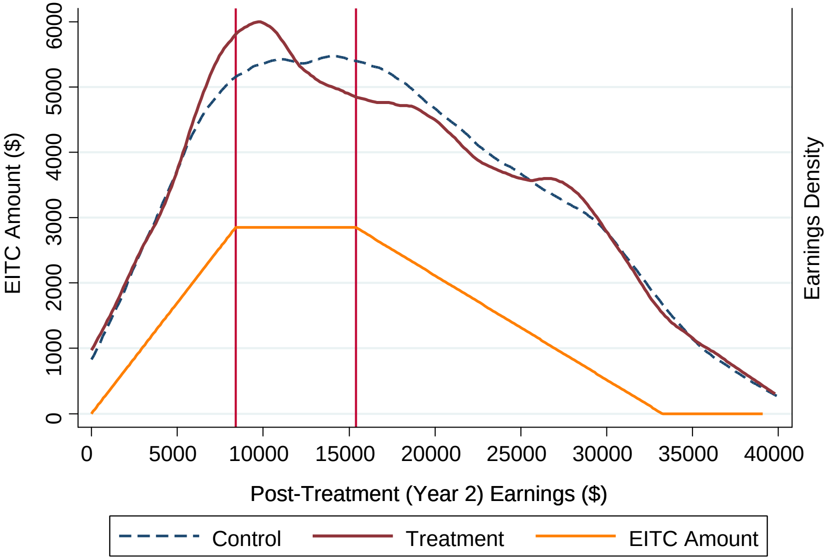
You are in the increasing range of the **EIC**. *Think about it like this:*

- (increasing) Suppose you earn $10 an hour, then you are really making $14.00 an hour.
- (peak) Your earnings are maxing-out the **EIC** amount
- (decreasing) If you earn $10 more, your **EIC** is reduced by $2.10

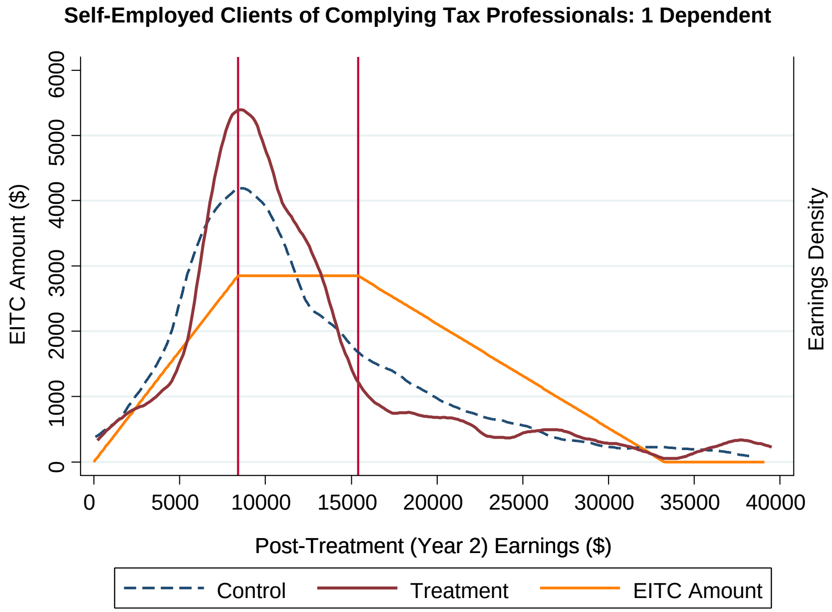
3. Table

EIC Range	If you earn between	EIC refund will be	If you earn $10 more, the EIC...
Increasing	$0-$11,790	$0 up to $4,716	Increases by $4
Peak	$11,790-$15,390	$4,716	Stays the same
Decreasing	$15,390-$37,780	$4,716 down to $0	Decreases by $2.10

Year 2 Earnings Distributions: 1 Dep., Clients of Complying Tax Preparers

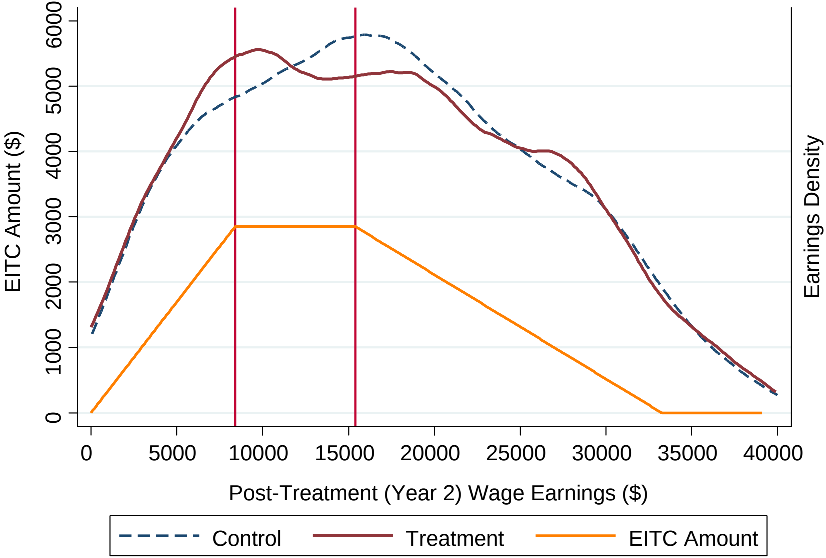


Source: Chetty and Saez 2009



Source: Chetty and Saez 2009

Year 2 Wage Earnings Distributions: Complying Tax Preparers, 1 Dependent



Source: Chetty and Saez 2009

Calibration: Magnitude of Information Effects

- How big is the behavioral response to the information relative to effects of conventional policy instruments?
- Existing literature implies intensive margin elasticity of earnings w.r.t. 1-MTR of at most $\varepsilon = 0.25$
- Complying tax pros increase treated clients' EITC by \$58
    - EITC expansion of 33 percent would generate same response

Labor Supply Elasticities: Implications for Preferences

- Labor supply elasticities central for tax policy because they determine efficiency costs

- But optimal income tax policy also depends on benefits of redistribution (curvature of utility fn.)

$$u(c) - \psi(l)$$

- Curvature of $u(c)$: $\gamma = \frac{-u_{cc}}{u_c} c$ determines how much more low income individuals value \$1 relative to higher income individuals

- Risk aversion parameter γ also central for social insurance literature and macro models

- Evidence on labor supply elasticities also contains information about γ (King, Plosser, Rebelo 1988; Basu and Kimball 2002; Chetty 2006)

Chetty 2006

- Suppose marginal utility of consumption declines quickly, i.e. γ large
- Then as wages rise, individuals should quickly become sated with goods
- Therefore, they should opt to consume much more leisure when wages rise
- But this would imply $\varepsilon_{l,w} << 0$
  - Ex: if marginal utility of consumption drops to zero, agent *reduces* labor supply 1-1 as wage rises
- But we know that increases in wages do not cause sharp reductions in labor supply ($\varepsilon_{l,w} > -0.1$)
- Places an upper bound on size of γ

Chetty: Formula for Risk Aversion

- Let $y =$ unearned inc, $w =$ wage, $l =$ labor supply and $u(c, l) =$ utility

- At an interior optimum, l must satisfy

$$wu_c(y + wl, l) = -u_l(y + wl, l) \tag{1}$$

- Work until point where marginal utility of an additional dollar is offset by marginal disutility of work required to earn that dollar

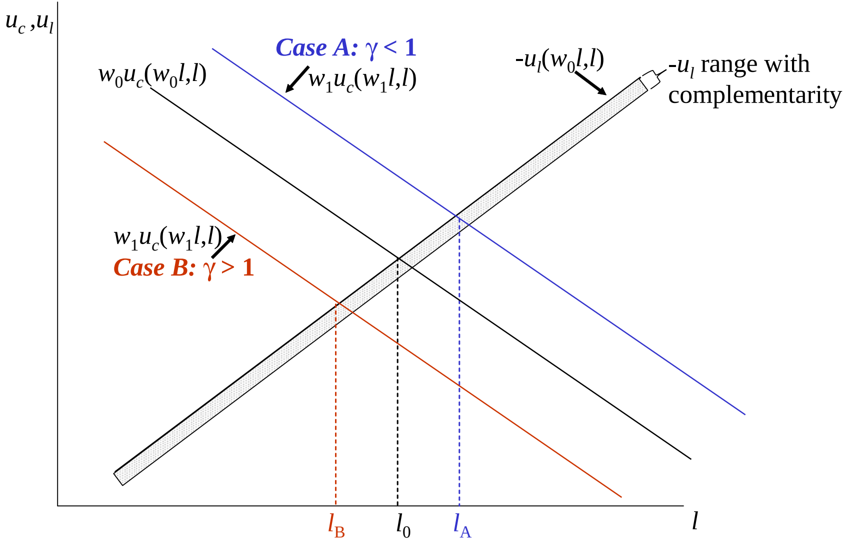
- Comparative statics of this condition implies (if $u_{cl} = 0$):

$$\gamma = -(1 + \frac{wl}{y})\frac{\varepsilon_{l,y}}{\varepsilon_{l^c,w}}$$

- Risk aversion directly related to ratio of income effect to substitution effect

Chetty 2006: Graphical Intuition

- Assume $y = 0$. At initial wage w_0, agent works l_0 hours
- Consider effect of increasing w by 1% to w_1
  - Shifts wu_c curve up by 1% (substitution effect)
  - Shifts wu_c curve down by $\frac{\partial \log u_c}{\partial \log w} = \gamma\%$ because γ is elasticity of MU w.r.t. c (income effect)
- Therefore, $\gamma < 1 \Longleftrightarrow \varepsilon_{l,w} > 0$
- If $u_{cl} \neq 0$, then $-u_l$ curve shifts when w changes
- But the shift is $-u_l$ relatively small, so change in l can still be used to get a bound on γ



Source: Chetty 2006

Labor Supply Elasticities and Implied Coefficients of Relative Risk Aversion

Study (1)	Sample (2)	Identification (3)	Income Elasticity (4)	Compensated Wage Elasticity (5)	γ Additive (6)	γ $\Delta c/c$=0.15 (7)
A. Hours						
MaCurdy (1981)	Married Men	Panel	-0.020	0.130	0.46	0.60
Blundell and MaCurdy (1999)	Men	Various	-0.120	0.567	0.63	0.82
MaCurdy, Green, Paarsch (1990)	Married Men	Cross Section	-0.010	0.035	1.47	1.81
Eissa and Hoynes (1998)	Married Men, Inc < 30K	EITC Expansions	-0.030	0.192	0.88	1.08
	Married Women, Inc < 30K	EITC Expansions	-0.040	0.088	0.64	1.34
Friedberg (2000)	Older Men (63-71)	Soc. Sec. Earnings Test	-0.297	0.545	0.93	1.46
Blundell, Duncan, Meghir (1998)	Women, UK	Tax Reforms	-0.185	0.301	0.93	1.66
Average					0.69	0.94
B. Participation						
Eissa and Hoynes (1998)	Married Men, Inc < 30K	EITC Expansions	-0.008	0.033	0.44	0.48
	Married Women, Inc < 30K	EITC Expansions	-0.038	0.288	0.15	0.30
Average					0.29	0.39
C. Earned Income						
Imbens, Rubin, Sacerdote (2001)	Lottery Players in MA	Lottery Winnings	-0.110			
Feldstein (1995)	Married, Inc > 30K	TRA 1986		1.040	0.32	0.41
Auten and Carroll (1997)	Single and Married, Inc>15K	TRA 1986		0.660	0.50	0.65
Average					0.41	0.53
D. Macroeconomic/Trend Evidence						
Blau and Kahn (2005)	Women	Cohort Trends	-0.278	0.646	0.60	1.29
Davis and Henrekson (2004)	Europe/US aggregate stats	Cross-Section of countries	-0.251	0.432	1.74	2.25
Prescott (2004)	Europe/US aggregate stats	Cross-Country time series	-0.222	0.375	1.78	2.30
Average					1.37	1.95
Overall Average					**0.71**	**0.97**

Source: Chetty 2006

Chetty 2006: Results

- Labor supply evidence justifies use of $u(c) = \log c$
- Formula $\gamma = -(1 + \frac{wl}{y})\frac{\varepsilon_{l,y}}{\varepsilon_{l^c,w}}$ useful in tax, insurance, and other applications

Public Economics Lectures
Part 6: Social Insurance

Raj Chetty and Gregory A. Bruich

Harvard University

Fall 2009

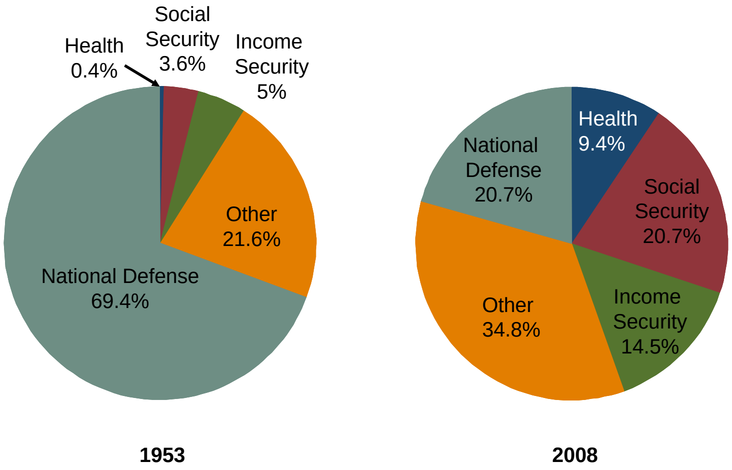
Outline

1. Motivations for Social Insurance
2. Unemployment Insurance
3. Workers' Compensation
4. Disability Insurance
5. Health Insurance

Definition of Social Insurance

- Transfers based on **events** such as unemployment, disability, or age
- Contrasts with welfare: **means-tested** transfers
- SI is the biggest and most rapidly growing part of government expenditure today

Growth of Social Insurance in the U.S.

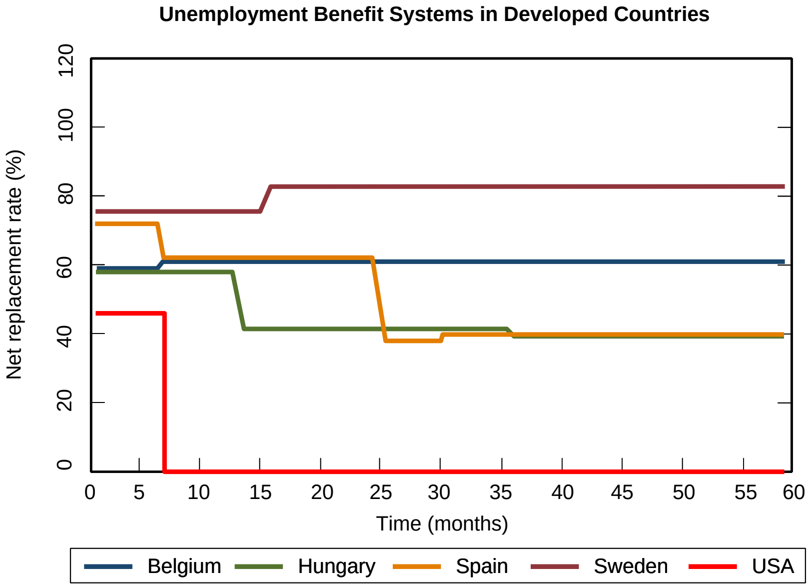


Source: Office of Management and Budget, historical tables, government outlays by function

Social Insurance Spending, 1996

	% of GDP	% of Central Government Expenditures	% of Total Government Expenditures
Sweden	32.47%	86.60%	49.58%
Germany	28.05%	82.91%	49.44%
Mexico	1.36%	8.82%	6.39%
Columbia	6.61%	43.33%	N/A
United Kingdom	17.53%	43.13%	33.77%
United States	12.22%	59.76%	30.02%
Japan	2.50%	19.44%	16%
Czech Republic	11.89%	38.90%	25.75%

Source: Krueger and Meyer 2002



Source: OECD Benefits and Wages 2002

Main Questions in Social Insurance

1. Why have social (as opposed to private, or any) insurance?
2. What type of SI system maximizes social welfare?

- Tradeoff between two forces:
  - **Benefits** – reducing risk (fluctuations in consumption)
  - **Distortion** – changes in incentives for workers and firms –> inefficient behavior and DWL
- Generate new distortions as you fix the problem you set out to solve –> second-best solution
- Identify optimal policy by combining theoretical models of social insurance with empirical evidence on program effects

Useful Background Reading

1. Institutional details: see handout posted on course website
2. Expected utility theory: See MWG or other graduate texts
3. Empirical program evaluation methods: Duflo handout on website
4. Survival analysis: Kiefer (1988 JEL)
5. Surveys: Krueger and Meyer Handbook 2002 (empirics), Chetty Ann Rev. 2009 (theory)

Why have social insurance?

- Motivation for insurance: reduction in risk for risk-averse individuals
    - Unemp Ins: risk of involuntary unemployment
    - Workers' comp and DI: risk of injuries/disabilities
    - Social Security annuity: risk of living too long
- But why is **government** intervention needed to provide this insurance?
- Possible sources of market failure here:
    1. Informational problems (adverse selection)
    2. Individual optimization failures (myopia/improper planning)
    3. Macroeconomic shocks

Adverse Selection as a Motivation for SI

- Key paper: Rothschild and Stiglitz (1976); see MWG Ch. 13 for a good review
- Consider an environment with asymmetric information, e.g. individuals know risk of losing job but insurer does not
- Main result: can lead to market failure where no equilibrium supports provision of insurance
- Government intervention through mandated insurance can increase welfare

Rothschild-Stiglitz model

- Economy with two types, low-risk (L) and high-risk (H)
- A fraction f of the individuals are high-risk
- Type L has a chance p_L of becoming unemployed in a given year
- Type H has a chance $p_H > p_L$ of becoming unemployed.
- In good state (state 1), income is E_1 for both types; in bad state, income is $E_2 < E_1$.

Rothschild-Stiglitz: Key Assumptions

1. **Static model**: individuals arrive in the period either employed or unemployed; no savings/dynamics.
2. **No moral hazard**: agents choose insurance contract but make no choices after signing a contract.
3. Insurance market is **perfectly competitive**, so firms earn zero profits in equilibrium.

Rothschild-Stiglitz: Contracts

- An insurance contract is described by a vector $\alpha = (\alpha_1, \alpha_2)$
  - Consumption in the two states: $(E_1 - \alpha_1, E_2 + \alpha_2)$
- Type i's expected utility is

$$V_i(\alpha) = (1 - p_i)u(E_1 - \alpha_1) + p_i u(E_2 + \alpha_2)$$

- Any contract that earns non-negative profits is feasible
  - Zero-profit condition $\Rightarrow$ firms price insurance s.t.

$$\alpha_2 = \frac{1-p}{p}\alpha_1$$

where p is risk rate of those who purchase contract.

Rothschild-Stiglitz: Equilibrium

Definition

An **equilibrium** is defined by a set of insurance contracts such that
(1) individuals optimize: both types cannot find a better contract than the ones they chose
(2) firms optimize: all firms earn zero profits

- Two types of equilibrium:

1. **Pooling**: both types are offered the same contract α.
2. **Separating**: high-risk types choose a contract α_H while low-risk types choose a different contract α_L.

Rothschild-Stiglitz: First Best Solution

- In first best, insurer can distinguish types (perfect information)
  - In this case, equilibrium is separating
- Plugging in $\alpha_2 = \frac{1-p_i}{p_i}\alpha_1$, each type solves

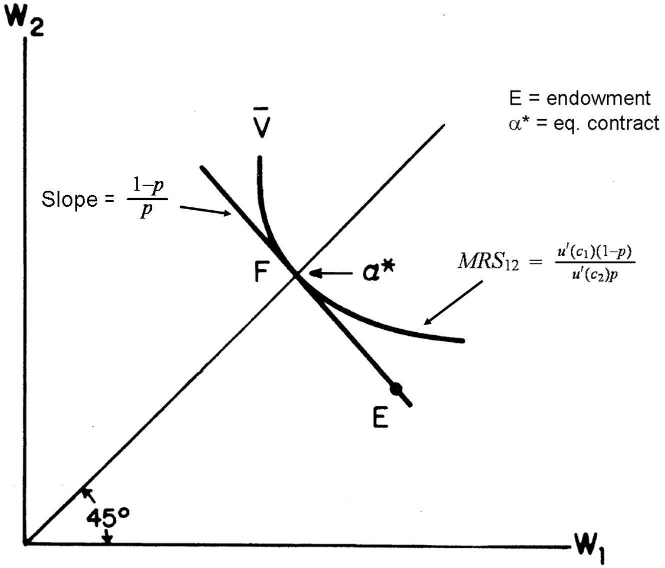
$$\max_{\alpha_1}(1-p_i)u(w-\alpha_1)+p_i u(w+\frac{1-p_i}{p_i}\alpha_1).$$

Solution

Set $MRS_{12} = \frac{1-p_i}{p_i}$, *i.e.* $u'(c_1) = u'(c_2)$, *i.e. full insurance*

- Both types are perfectly insured: earn their expected income $(1-p_i)w$ regardless of the state.

Equilibrium with Perfect Information



Source: Rothschild and Stiglitz 1976

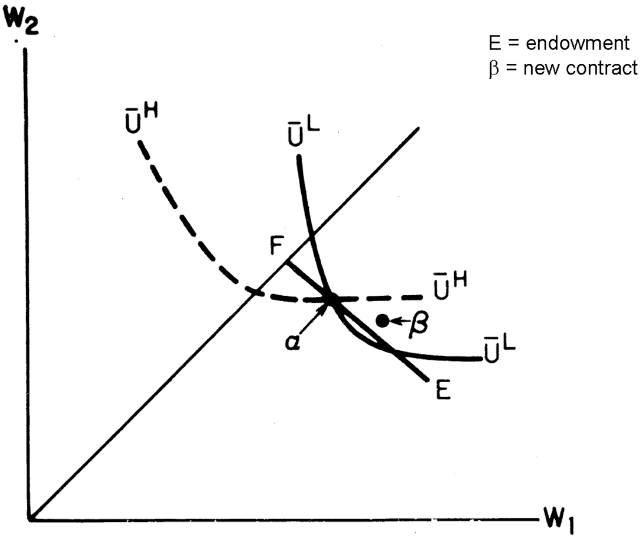
Rothschild-Stiglitz: Second Best Problem

- Firms cannot distinguish types in practice, because they cannot determine true layoff risks, illness history, etc.
- With contracts above, all the high risk types buy the low risk's contract and insurer goes out of business
- Hence optimal contracts differ when information is asymmetric

Rothschild-Stiglitz: Second Best Solution

- **Result #1:** no pooling equilibrium exists
- If H and L types are pooled in a contract α,low-risk types lose money in expectation.
- Zero-profit condition requires $\alpha_2 = \frac{1-\overline{p}}{\overline{p}}\alpha_1$ but $\overline{p} > p_L$.
  - Low-risk type gets fewer dollars in state 2 than he should if the insurance were fair for him.
- Creates an opportunity for a new insurer to enter and "pick off" low risk types by offering slightly less insurance at a better price: higher c_1, lower c_2
  - Only low risk types switch, because they value c_1 more.

No Pooling Equilibrium with Asymmetric Information



Source: Rothschild and Stiglitz 1976

Rothschild-Stiglitz: Second Best Solution

- **Result #2:** in a separating eq, Type H obtains full insurance and Type L is under-insured
- **Intuition:** in any sep. eq., both types are getting actuarially fair insurance because of the zero-profits condition
  - For H, no cost to firm in providing full ins. (worst that can happen is that L will join the pool, raising profits)
  - But for L, full ins. would create an incentive for H to buy this (cheaper) policy, forcing firm into negative profits
- Incentive constraints always bind downward – "no distortion at the top" result in standard asymmetric info. models
- In eq., L gets as much ins as possible without inducing H to deviate and pretend to be low-risk

Rothschild-Stiglitz: Gains from Government Mandate

- There can be gains from government intervention through mandated insurance
- Consider an example where

$$\begin{aligned} E_1 &= 100, E_2 = 0 \\ u(c) &= \sqrt{c}, p_L = \frac{1}{4}, p_H = \frac{3}{4}, f = 10\% \end{aligned}$$

- In candidate separating eq., type H gets perfect insurance:

$$EU_H = u(100(1-p_H)) = \sqrt{100 \cdot \frac{1}{4}} = 5$$

Rothschild-Stiglitz: Second Best Solution

- Type L gets as much ins. as possible without making H want to deviate at actuarially fair rate for L:

$$5 = \frac{1}{4}\sqrt{100 - \alpha_1^L} + \frac{3}{4}\sqrt{\frac{1 - p_L}{p_L}\alpha_1^L}$$

- Solving gives $\alpha_1^L = \$3.85, \alpha_2^L = \$11.55$ – nowhere near full insurance for low risk type.

- Note that expected utility for low risk type is

$$EU_L = \frac{3}{4}\sqrt{100 - 3.85} + \frac{1}{4}\sqrt{3 \cdot 3.85} = 8.2.$$

Rothschild-Stiglitz: Second Best Solution

- Now suppose govt. comes in and mandates pooled insurance at actuarial rate. Everyone gets an income of

$$(\frac{9}{10}\frac{3}{4} + \frac{1}{10}\frac{1}{4})100 = \frac{7}{10}100 = 70.$$

- H benefits from this: now pooling with less risky people
- But L benefits too! Expected utility is $\sqrt{70} > 8.2$

Rothschild-Stiglitz: Second Best Solution

- Because there are relatively few high risk types, L types benefit from pooling with them and getting full insurance coverage.
- Note: pooled contract of 70 could be offered by a private firm, destroying separating eq. proposed above
    - Hence there is actually no equilibrium in this example

Adverse Selection as a Motivation for SI

- More generally, consider an economy in which people differ in their risks of becoming unemployed
- Adverse selection can destabilize the market:
  - Firm provides UI but lowest-risk (tenured people) drop out $\Rightarrow$ rates have to rise
  - But then even moderate-risk types opt out $\Rightarrow$ rates rise further, more drop out, ...
  - Could cause unraveling to the point where virtually no one is insured by private market
  - UI program that pools everyone can lead to (ex-ante) welfare improvements
- What tool does the govt. have that private sector does not? Ability to mandate

Adverse Selection: Empirical Evidence

- Empirical evidence shows that adverse selection is a real source of market failures in practice
- General test: "positive correlation" property in equilibrium
  - Are those who buy more insurance more likely to file claims?
  - Could be driven by both moral hazard + AS but not in certain contexts such as death
- Example: Finkelstein and Poterba (2004): adverse selection in U.K. annuity market.
  - Annuities = ins. against the risk of living too long.

Finkelstein and Poterba 2004

- Study two types of annuity markets: compulsory vs. voluntary.
- Examine two features of annuity contracts
  - degree of **backloading** (inflation indexing and escalation of payments over time)
  - **payments to estate** in event of death (guarantees and capital protection).
- Test for positive correlation in two ways
  1. In eq., those who purchase backloaded annuities have lower mortality rates
  2. In eq., those who purchase annuities with payment to estate have higher mortality rates

  - Both effects should be stronger in voluntary markets

TABLE 2
SELECTION EFFECTS AND ANNUITY PRODUCT CHARACTERISTICS

Explanatory Variable	Estimates from Hazard Model of Mortality after Purchasing an Annuity: Compulsory Market (1)	Estimates from Hazard Model of Mortality after Purchasing an Annuity: Voluntary Market (2)	Estimates from Linear Probability Model of Probability of Dying within Five Years: Compulsory Market (3)	Estimates from Linear Probability Model of Probability of Dying within Five Years: Voluntary Market (4)
Index-linked	−.839***	−.894**	−.053***	−.185***
	(.217)	(.358)	(.019)	(.050)
Escalating	−1.085***	−1.497***	−.072***	−.152***
	(.113)	(.253)	(.010)	(.030)
Guaranteed	.019	.216***	.007*	.046***
	(.029)	(.060)	(.004)	(.016)
Capital-protected	...	.056	...	.064***
		(.051)		(.016)
Payment (£100s)	−.003***	.001**	−.0003***	.0003***
	(.0006)	(.0004)	(.0001)	(.0001)
Male Annuitant	.640***	.252***	.044***	.044***
	(.039)	(.051)	(.005)	(.014)
Observations	38,362	3,692	24,481	3,575
Number of deaths in sample	6,311	1,944	2,693	822

Source: Finkelstein and Poterba 2004

Individual Optimization Failures as a Motivation for SI

- Given adverse selection, expect individuals to “self-insure” against temp. shocks by building up savings
- With such buffer stocks, still no need for large social safety nets to insure against temporary shocks such as unemployment
- In practice, individuals appear to be very liquidity constrained when hit by shocks: median job loser has $<$\$200 in assets
- Suggests 1st Welfare thm also does not hold due to *individual* failures to optimize
  - Individuals may misperceive the probability of a layoff
  - Firms may not be able to debias people in equilibrium, leading to role for govt. (Spinnewijn 2009)

Aggregate Shocks as a Motivation for SI

- Private ins. (cross-sectional pooling) relies on idiosyncratic risks so those who are well off can pay those who are poor
- Government is the only entity able to coordinate risk-sharing across different groups that are all affected by negative shocks
  - Inter-generational risk sharing required if everyone is poor at the same time
- Particularly relevant for UI and maybe social security
- Less so for health-related shocks

Optimal Social Insurance

- Now turn to question of optimal design of SI policies
- Take as given that market provides no insurance for some reason
- In the simple Rothschild-Stiglitz model, perfect insurance is optimal
  - But this abstracts from the central moral hazard problem
  - Individuals will not work if they have perfect unemp insurance
  - Must take this distortion into account to find optimal level of social insurance

Unemployment Insurance

- Start with UI: approx. $40 bn/yr. paid to people who get laid off
- **Potential benefits**

1. Smoother path of consumption
2. Better job matches

- **Potential distortions:**

1. Less job search, higher unemployment rate
2. Workers' preferences distorted toward unstable jobs
3. Shirking because fear of job loss not as great
4. Less savings

Optimal UI: Outline

1. Optimal level of UI benefits ignoring firm responses [Baily-Chetty model]
    - Theory applies to all the income security programs discussed later
2. Distortions to firms' layoff decisions due to imperfect exp rating [Feldstein model]
3. Other issues: Post-unemployment outcomes, general equilibrium effects

Replacement Rate

- Common measure of program's size is its "replacement rate"

$$r = \frac{\text{net benefit}}{\text{net wage}}$$

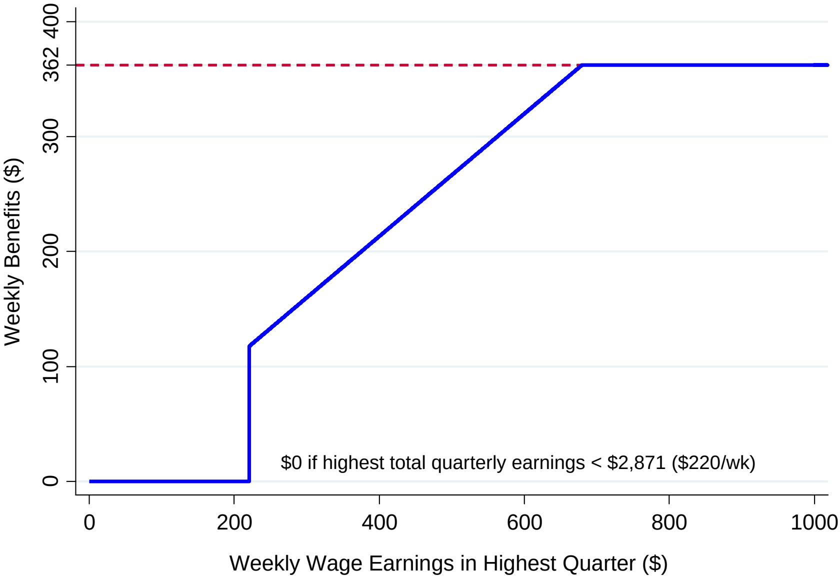
- UI reduces agents' effective wage rate from finding a new job to $w(1-r)$

- Feldstein (1978): UI makes effective wages very low because of interaction with tax system:

  - 1970: No tax $\Rightarrow r = \frac{(0.5)w}{(1-.18-.05-.07)w} = 72\%$
  - Incentives worse for some subgroups: secondary income earner faces MTR of 50% $\Rightarrow r = 1.3$

- Today, federal income taxes paid on UI benefits, so rep. rate is 50-60%

Unemployment Insurance Benefit Schedule in Michigan, 2009



Source: Michigan Department of Energy, Labor, and Economic Growth 2009

Baily-Chetty model

- Canonical analysis of optimal level of UI benefits: Baily (1978)
- Shows that the optimal benefit level can be expressed as a fn of a small set of parameters in a static model.
- Once viewed as being of limited practical relevance because of strong assumptions
- Chetty (2006) shows formula actually applies with arbitrary choice variables and constraints.
- Parameters identified by Baily are sufficient statistics for welfare analysis $\Rightarrow$ robust yet simple guide for optimal policy.

Baily-Chetty model: Assumptions

1. Fixed wages – no GE effects
2. No distortions to firm behavior (temporary layoffs); implicitly assume perfect experience rating
3. No externalities such as spillovers to search

Baily-Chetty model: Setup

- Static model with two states: high (employed) and low (unemployed)
- Let w_h denote the individual's income in the high state and $w_l < w_h$ income in the low state
- Let A denote wealth, c_h consumption in the high state, and c_l consumption in the low state
- Agent is initially unemployed. Controls probability of being in the bad state by exerting search effort e at a cost $\psi(e)$
- Choose units of e so that the probability of being in the high state is given by $p(e) = e$

Baily-Chetty model: Setup

- UI system that pays constant benefit b to unemployed agents
- Benefits financed by lump sum tax $t(b)$ in the high state
- Govt's balanced-budget constraint:

$$e \cdot t(b) = (1 - e) \cdot b$$

- Let $u(c)$ denote utility over consumption (strictly concave)
- Agent's expected utility is

$$eu(A + w_h - t(b)) + (1 - e)u(A + w_l + b) - \psi(e)$$

Baily-Chetty model: First Best Problem

- In first best, there is no moral hazard problem
- To solve for FB, suppose government chooses b and e joints to maximize agent's welfare:

$$\max_{b,e} e(A + w_h - t) + (1-e)u(A + w_l + b) - \psi(e)$$

$$\text{s.t. } t = \frac{1-e}{e} b$$

- Solution to this problem is $u'(c_e) = u'(c_u) \Rightarrow$ full insurance

Baily-Chetty model: Second Best Problem

- In second best, cannot eliminate moral hazard problem because effort is unobserved by govt.
- Problem: Agents only consider *private* marginal costs and benefits when choosing e
  - Social marginal product of work is w private marginal product is $w - b$
  - Agents therefore search too little from a social perspective, leading to efficiency losses

Baily-Chetty model: Second Best Problem

- Agents maximize expected utility, taking b and $t(b)$ as given

$$\max_{e} eu(A + w_h - t) + (1 - e)u(A + w_l + b) - \psi(e)$$

- Let indirect expected utility be denoted by $V(b, t)$
- Government's problem is to maximize agent's expected utility, taking into account agent's behavioral responses:

$$\max_{b,t} V(b, t)$$
$$\text{s.t. } e(b)t = (1 - e(b))b$$

Baily-Chetty model: Second Best Problem

Problem

Optimal Social Insurance

$$\max_b V(b, t(b))$$
$$s.t.\ \ e(b)t(b) = (1 - e(b))b$$
$$e(b) = \arg\max_e e \cdot u(A + w_h - t) + (1 - e) \cdot u(A + w_l + b) - \psi(e)$$

- Formally equivalent to an optimal Ramsey tax problem with state-contingent taxes

Two Approaches to Optimal Social Insurance

1. **Structural**: specify complete models of economic behavior and estimate the primitives

   - Identify b^* as a fn. of discount rates, nature of borrowing constraints, informal ins. arrangements.
   - Challenge: difficult to identify all primitive parameters in an empirically compelling manner given unobserved heterogeneity

2. **Sufficient Statistic:** derive formulas for b^* as a fn. of high-level elasticities

   - These elasticities can be estimated using reduced-form methods
   - Estimate statistical relationships using quasi-experimental research designs
   - Baily-Chetty solution described below is one example

Baily-Chetty model: Second Best Solution

- At an interior optimum, the optimal benefit rate must satisfy

$$dV/db(b^*) = 0$$

- To calculate this derivative, write $V(b)$ as

$$V(b) = \max_e eu(A + w_h - t(b)) + (1 - e)u(A + w_l + b) - \psi(e)$$

- Since fn has been optimized over e, Envelope Thm. implies:

$$\frac{dV(b)}{db} = (1 - e)u'(c_l) - \frac{dt}{db}eu'(c_h)$$

- Key: can neglect $\frac{\partial e}{\partial b}$ terms

Envelope Condition

- Why can $\frac{\partial e}{\partial b}$ be ignored? Because $\frac{\partial V}{\partial e} = 0$ by agent optimization.

- Contrast with total derivative ignoring optimization of e:

$$\begin{aligned}\frac{dV(b)}{db} &= (1-e)u'(c_l) - \frac{dt}{db}eu'(c_h) \\ &\quad + \frac{\partial e}{\partial b}[(u(c_h) - u(c_l) - \psi'(e)]\end{aligned}$$

- Second term drops out because f.o.c. for e is

$$u(c_h) - u(c_l) = \psi'(e)$$

Kaplan 2009

- Exploiting f.o.c.'s from agent optimization particularly useful in more complex models
- Kaplan (2009): unemployed youth move back in with their parents.
  - How does this affect optimal UI?
- Kaplan takes a structural approach and estimates a dynamic model of the decision to move back home

Sufficient Statistic Approach to Kaplan 2009

- Suppose moving home raises consumption by H and has a cost $g(H)$:

$$
\begin{aligned}
V(b) &= \max_{e,H} eu(A + w_h - t(b)) \\
&\quad + (1-e)[u(A + w_l + b + H) - g(H)] - \psi(e)
\end{aligned}
$$

- Variable H drops out, as did e, because of agent optimization

- Formula derived for $\frac{dV(b)}{db}$ is unaffected by ability to move home:

$$
\frac{dV(b)}{db} = (1-e)u'(c_l) - \frac{dt}{db} eu'(c_h)
$$

where c_l is measured in the data as including home consumption (H)

Baily-Chetty model: Second Best Solution

- The government's UI budget constraint implies

$$\begin{aligned} \frac{dt}{db} &= \frac{1-e}{e} - \frac{b}{e^2}\frac{de}{db} = \frac{1-e}{e}(1+\frac{\varepsilon_{1-e,b}}{e}) \\ \Longrightarrow & \frac{dV(b)}{db} = (1-e)\{u'(c_l) - (1+\frac{\varepsilon_{1-e,b}}{e})u'(c_h)\} \end{aligned}$$

- Setting $dV(b)/db = 0$ yields the optimality condition

$$\frac{u'(c_l) - u'(c_h)}{u'(c_h)} = \frac{\varepsilon_{1-e,b}}{e}$$

- LHS: benefit of transferring \$1 from high to low state

- RHS: cost of transferring \$1 due to behavioral responses

Baily-Chetty model: Second Best Solution

$$\frac{u'(c_l) - u'(c_h)}{u'(c_h)} = \frac{\varepsilon_{1-e,b}}{e}$$

- This equation provides an exact formula for the optimal benefit rate
- Implementation requires identification of $\frac{u'(c_l)-u'(c_h)}{u'(c_h)}$
- Three ways to identify $\frac{u'(c_l)-u'(c_h)}{u'(c_h)}$ empirically
  1. Baily (1978), Gruber (1997), Chetty (2006): cons-based approach
  2. Shimer and Werning (2007): reservation wages
  3. Chetty (2008): moral hazard vs liquidity

Baily-Chetty model: Consumption-Based Formula

- Write marginal utility gap using a Taylor expansion

$$u'(c_l) - u'(c_h) \approx u''(c_h)(c_l - c_h)$$

- Defining coefficient of relative risk aversion $\gamma = \frac{-u''(c)c}{u'(c)}$, we can write

$$\begin{aligned} \frac{u'(c_l) - u'(c_h)}{u'(c_h)} &\approx -\frac{u''}{u'} c_h \frac{\Delta c}{c} \qquad (1) \\ &= \gamma \frac{\Delta c}{c} \end{aligned}$$

- Gap in marginal utilities is a function of curvature of utility (risk aversion) and consumption drop from high to low states

Baily-Chetty Consumption-Based Formula

Theorem

The optimal unemployment benefit level b^* satisfies

$$\gamma \frac{\Delta c}{c}(b^*) \approx \frac{\varepsilon_{1-e,b}}{e}$$

where

$$\begin{aligned}
\frac{\Delta c}{c} &= \frac{c_h - c_l}{c_h} = \textit{consumption drop during unemployment} \\
\gamma &= -\frac{u''(c_h)}{u'(c_h)} c_h = \textit{coefficient of relative risk aversion} \\
\varepsilon_{1-e,b} &= \frac{d \log 1 - e}{d \log b} = \textit{elast. of probability of unemp. w.r.t. benefits}
\end{aligned}$$

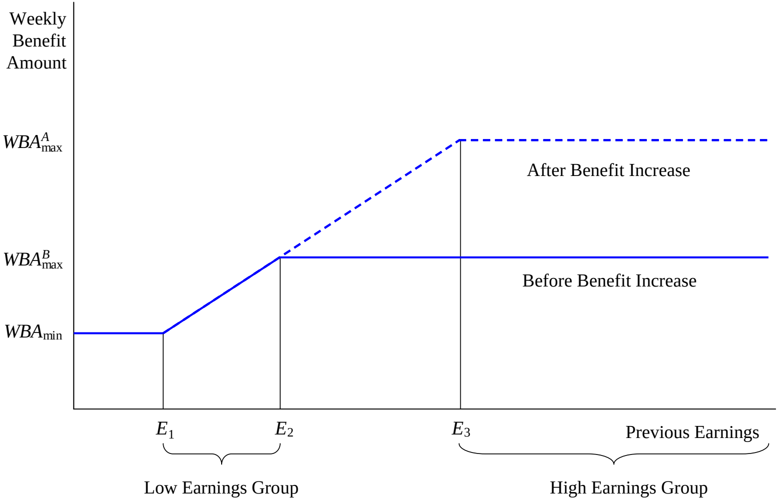
Baily-Chetty Consumption-Based Formula

$$\gamma \frac{\Delta c}{c}(b^*) \approx \frac{\varepsilon_{1-e,b}}{e}$$

- Intuition for formula: LHS is marginal social benefit of UI, RHS is marginal social cost of UI
- Extends to model where agent chooses N other behaviors and faces M other constraints, subject to some regularity conditions (Chetty 2006).
  - Envelope conditions used above still go through with arbitrary choice vars.
- Empirical work on UI can essentially be viewed as providing estimates of the three key parameters $(\gamma, \frac{\Delta c}{c}, \varepsilon)$.

Empirical Estimates: Duration Elasticity

- Early literature used cross-sectional variation in replacement rates.
- Problem: comparisons of high and low wage earners confounded by other factors.
- Modern studies use exogenous variation from policy changes (e.g. Meyer 1990)



Source: Krueger and Meyer 2002

Hazard Models

- Define hazard rate h_t = number that find a job at time t divided by number unemployed at time t
  - This is an estimate of the probability of finding a job at time t conditional on being unemployed for at least t weeks
- Standard specification of hazard model: Cox "proportional hazards"

$$h_t = \alpha_t \exp(\beta X)$$

- Here α_t is the non-parametric "baseline" hazard rate in each period t and X is a set of covariates
- Semi-parametric specification – allow hazards to vary freely across weeks and only identify coefficients off of variation across spells

Hazard Models

- Useful to rewrite expression as:

$$\log h_t = \log \alpha_t + \beta X$$

- Key assumption: effect of covariates proportional across all weeks

$$\frac{d \log h_t}{dX} = \beta = \frac{d \log h_s}{dX} \forall t, s$$

- If a change in a covariate doubles hazard in week 1, it is forced to double hazard in week 2 as well

- Restrictive but a good starting point; can be relaxed by allowing for time varying covariates X_t

Meyer 1990

- Meyer includes log UI benefit level as a covariate:

$$\log h_t = \log \alpha_t + \beta_1 \log b + \beta_2 X$$

- In this specification,

$$\frac{d \log h_t}{d \log b} = \beta_1 = \varepsilon_{h_t,b}$$

- Note: in exponential survival (constant-hazard) models, $\varepsilon_{h_t,b} = -\varepsilon_{1-e,b}$
- Meyer estimates $\varepsilon_{h_t,b} = -0.9$ using administrative data for UI claimants
- Subsequent studies get smaller estimates; consensus: $\varepsilon_{h_t,b} = -0.5$ (Krueger and Meyer 2002)

Hazard Model Estimates[a]

Variable	Specification (1)	(2)	(3)	(4)	(5)
Number of dependents	−.0418	−.0422	−.0416	−.0386	−.0386
	(0.0169)	(0.0171)	(0.0168)	(0.0239)	(0.0242)
1 = married, spouse present	.1302	.1221	.1315	.1006	.1001
	(0.0508)	(0.0515)	(0.0507)	(0.0722)	(0.0730)
1 = white	.2097	.2230	.2171	.2337	.2364
	(0.0572)	(0.0579)	(0.0568)	(0.0834)	(0.0841)
Years of schooling	−.0276	−.0275	−.0272	−.0177	−.0176
	(0.0083)	(0.0084)	(0.0083)	(0.0123)	(0.0124)
Log UI benefit level	−.8782	−.8157	−.8478	−.8685	−.8757
	(0.1091)	(0.1096)	(0.1088)	(0.2042)	(0.2065)
Log pre-UI after tax wage	.5630	.5651	.5530	.7289	.7411
	(0.0855)	(0.0860)	(0.0848)	(0.1415)	(0.1433)

Source: Meyer 1990

Consumption Smoothing Benefits of UI

- Gruber (1997) takes the Baily formula to the data by estimating consumption smoothing response.
- Same methodology as Meyer
  - Uses cross-state and time variation and uses drop in food consumption as the LHS variable.
  - Data: PSID food consumption

Gruber 1997

- Gruber estimates
$$\frac{\Delta c}{c} = \beta_1 + \beta_2 \frac{b}{w}$$
- Finds $\beta_1 = 0.24$, $\beta_2 = -0.28$
- Without UI, cons drop would be about 24%
- Mean drop with current benefit level ($b = 0.5$) is about 10%
- Implies a 10 pp increase in UI replacement rate causes 2.8 pp reduction in cons. drop
- Convincing evidence that ins. markets are not perfect and UI does play a consumption smoothing role

Consumption Smoothing Benefits of UI

- What is substituting for/getting crowded out by UI?
- Cullen and Gruber (2000) emphasize spousal labor supply
  - Study wives of unemployed husbands
  - Examine wives' labor supply as a fn of level of husbands' UI benefits
  - For a $100/wk increase in UI benefit, wives work 22 hrs less per month
  - In the absence of UI, wives would work 30% more during the spell than they do now
- Engen and Gruber (1995) document that higher UI benefits lower ex-ante savings, another crowdout channel

Calibrating the Model

- Gruber calibrates Baily's model using his and Meyer's estimates:

$$\gamma \frac{\Delta c}{c} \approx \frac{\varepsilon_{1-e,b}}{e}$$

$$\gamma(\beta_1 + \beta_2 \frac{b^*}{w}) = \frac{\varepsilon_{1-e,b}}{e}$$

- Solving for the optimal replacement rate yields:

$$\frac{b^*}{w} = \frac{\varepsilon_{1-e,b}/e}{\beta_2}(\frac{1}{\gamma}) - \frac{\beta_1}{\beta_2}$$

- Plugging in $\varepsilon_{1-e,b} = .43$ as in Gruber (1997) and $e = .95$ (5% unemployment rate) yields:

$$\frac{b^*}{w} = -(\frac{.43/.95}{.28} \cdot \frac{1}{\gamma}) - \frac{(-.24)}{.28}$$

Calibrating the Model

- Results: $\frac{b}{w}^*$ varies considerably with γ

γ	1	2	3	4	5	10
$\frac{b}{w}^*$	0	0.05	0.31	0.45	0.53	0.7

- Gruber: introspection and existing evidence suggests $\gamma < 2$

  $\Rightarrow$ optimal program small (i.e. replacement rates should be much lower than is observed)

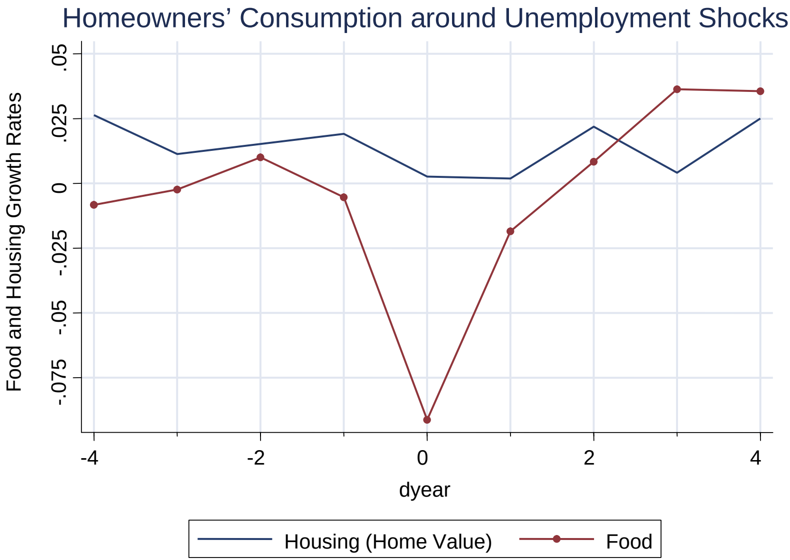
- Surprising result in view of $200bn income security expenditure

- Parameter that is most poorly identified: γ

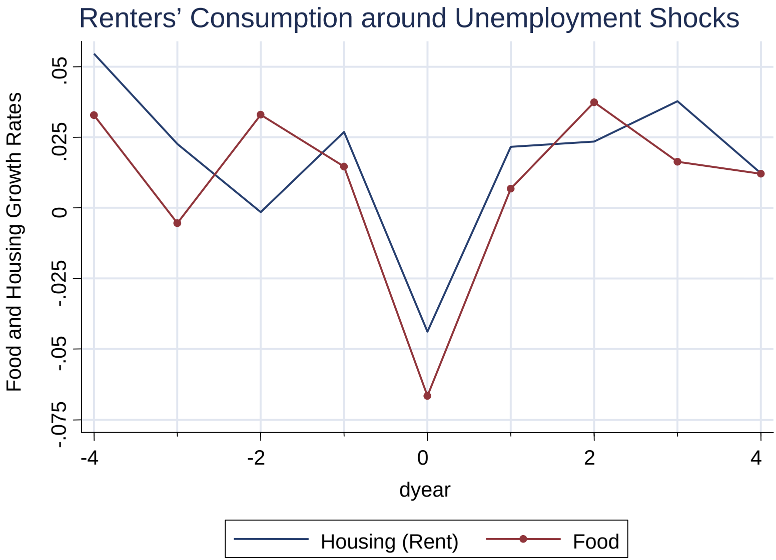
  - Risk preferences appear to vary substantially according to situation.

Chetty and Szeidl (2007): Consumption Commitments

- Standard expected utility model: one composite consumption good c
- Composite commodity assumes that people can cut back on all consumption goods at all times freely.
- E.g. when unemployed, cut consumption of food, housing, cars, furniture, etc.
- In practice, difficult to adjust many elements of consumption in short run because of fixed adjustment costs



Source: Chetty and Szeidl 2007



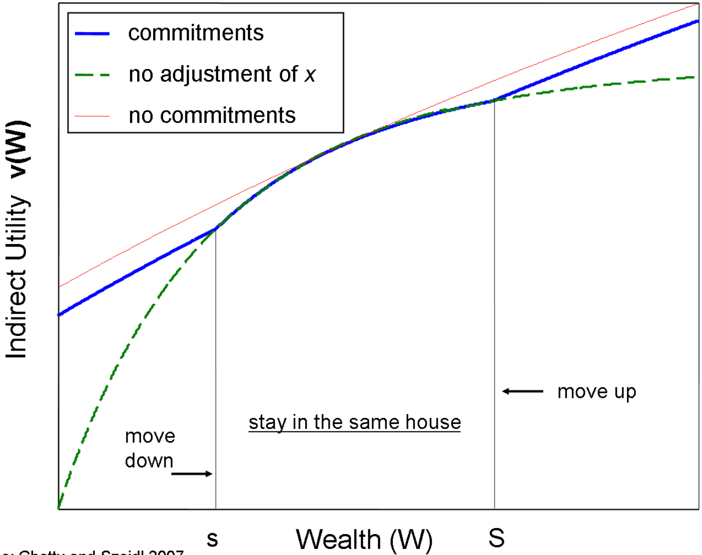
Source: Chetty and Szeidl 2007

Commitments and Risk Aversion

- How do commitments affect risk aversion?
- Utility over two goods, food and housing:

$$U(f, h) = u(f) + v(h).$$

- Adjusting h requires payment of a fixed cost k
- Agent follows an (S, s) policy



Source: Chetty and Szeidl 2007

Commitments Model: Implications for UI

- Commitments amplify risk aversion
  - Ex: 50% food, 50% housing
  - Suppose unemployed agent forced to cut expenditure by 10%
  - Then have to cut food cons by 20%, leading to larger welfare loss
- Model of commitments suggests that γ might actually exceed 4 for unemployment shocks

γ	1	2	3	4	5	10
$\frac{b}{w}^*$	0	0.05	0.31	**0.45**	0.53	0.7

- Problem: γ hard to estimate precisely by context

Alternative Formulas for Optimal UI

- Since γ and $\frac{\Delta c}{c}$ are hard to identify, recent work has sought alternative ways of calculating optimal benefit.

- Two approaches

  1. Moral hazard vs. liquidity (Chetty 2008)
  2. Reservation wage response (Shimer Werning 2007)

- Note that any formula is only one representation of optimal benefit

Chetty 2008: Moral Hazard vs. Liquidity

- Discrete time dynamic search model
- Individual lives for T periods
- Interest rate and discount rate equal to 0
- Individual loses job in period $t = 0$
- Let $u(c_t)$ denote flow utility over cons.
- Dynamic budget constraint:

$$A_{t+1} = A_t + y_t - c_t$$

- Asset limit: $A_t \geq L$

Chetty 2008: Baseline Assumptions

1. Assets prior to job loss exogenous
2. No heterogeneity
3. Fixed wages: choose only search intensity, not reservation wage

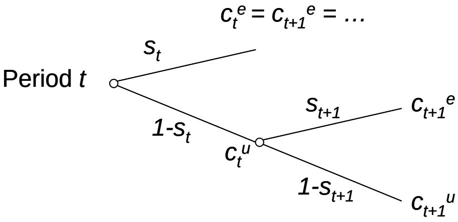
- Each of these is relaxed in paper, so model nests search models used in structural literature (e.g. Wolpin 1987)

Chetty 2008: Job Search Technology

- If unemployed in period t, worker first chooses search intensity s_t
- Finds a job that begins immediately in period t with probability s_t
- If job found, consumes c_t^e. Jobs are permanent, pay wage $w_t - \tau$.

Chetty 2008: Job Search Technology

- If no job found: receives benefit b_t, consumes c_t^u, enters $t+1$ unemployed
- Cost of job search: $\psi(s_t)$



Chetty 2008: Value Functions

- Value function for agent who finds a job in period t:

$$V_t(A_t) = \max_{A_{t+1} \geq L} u(A_t - A_{t+1} + w - \tau) + V_{t+1}(A_{t+1})$$

- Value function for agent who does not find a job in period t:

$$U_t(A_t) = \max_{A_{t+1} \geq L} u(A_t - A_{t+1} + b_t) + J_{t+1}(A_{t+1})$$

where $J_{t+1}(A_{t+1})$ is value of entering next period unemployed.

- Agent chooses s_t to maximize expected utility

$$J_t(A_t) = \max_{s_t} s_t V_t(A_t) + (1 - s_t) U_t(A_t) - \psi(s_t)$$

Chetty 2008: Optimal Search Behavior

- First order condition for optimal search intensity:

$$\psi'(s_t^*) = V_t(A_t) - U_t(A_t)$$

- Intuitively, s_t is chosen to equate the marginal cost of search effort with the marginal value of search effort.

- Effect of benefits on durations:

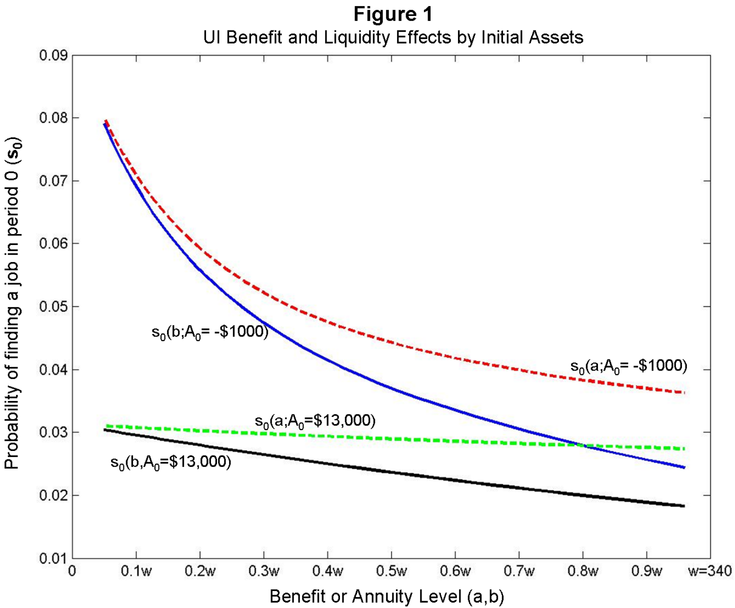
$$\partial s_t / \partial b_t = -u'(c_t^u) / \psi''(s_t)$$

Chetty 2008: Moral Hazard vs. Liquidity Decomposition

- Benefit effect can be decomposed into two terms:

$$\begin{aligned} \partial s_t/\partial A_t &= \{u'(c_t^e) - u'(c_t^u)\}/\psi''(s_t) < 0 \\ \partial s_t/\partial w_t &= u'(c_t^e)/\psi''(s_t) > 0 \\ \Rightarrow \partial s_t/\partial b_t &= \partial s_t/\partial A_t - \partial s_t/\partial w_t \end{aligned}$$

- $\partial s_t/\partial A_t$ is "liquidity effect"
- $\partial s_t/\partial w_t$ is "moral hazard" or price effect
- Liquidity and total benefit effects smaller for agents with better consumption smoothing capacity



Source: Chetty 2008

Chetty 2008: Formula for Optimal UI

$$\begin{aligned}
\partial s_t/\partial A_t &= \{u'(c_t^e) - u'(c_t^u)\}/\psi''(s_t) \geq 0 \\
\partial s_t/\partial w_t &= u'(c_t^e)/\psi''(s_t) > 0 \\
\Rightarrow & \frac{\partial s_t/\partial A_t}{\partial s_t/\partial w_t} = \frac{\text{LIQ}}{\text{MH}} = \frac{u'(c_t^u) - u'(c_t^e)}{u'(c_t^e)}
\end{aligned}$$

- Can show that the Baily formula holds in this model:

$$\frac{u'(c_t^u) - u'(c_t^e)}{u'(c_t^e)} = \frac{\varepsilon_{1-e,b}}{e}$$

- Combining yields formula that depends solely on duration elasticities:

$$\begin{aligned}
\frac{\partial s_t^*/\partial A_t}{\partial s_t^*/\partial b_t - \partial s_t^*/\partial A_t} &= \frac{\varepsilon_{1-e,b}}{e} \\
\frac{\varepsilon_{1-e,A}}{\varepsilon_{1-e,b}\frac{A}{b} - \varepsilon_{1-e,A}} &= \frac{\varepsilon_{1-e,b}}{e}
\end{aligned}$$

Intuition for Moral Hazard vs. Liquidity Formula

- Formula is a "revealed preference" approach to valuing insurance
  - Infer value of UI to agent by observing what he would do if money given as a cash-grant without distorted incentives
  - If agent would not use money to extend duration, infer that only takes longer because of price subsidy (moral hazard)
  - But if he uses cash grant to extend duration, indicates that UI facilitates a choice he would make if markets were complete
- Same strategy can be used in valuing other types of insurance
  - Make inferences from agent's choices instead of directly computing costs and benefits of the policy
  - Key assumption: perfect agent optimization

Moral Hazard vs. Liquidity: Evidence

- Two empirical strategies

    1. Divide agents into liquidity constrained and unconstrained groups and estimate effect of benefits on durations using changes in UI laws.

    2. Look at lump-sum severance payments to estimate liquidity effect.

TABLE 1

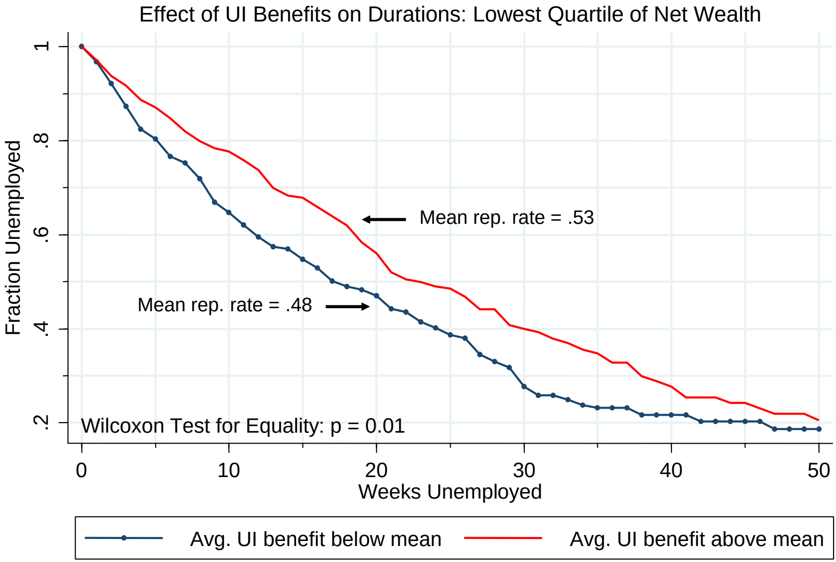
Summary Statistics by Wealth Quartile for SIPP Sample

	Net Liquid Wealth Quartile			
	1 (< -$1,115)	2 (-$1,115-$128)	3 ($128-$13,430)	4 (>$13,430)
Median Liq. Wealth	$466	$0	$4,273	$53,009
Median Debt	$5,659	$0	$353	$835
Median Home Equity	$2,510	$0	$11,584	$48,900
Median Annual Wage	$17,188	$14,374	$18,573	$23,866
Mean Years of Education	12.21	11.23	12.17	13.12
Mean Age	35.48	35.18	36.64	41.74
Fraction Renters	0.43	0.61	0.35	0.16
Fraction Married	0.64	0.59	0.60	0.63

All monetary variables in real 1990 dollars

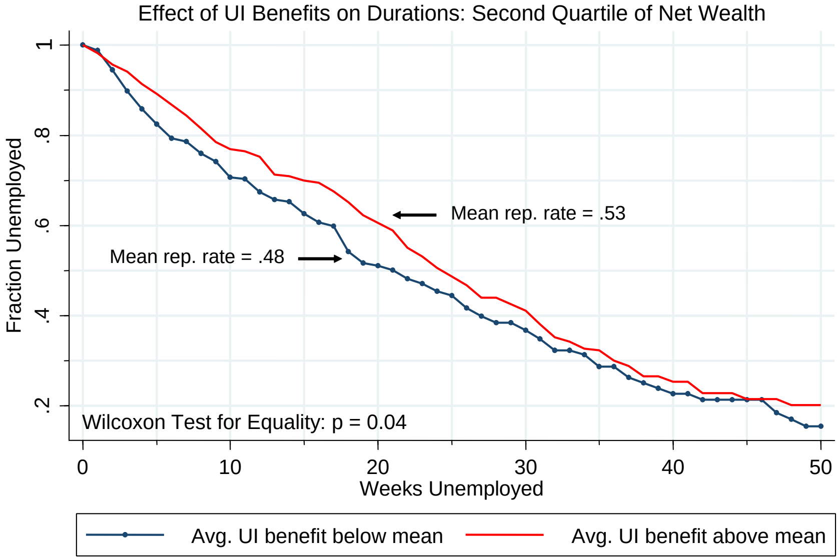
Source: Chetty 2008

Figure 3a



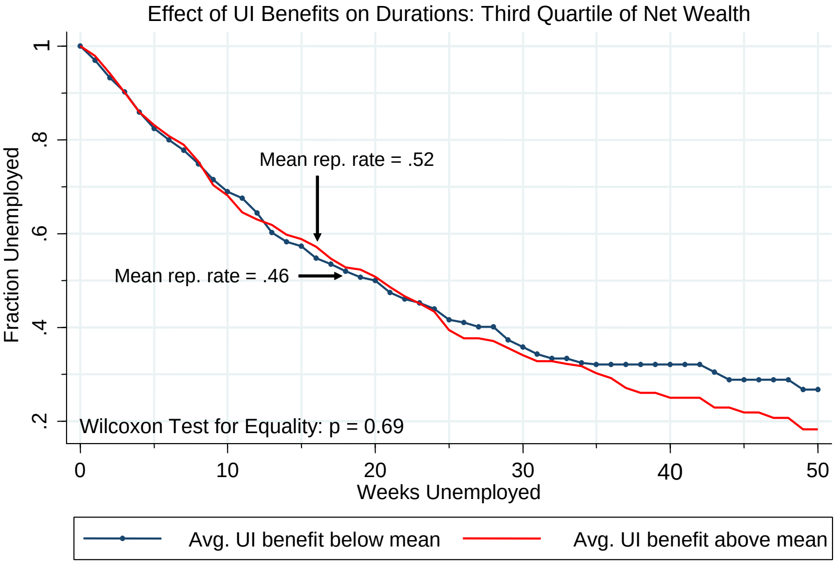
Source: Chetty 2008

Figure 3b



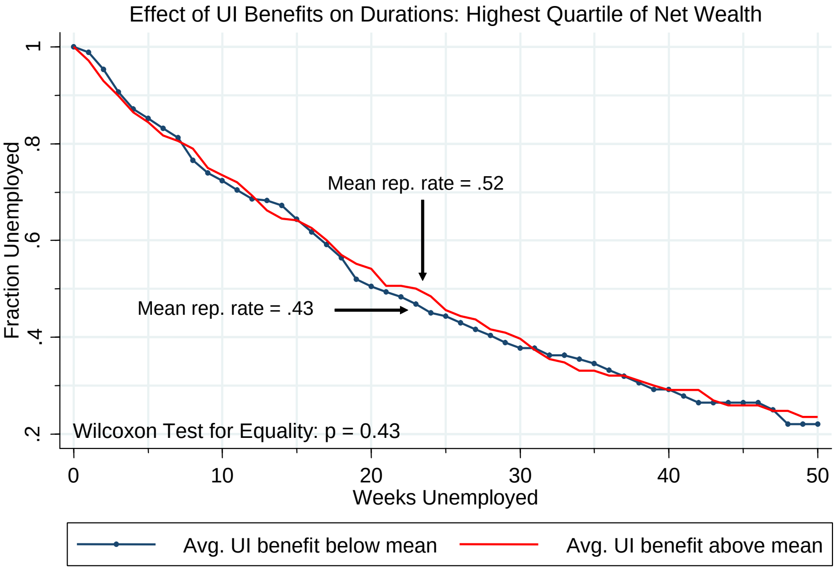
Source: Chetty 2008

Figure 3c



Source: Chetty 2008

Figure 3d



Source: Chetty 2008

TABLE 2
Effect of UI Benefits: Cox Hazard Model Estimates

	(1)	(2)	(3)	(4)	(5)
	Pooled	Stratified	Stratified with Full Controls		
	Full cntrls	No cntrls	Avg WBA	Max WBA	Ind. WBA
log UI ben	**-0.527**				
	(0.267)				
Q1 x log UI ben		**-0.721**	**-0.978**	**-0.727**	**-0.642**
		(0.304)	(0.398)	(0.302)	(0.241)
Q2 x log UI ben		**-0.699**	**-0.725**	**-0.388**	**-0.765**
		(0.484)	(0.420)	(0.303)	(0.219)
Q3 x log UI ben		**-0.368**	**-0.476**	**-0.091**	**-0.561**
		(0.309)	(0.358)	(0.370)	(0.156)
Q4 x log UI ben		**0.234**	**0.103**	**0.304**	**0.016**
		(0.369)	(0.470)	(0.339)	(0.259)
Q1=Q4 p-val		0.039	0.013	0.001	0.090
Q1+Q2=Q3+Q4 p-val		0.012	0.008	0.002	0.062
Number of Spells	4529	4337	4054	4054	4054

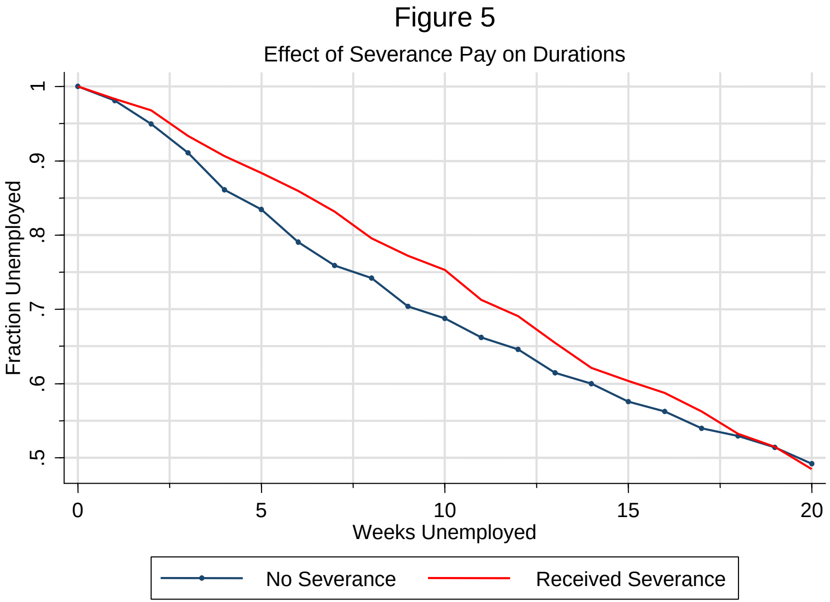
Source: Chetty 2008

TABLE 3

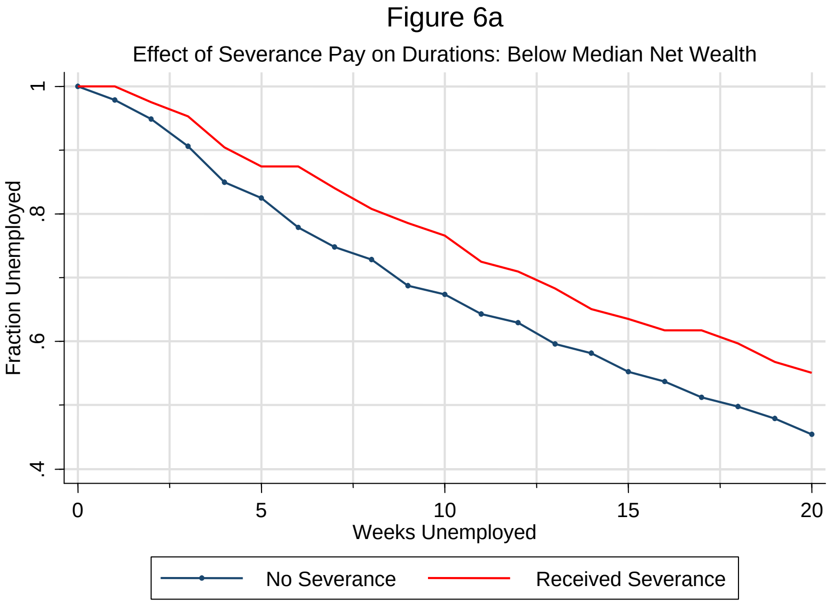
Summary Statistics for Mathematica Data

	Pooled	No Severance (0.83)	Severance (0.17)
Percent dropouts	14%	15%	6%
Percent college grads	17%	13%	34%
Percent married	58%	56%	68%
Mean age	36.2	35.2	40.6
Median pre-unemp annual wage	$20,848	$19,347	$30,693
Median job tenure (years)	1.9	1.5	4.8

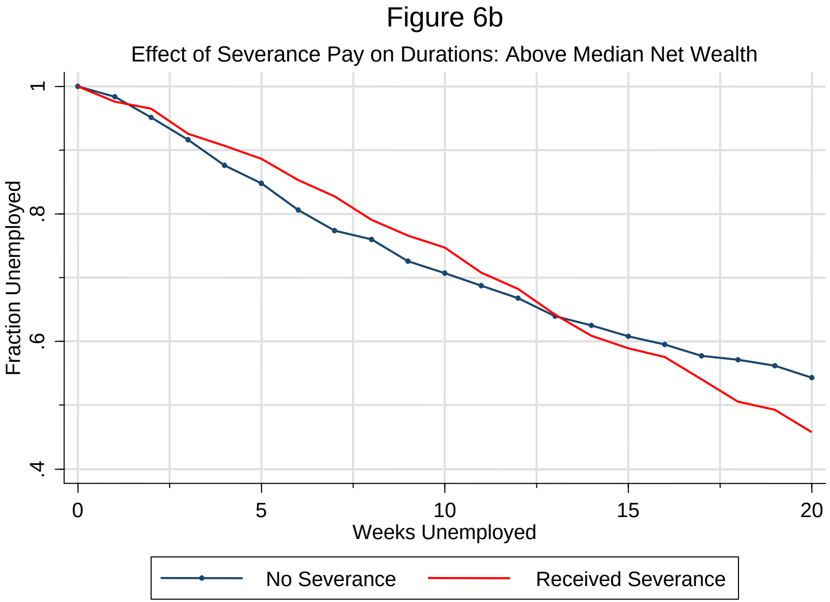
Source: Chetty 2008



Source: Chetty 2008



Source: Chetty 2008



Source: Chetty 2008

TABLE 4
Effect of Severance Pay: Cox Hazard Model Estimates

	Pooled	By Liquid Wealth	By Sev. Amt.
Severance Pay	**-0.233**		
	(0.071)		
(Netliq < Median) x Sev Pay		**-0.457**	
		(0.099)	
(Netliq > Median) x Sev Pay		**-0.088**	
		(0.081)	
(Tenure < Median) x Sev Pay			**-0.143**
			(0.055)
(Tenure > Median) x Sev Pay			**-0.340**
			(0.119)
Equality of coeffs p-val		<0.01	0.03

N=2428; all specs. include full controls.

Source: Chetty 2008

Chetty 2008: Implications for Optimal UI

- Plug reduced-form estimates of de/dA and de/db into formula to calculate dW/db
- Welfare gain from raising benefit level by 10% from current level in U.S. (50% wage replacement) is \$5.9 bil = 0.05% of GDP
  - Small but positive
- In structural models calibrated to match sufficient statistics, dW/db falls rapidly with b
  - Small dW/db suggests we are currently near optimal benefit level

Card, Chetty, and Weber 2007

- Use discontinuities in Austria's unemployment benefit system to estimate liquidity effects
- Severance payment is made by firms out of their own funds
- Formula for sev. pay amount for all non-construction workers:

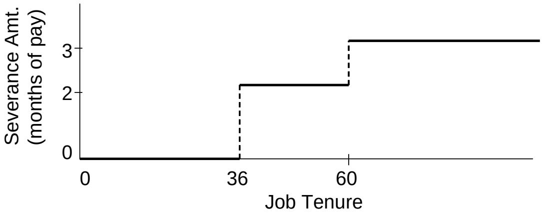
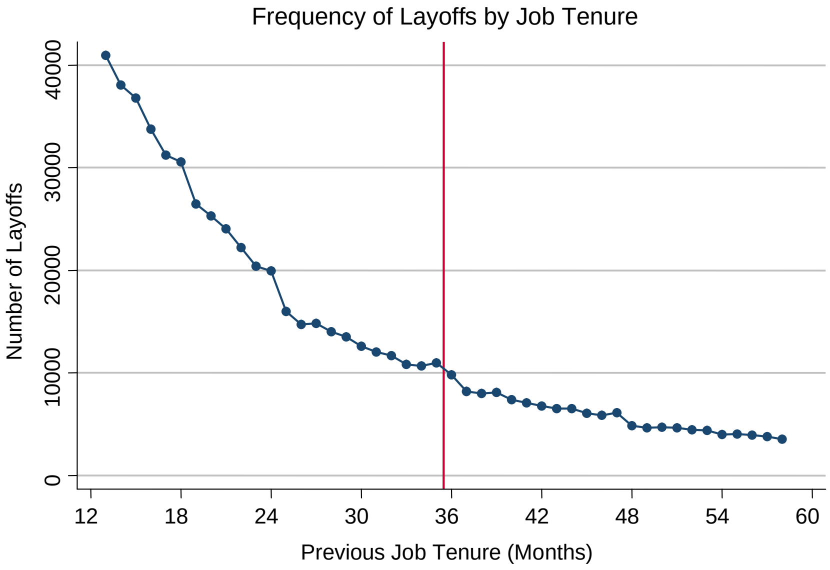


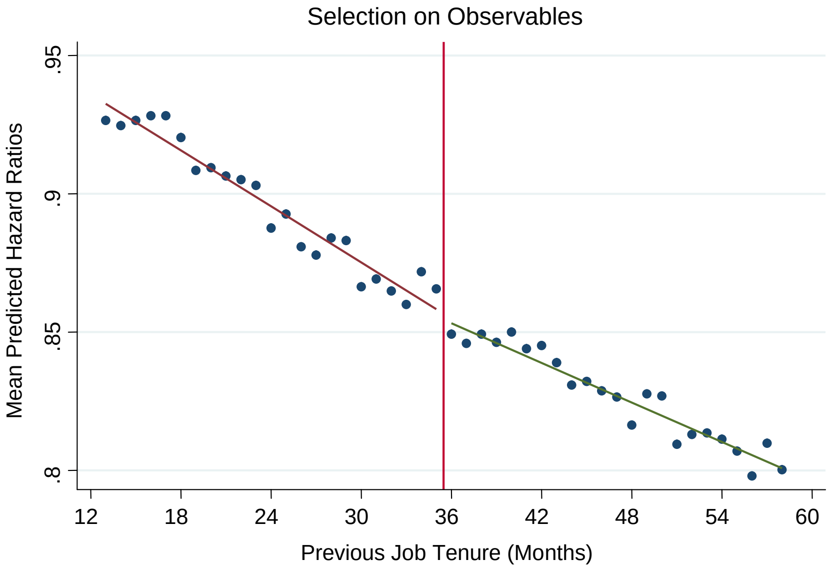
Figure 3



Source: Card, Chetty, and Weber 2007

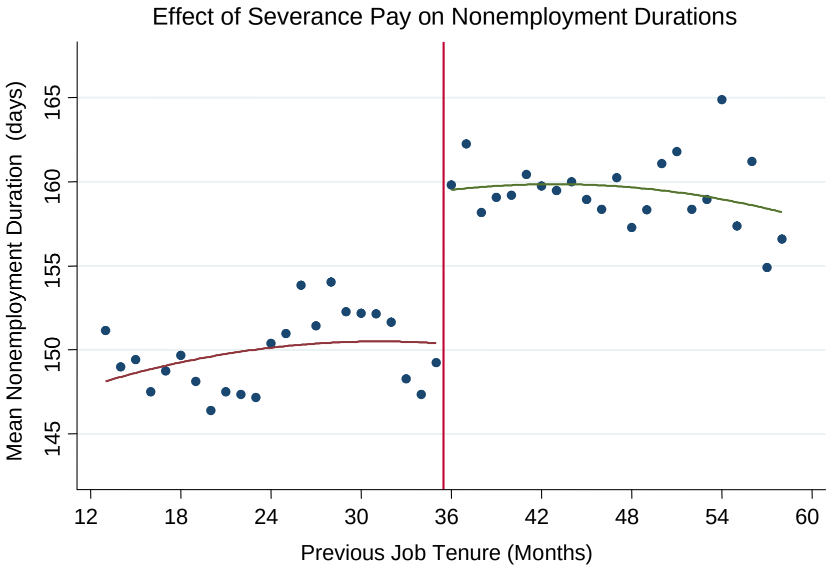
Source: Card, Chetty, and Weber 2007

Figure 4



Source: Card, Chetty, and Weber 2007

Figure 5a



Source: Card, Chetty, and Weber 2007

TABLE 3a

Effects of Severance Pay and EB on Durations: Hazard Model Estimates

	(1) Restricted Sample	(2) Restricted Sample	(3) Full Sample
Severance pay	-0.122		-0.125
	(0.019)		(0.017)
Extended benefits		-0.084	-0.093
		(0.018)	(0.016)
Sample size	512,767	512,767	650,922

NOTE--All specs are Cox hazard models that include cubic polynomials with interactions with sevpay and/or extended benefit dummy.

Source: Card, Chetty, and Weber 2007

Shimer and Werning 2007: Reservation-Wage Model

- Reservation wage model: probability of finding job (e) determined by decision to accept or reject a wage offer, not search effort

- Wage offers drawn from distribution $w \sim F(x)$

- Agent rejects offer if net wage $w - t$ is less than outside option b, implying that probability of finding a job is $e = 1 - F(b + t)$

- Agent's expected value prior to job search:

$$W(b) = (1 - F(b + t))E[u(w - t)|w - t > b] + F(b + t)u(b)$$

- Reservation wage prior to job search satisfies

$$u(\bar{w}_0 - t) = W(b)$$

Shimer and Werning 2007: Reservation-Wage Formula

- Government's problem is

$$\max W(b) = \max u(\bar{w}_0 - t) = \max \bar{w}_0 - t$$

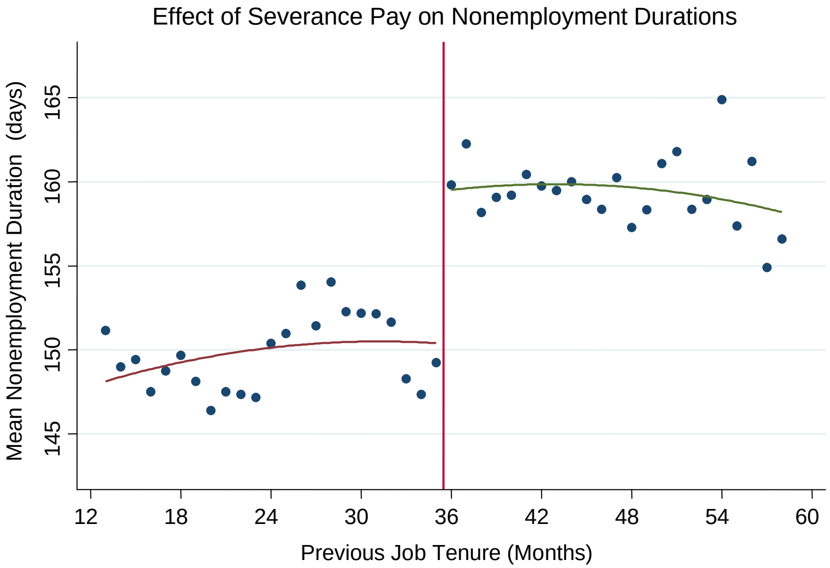
- It follows that

$$\begin{aligned} \frac{dW}{db} &= \frac{d\bar{w}_0}{db} - \frac{dt}{db} \\ &= \frac{d\bar{w}_0}{db} - \frac{1-e}{e} \cdot (1 + \frac{1}{e} \cdot \varepsilon_{1-e,b}) \end{aligned}$$

Shimer and Werning 2007: Reservation-Wage Formula

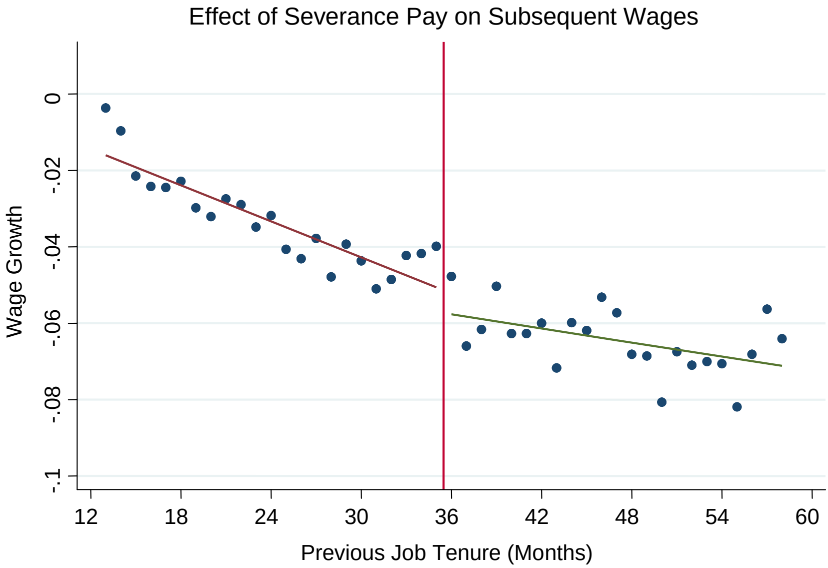
- Implement formula using estimates of $\frac{d\bar{w}_0}{db}$ reported by Feldstein and Poterba (1984)
  - Find gains from raising UI benefits 5 times larger than Chetty (2008)
- But reservation wage elasticity estimates questionable
- Do greater benefits $\rightarrow$ longer durations $\rightarrow$ better outcomes later on? No.
  - Ex: evidence from Austrian discontinuity (Card, Chetty, Weber 2007)
  - Note: all the formulas above take such match quality gains into account via envelope conditions

Figure 5a



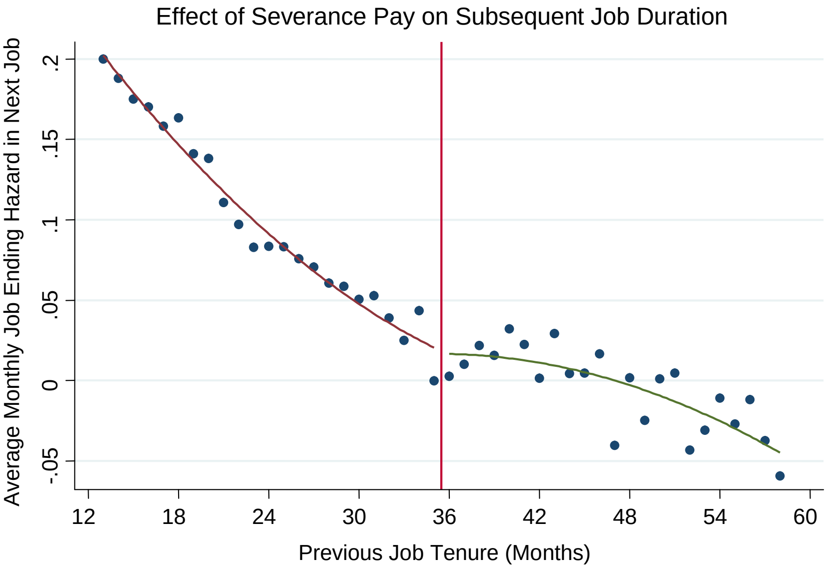
Source: Card, Chetty, and Weber 2007

Figure 10a



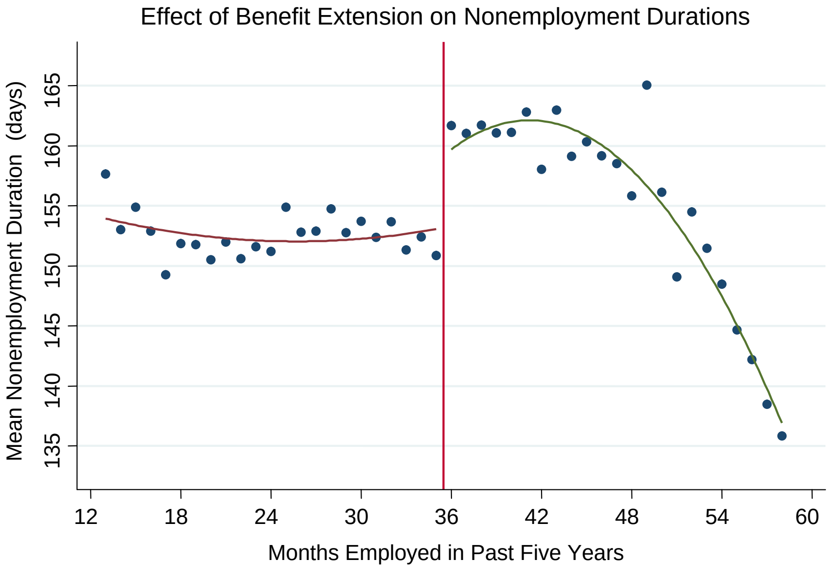
Source: Card, Chetty, and Weber 2007

Figure 10b

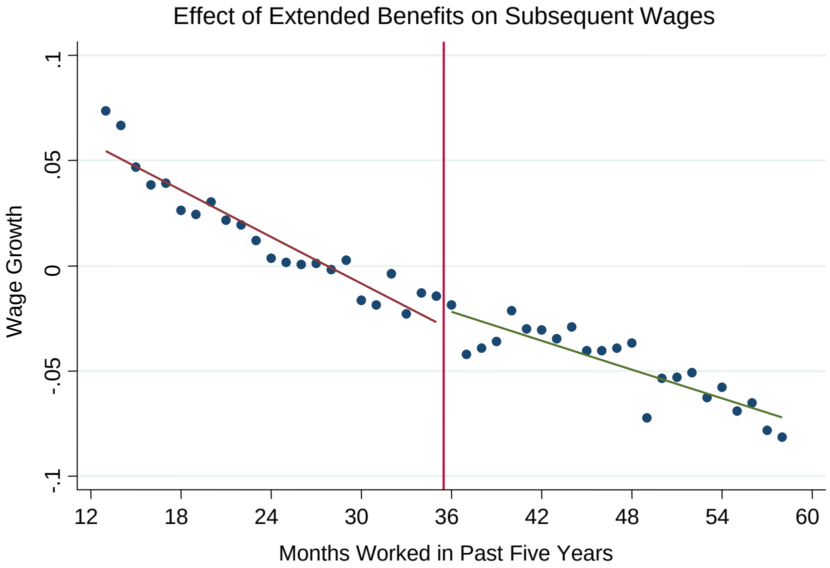


Source: Card, Chetty, and Weber 2007

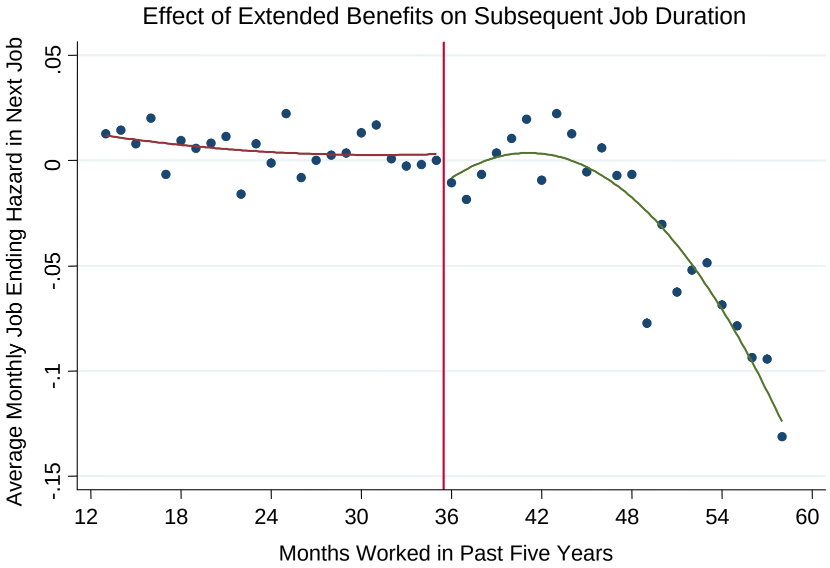
Figure 9a



Source: Card, Chetty, and Weber 2007



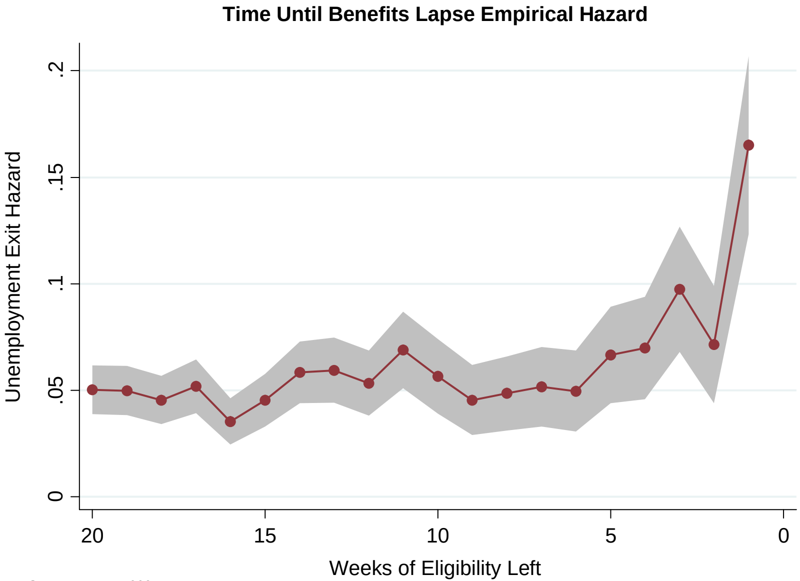
Source: Card, Chetty, and Weber 2007



Source: Card, Chetty, and Weber 2007

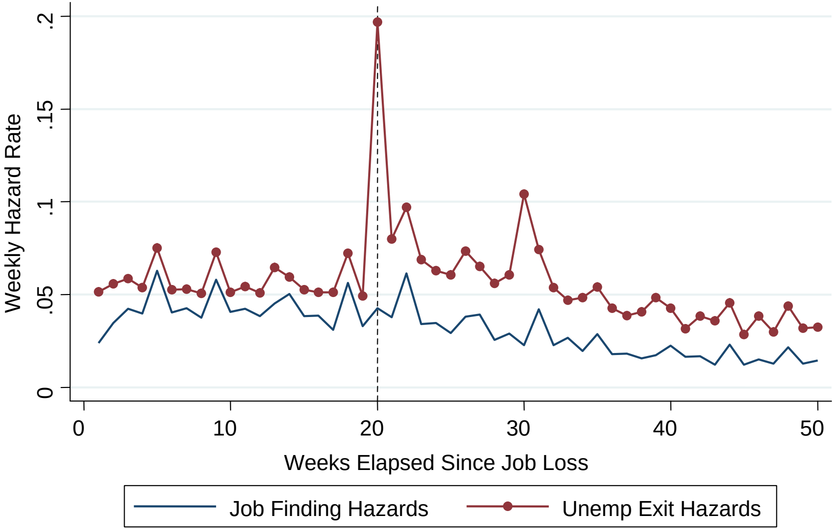
Spike at Benefit Exhaustion

- Most striking evidence for distortionary effects of social insurance: "spike" in hazard rate at benefit exhaustion
  - Katz and Meyer (1990), Meyer (1990), ...
- Traditional measure of hazard: exiting UI system
- Preferred measure based on theory: finding a job
- The two could differ if workers transit off of UI but are still jobless
  - Ex. may not go to pick up last unemployment check
  - Particularly important in European context, where you can remain registered on UI indefinitely



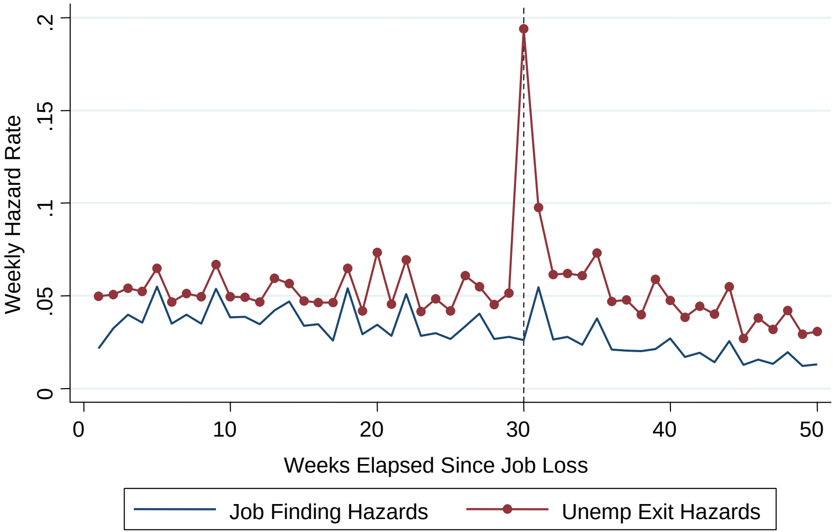
Source: Meyer 1990

Job Finding vs. Unemployment Exit Hazards: 20 Week UI

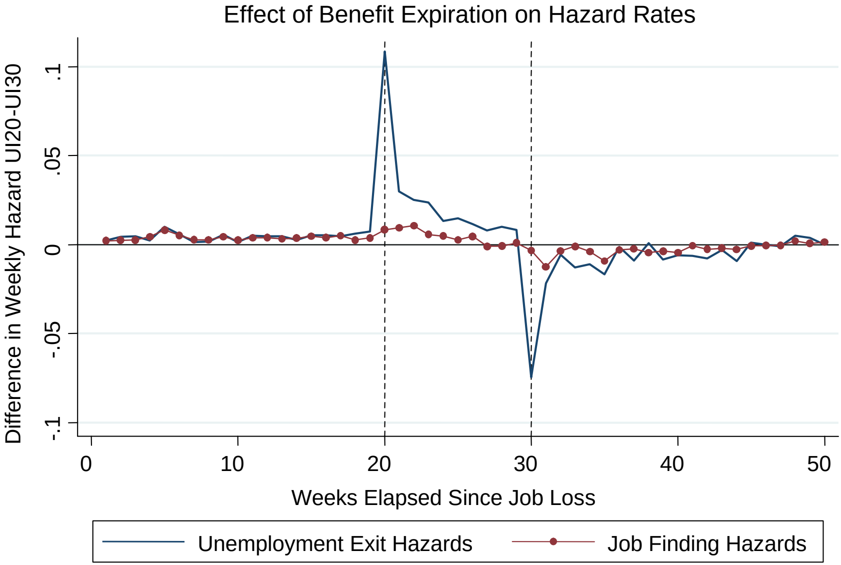


Source: Card, Chetty, Weber 2007b (AER P&P)

Job Finding vs. Unemployment Exit Hazards: 30 Week UI



Source: Card, Chetty, Weber 2007b (AER P&P)

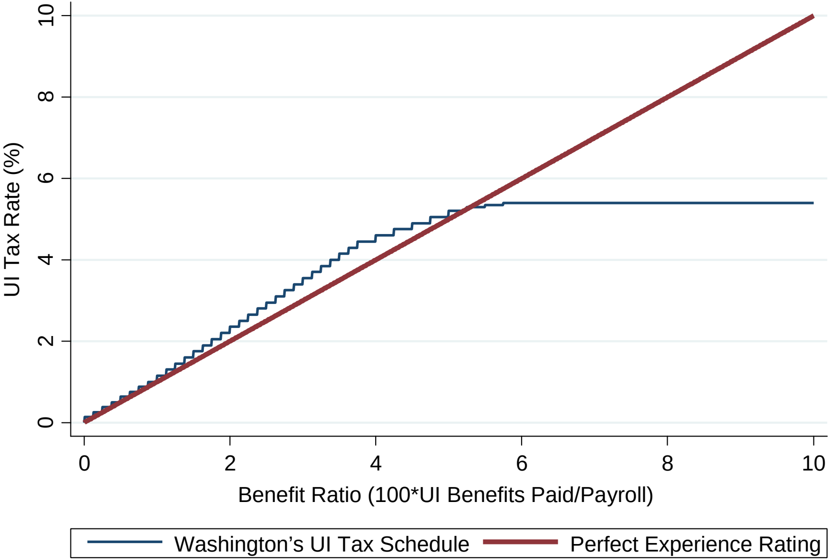


Source: Card, Chetty, Weber 2007b (AER P&P)

UI and Firm Behavior

- Preceding discussion assumed perfect experience rating of UI
  - Firms' layoff incentives are not distorted
- But in practice, UI is not perfectly experience rated
- Feldstein (1976, 1978) shows:
  - Theoretically that imperfect experience rating effect can raise rate of temporary layoffs
  - Empirically that this effect is large in practice

Experience Rating in Washington, 2005



Source: Washington State Joint Legislative Task Force on Unemployment Insurance Benefit Equity 2005

UI and Firm Behavior: Feldstein 1976 model

- Firms offer workers stochastic contracts, with wage and probability of temporary layoff
- Two states: high demand and low demand
- In equilibrium, competitive firms will offer contract that pays worker his marginal product in expectation over two states at cheapest cost to firm
- Firm profits by laying off workers with imperfect exp rating
- Layoffs generate first-order gain in profits at a second-order cost from added risk to worker
- In an imperfectly experience-rated economy, firms choose a positive rate of layoffs in low output state

Feldstein 1978: Empirical Results

- First observation: more than half of firms are above the max rate or below the min rate
  - No marginal incentive for these firms to reduce layoffs.
- Uses cross-state/time variation in UI benefits
- 10% increase in UI benefits causes a 7% increase in temp layoff unemployment
- Effect is twice as large for union members as non-union, suggesting worker-firm coordination.

Topel 1983

- Feldstein does not directly show that imperfect exp rating is to blame for more temp layoffs b/c not using variation in experience rating itself
- Topel (1983) uses state/industry variation in financing of UI
  - Variation in tax rate on firms from min/max thresholds for exp rating
  - Finds that imperfect subsidization accounts for 31% of all temp layoff unemployment, a very large effect
- See Krueger and Meyer (2002) for review of more recent studies, which find similar results but smaller magnitudes

UI Savings Accounts

- Alternative to UI transfer-based system (Feldstein and Altman 2007)
  - Instead of paying UI tax to government, pay into a UI savings account.
  - If unemployed, deplete this savings account according to current benefit schedule
  - If savings exhausted, government pays benefit as in current system (financed using a tax).
- Idea: people internalize loss of money from staying unemp longer.
  - Reduces distortion from UI while providing benefits as in current system.
  - But modelling this formally is difficult: to internalize incentives at retirement, people must be forward looking, but then no need to force them to save.

Feldstein and Altman 2007

- Address feasibility: How many people hit negative balance on UI account and just go back to old system?
- Simulate how UI savings accounts would evolve using actual earnings histories from PSID.
- Calculations imply that only 1/3 of spells will occur with negative balances, so most people still have good incentives while unemployed.
- Total tax payments are less than half what they are in current system.
- In their simulation, benefits are identical; only question is how costs change.

Feldstein and Altman 2007

- Calculation of changes in present value of lifetime wealth from switch to UISA by income quintile:

	Q1	Q2	Q3	Q4	Q5
Present Value Gain:	**-$95**	**+$22**	**-$67**	**+$94**	**+$468**

- Net PVG is positive
- Without change in behavior, how is the pie larger?
    - Reason: discounting at 2% but earning 5.5% interest

Takeup

- Mean takeup rate is very low – a major puzzle in this literature (Currie 2004)
  - Why leave money on the table?
- Andersen and Meyer (1997) show that after-tax UI replacement rate affects level of takeup.
  - So at least some seem to be optimizing at the margin.
- Takeup low in many govt. programs. (UI, food stamps, EITC, etc.)
- Possible explanations: myopia, stigma, hassle, lack of info.

Black, Smith, Berger, and Noel 2003

- Experiment in KY where some UI claimants were randomly assigned to receive re-employment services
  - E.g., assisted job search, employment counseling, job search workshops, retraining programs
  - Treatment [$N = 1236$] required to receive services in order to get UI benefits
  - Control [$N = 745$]: exempt from services

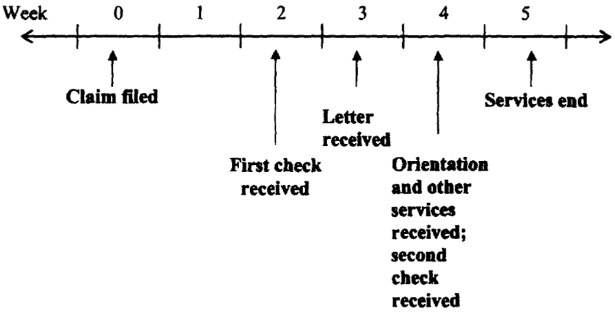


FIGURE 1. TIMELINE FOR TYPICAL UI CLAIMANT IN KENTUCKY WPRS PROGRAM

Source: Black, Smith, Berger, and Noel 2003

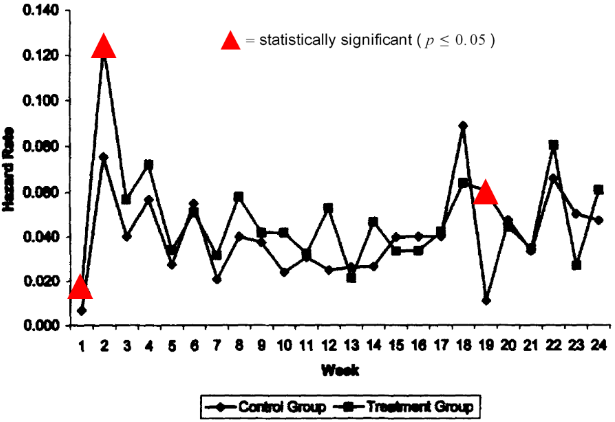


FIGURE 2. HAZARD FUNCTIONS OF THE TREATMENT AND CONTROL GROUPS, KENTUCKY WPRS EXPERIMENT, OCTOBER 1994 TO JUNE 1996

Source: Black, Smith, Berger, and Noel 2003

TABLE 2—IMPACT OF TREATMENT ON DURATION OF BENEFITS AND EARNINGS: KENTUCKY WPRS EXPERIMENT, OCTOBER 1994 TO JUNE 1996

Outcome measures	(1) Fixed-effect regression estimates	(2) Matching estimates
Number of weeks receiving UI benefits	−2.241	−2.045
	(0.509)	(0.411)
	[0.000]	[0.000]
UI benefits received	−143.18	−81.44
	(100.3)	(81.6)
	[0.077]	[0.159]
Fraction exhausting benefits	−0.024	−0.030
	(0.023)	(0.0019)
	[0.152]	[0.0055]
Earnings in the year after the start of the UI claim	1,054.32	1,599.99
	(588.0)	(475.2)
	[0.037]	[0.001]
N	1,981	1,981

Source: Black, Smith, Berger, and Noel 2003

Black, Smith, Berger, and Noel 2003: Results

- Treatment group exit UI system earlier, receiving 2.2 fewer weeks of benefits on average
- Most significant increase in exits in wks 2-3, when notified of mandatory services

General Equilibrium: Acemoglu and Shimer 1999

- UI can be efficiency-enhancing in equilibrium.
- Standard models focus only on distortionary costs, and assume that total output always lower when UI is provided.
- But this ignores potentially important GE effect: more risky jobs provided in eq. if workers are insured.
- Provision of UI raises availability of risky jobs (e.g. tech jobs) and can raise efficiency in equilibrium
- So if workers are risk averse, tradeoff may not be very hard – both raise output and insure them better.

Dynamics: Path of UI Benefits

- Classic reference is Shavell and Weiss (1979), who solved for optimal path of benefits in a 3 period model.
- Tradeoff: upward sloping path $\rightarrow$ more moral hazard but more consumption-smoothing benefits.
- Recent literature that is very active in this area: “new dynamic public finance” – optimal path of unemployment and disability programs.
  - Hopenhayn and Nicolini (1997) – numerical simulations for case where govt can control consumption
  - Shimer and Werning (2008) – with perfect liquidity and CARA utility, optimal benefit path is flat

Optimal Insurance in Behavioral Models

- We do not have a model consistent with the data that can explain both savings behavior pre-unemployment and search behavior post-unemployment
  - Evidence that unemployment is indeed costly and benefits can improve welfare a lot for certain liquidity-constrained groups
  - Simple rational model cannot rationalize level of savings that people have when they get unemployed
- Interesting direction for future research: optimal SI with behavioral considerations (see e.g., Spinnewijn 2009)

Workers Compensation

- Insurance against injury at work
- Covers both lost wages and medical benefits
- Rationales for govt. intervention:
  - Market may fail due to adverse selection
  - Workers may be unaware of risks on the job
  - Litigation costs (origin of system in 1920s)
- Substantial variation in benefits across states for different injuries

Maximum Indemnity Benefits (2003)

State	Type of permanent impairment					Temporary Injury
	Arm	Hand	Index finger	Leg	Foot	(10 weeks)
California	$108,445	$64,056	$4,440	$118,795	$49,256	$6,020
Hawaii	180,960	141,520	26,800	167,040	118,900	5,800
Illinois	301,323	190,838	40,176	276,213	155,684	10,044
Indiana	86,500	62,500	10,400	74,500	50,500	5,880
Michigan	175,657	140,395	24,814	140,395	105,786	6,530
Missouri	78,908	59,521	15,305	70,405	52,719	6,493
New Jersey	154,440	92,365	8,500	147,420	78,200	6,380
New York	124,800	97,600	18,400	115,200	82,000	4,000

Source: Gruber (2007)

Theory of Workers' Compensation

- Formally very similar to that of unemployment insurance
  - If prob of injury **cannot** be controlled, model same as Baily-Chetty
  - If prob of injury **can** be controlled, that distortion must be taken into account in calculation
  - Leisure now includes benefits of having more time to heal
- Similar formal theory, so literature is mostly empirical

Outline of Empirical Evidence

1. Monday effects and impact on worker behavior
2. Firm side responses
3. Effect on equilibrium wage

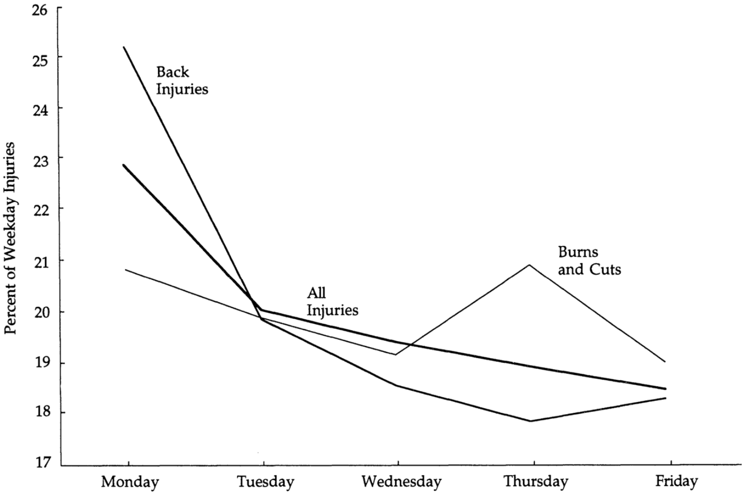


Figure 1. Distribution of Weekday Injuries.

Source: Card and McCall 1996

Day of the Week Effect

- Intertemporal distortions, moral hazard effect of workers' comp.
- Card & McCall (1994): test if weekend injuries lead to Monday effect.
  - Look at uninsured workers, who should have bigger Monday effect.
  - Find no difference in effect between insured and uninsured.
- Other explantations:
  - Gaming system for more days off.
  - Pure reporting effect if pain does not go away.
- Suggests that incentives matter a lot.

Effects of Benefits on Injuries

- Potential incentive effects to look for on worker's side:
  - Number of claims of injury
  - Duration of injuries
- Meyer, Viscusi, and Durbin (1995):
  - Implement DD analysis for workers' comp durations
  - Find large effects on duration using reforms in MI and KY

Variable	Kentucky			Michigan		
	Before increase (1)	After increase (2)	Percentage change (3)	Before increase (4)	After increase (5)	Percentage change (6)
Maximum benefit ($)	131.00	217.00	65.65	181.00	307.00	69.61
Replacement rate, high earnings (percent)	32.70 (0.25)	51.02 (0.37)	56.02 (1.65)	30.01 (0.35)	44.15 (0.48)	47.14 (2.33)
Replacement rate, low earnings (percent)	66.42 (0.20)	66.66 (0.22)	0.36 (0.44)	66.64 (0.24)	66.35 (0.30)	−0.45 (0.58)

Source: Meyer, Viscusi, Durbin 1995

TABLE 4—KENTUCKY AND MICHIGAN: DURATION AND MEDICAL COSTS OF TEMPORARY TOTAL DISABILITIES DURING THE YEARS BEFORE AND AFTER BENEFIT INCREASES

	High earnings		Low earnings		Differences		Difference in differences
Variable	Before increase (1)	After increase (2)	Before increase (3)	After increase (4)	[(2)−(1)] (5)	[(4)−(3)] (6)	[(5)−(6)] (7)
Mean duration (weeks)							
Kentucky	11.16	12.89	6.25	7.01	1.72	0.76	0.96
	(0.83)	(0.83)	(0.30)	(0.41)	(1.17)	(0.51)	(1.28)
Michigan	14.76	19.42	10.94	13.64	4.66	2.70	1.96
	(2.25)	(2.67)	(1.09)	(1.56)	(3.49)	(1.90)	(3.97)
Median duration (weeks)							
Kentucky	4.00	5.00	3.00	3.00	1.00	0.00	1.00
	(0.14)	(0.20)	(0.11)	(0.12)	(0.25)	(0.16)	(0.29)
Michigan	5.00	7.00	4.00	4.00	2.00	0.00	2.00
	(0.45)	(0.67)	(0.22)	(0.28)	(0.81)	(0.35)	(0.89)
Median medical cost (dollars)							
Kentucky	393.51	411.49	238.96	254.40	17.98	15.44	2.55
	(19.29)	(22.72)	(8.48)	(9.11)	(29.80)	(12.44)	(32.30)
Michigan	689.73	765.00	390.63	435.00	75.27	44.38	30.89
	(77.30)	(134.53)	(32.80)	(33.09)	(155.16)	(46.59)	(162.00)

Source: Meyer, Viscusi, Durbin 1995

Firm Side Responses

- Purchasing insurance leads to imperfect experience rating and moral hazard
- Self-insured firms: stronger incentives to improve safety
  - Also, have incentive to ensure that workers return to work quickly
- Krueger (1990): compares behavior of self-insured firms with others
  - Finds self-insured have 10% shorter durations, but selection bias a concern

Effect on Equilibrium Wage

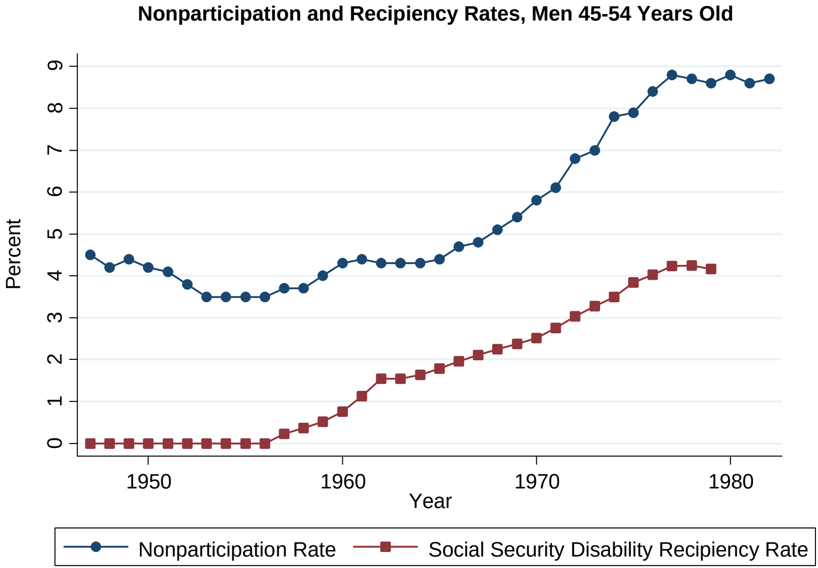
- Workers' compensation is a mandated benefit
  - When firms hire, adjust wage offered to workers downwards b/c they realize they must pay benefit
- Summers (1989):
  - If workers value benefits at cost, they bear the full incidence
  - If they do not value it, has same effect and DWL as a tax
- Gruber-Krueger (1991):
  - 85% of WC cost is shifted to workers, no significant employment effect
- Fishback-Kantor (1995):
  - Find 100% shift to workers' wages in initial implementation of prog
  - Suggests that benefits valued close to cost

Directions for Further Research on WC

- Decomposition into liquidity vs. moral hazard effects
- Better evidence on firm side responses
- Consumption smoothing benefits

Disability Insurance

- See Bound et. al (HLE 1999) for an overview
- Insures against long-term shocks that affect individuals at home or work
- Federal program that is part of social security
- Eligible if unable to "engage in substantial gainful activity" b/c of physical/mental impairment for at least one (expected) year
- Main focus of literature is sharp rise in the size of the program



Source: Parsons 1984 Table A1

Two Views on the Rise in DI

- Trend has continued since 1980s: DI share of non-elderly adults rose from 3.1% in 1984 to 5.4% in 2000
- One perspective on the rise: moral hazard from a lenient system that leads to inefficiency
- Another perspective: program is now helping more needy people who have high disutilities of work
- Empirical work attempts to distentangle these two views to some extent

TABLE 1—REASSESSMENTS OF INITIAL SOCIAL SECURITY DETERMINATIONS

A. *Bureau of Disability Insurance Review One Year After Initial Determination (Percentages):*

	Initial determination	
BDI assessment	Allowance	Denial
Allowance	78.8	21.1
Denial	22.5	77.5

Note: The sample sizes are 250 initial allowances and 248 initial denials.
Source: Smith and Lilienfeld (1971 p. 195).

Theory of Disability Insurance

- Key additional element relative to UI models is screening and waiting periods.
- Less relevant for unemployment because it is easy to identify who has a job and who does not.
- Diamond-Sheshinski (1995) build a model that incorporates screening.
- Characterize optimal properties of solution but do not derive an empirically implementable formula for optimal screening rule or benefit level.

Diamond and Sheshinski 1995

- Individuals have different disutilities of working ψ_i
- To max social welfare, not desirable for those with high ψ_i to work.
- First best: Individual i works iff

$$\text{Marginal product} > \psi_i$$

- But govt observes only an imperfect signal of $\psi_i \rightarrow$ sets a higher threshold for disability
- Result: lower benefit rate if screening mechanism has higher noise to signal ratio

Empirical Evidence: Bound-Parsons Debate

- Question: Did increase in DI benefits cause decline in labor supply?
- Well-known debate between Bound & Parsons in 1980s
- Parsons (1980)
  - Uses cross-sectional variation in replacement rates
  - Data on men aged 45-59 in 1966-69 NLSY
  - OLS regression:
    $$LFP_i = \alpha + \beta \text{DIreprate}_i + \varepsilon_i$$
    where DIreprate is calculated using wage in 1966
  - Finds elasticity of 0.6
  - Simulations using this elasticity imply that increase in DI can completely explain decline in elderly labor force participation

Empirical Evidence: Bound-Parsons Debate

- Criticizes Parsons for using an endogenous variable on RHS
  - Econometric problem: DIreprate $= f(wage(-); law)$ with no variation in law
  - Identification assumption: LFP rates equal across wage groups
  - Potential solution: "control" for wage on RHS. Does not make sense.
- Bound replicates Parson's regression on sample that never applied to DI and obtains a similar elasticity

Empirical Evidence: Bound-Parsons Debate

- Bound proposes a technique to bound effect of DI on LFP rate
- Uses data on LFP of *rejected* applicants as a counterfactual
- Idea: if rejected applicants do not work, then surely DI recipients would not have worked
  - Rejected applicants' LFP rate is an upper bound for LFP rate of DI recipients absent DI
- Results: Only 30% of rejected applicants return to work
  - Earn less than half of the mean non-DI wage
- Implies that at most 1/3 of the trend in male LFP decline can be explained by shift to DI

TABLE 2—EMPLOYMENT, EARNINGS, AND OTHER CHARACTERISTICS OF REJECTED DISABILITY INSURANCE APPLICANTS

	1972			1978		
	Population	Rejected Applicants	Beneficiaries	Population	Rejected Applicants	Beneficiaries
Labor Supply						
Percent Employed	77.7	32.6	3.2	69.3	28.7	2.3
Percent Worked 71/77	91.9	45.0	7.5	86.7	40.4	5.5
Percent Full Year ($\geq$ 50 Weeks)[a]	76.8	47.4	31.4	83.5	41.2	22.2
Percent Full Time ($\geq$ 35 Hours)[a]	95.4	75.9	25.0	92.4	79.6	38.3
Earnings Among Positive Earners						
Median Annual Earnings, 71/77[b]	\$9000	\$4000	\$700	\$14000	\$5300	\$1000

Source: Bound 1991

Gruber 2000

- Exploits differential law change in Quebec and rest of Canada as a natural experiment
- In 1987, 36% inc. in benefits in rest of Canada; in Quebec, no change
- Estimates effect of law change on labor force participation of men aged 45-59
- Uses DD method on NLFP rates of men aged 45-59

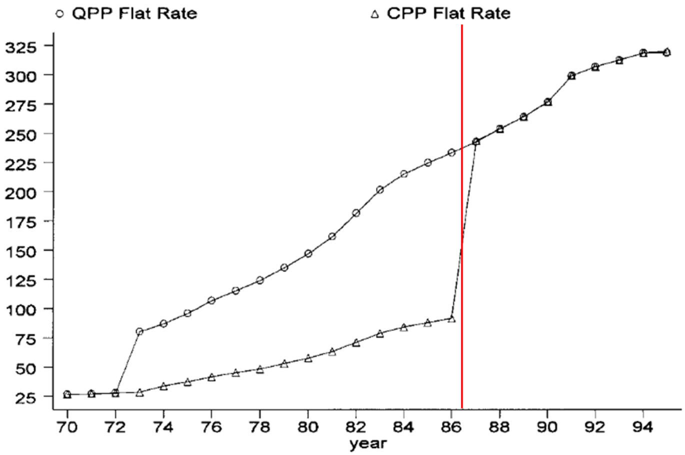


FIG. 1.—Flat-rate portion in Quebec and the rest of Canada

Source: Gruber 2000

TABLE 1
MEANS

	CPP		QPP		DIFERENCE IN DIFFERENCE
	Before (1)	After (2)	Before (3)	After (4)	(5)
Benefits	5,134	7,776	6,878	7,852	1,668 (17)
Replacement rate	.245	.328	.336	.331	.088 (.003)
Not employed last week	.200	.217	.256	.246	.027 (.013)
Married?	.856	.856	.817	.841	−.024
Any kids < 17?	.367	.351	.354	.336	.002
Less than 9 years of education	.303	.274	.454	.421	.004

Source: Gruber 2000

Gruber 2000

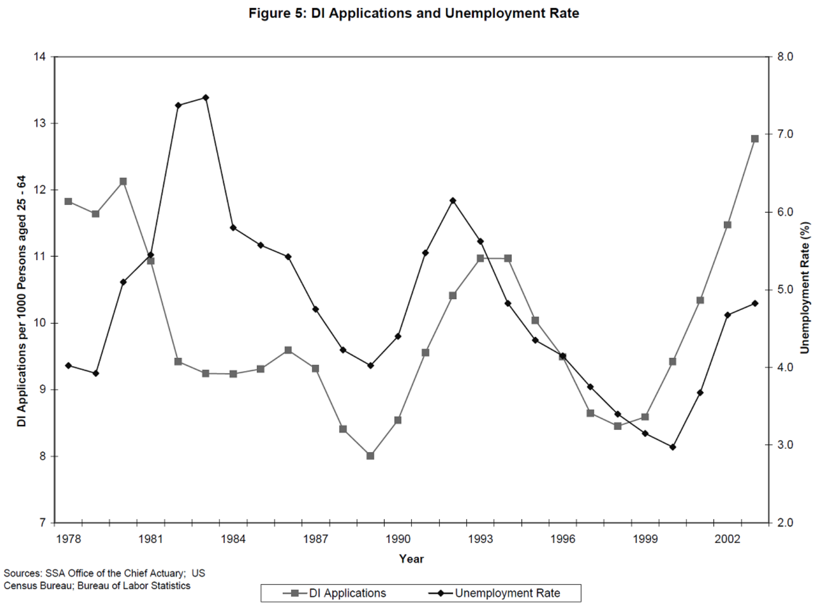
- Implied elasticity of NLFP rate w.r.t. DI benefit level: 0.25-0.3
- Agrees more with Bound than Parsons
- Limitation of Gruber study (like all DD studies): only estimates short run response

Autor and Duggan 2003

- Focus on interaction between DI and UI systems
- Observe that DI claims rise in recessions, may reduce measured unemployment rate
- Idea: consider a worker laid off in current recession
  - Given generosity of DI program, instead of claiming UI and searching for a job, he applies for DI
  - One less unemployed person $->$ unemployment rate lower
- But economic situation is the same: one less person working
- Test this hypothesis using cross-state variation in employment shocks

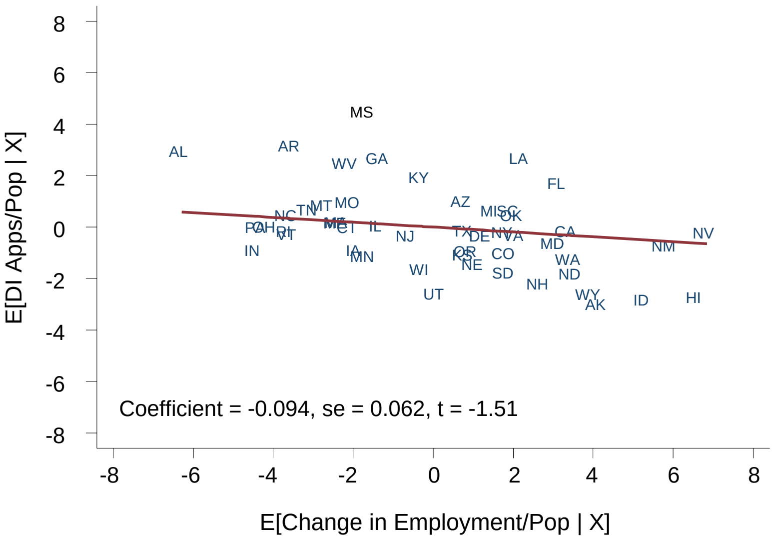
Autor and Duggan 2003

- Construction of state-level employment shocks over a five year window:
  - Calculate industry shares in a given state in base year
  - Calculate employment changes over five year period by industry using data on national employment (excluding state in question)
  - Project changes in each state's employment using national changes
  - Ex: if car industry declines over a five year period, assign a negative employment shock to Michigan
- Then correlate state employment shocks with DI applications



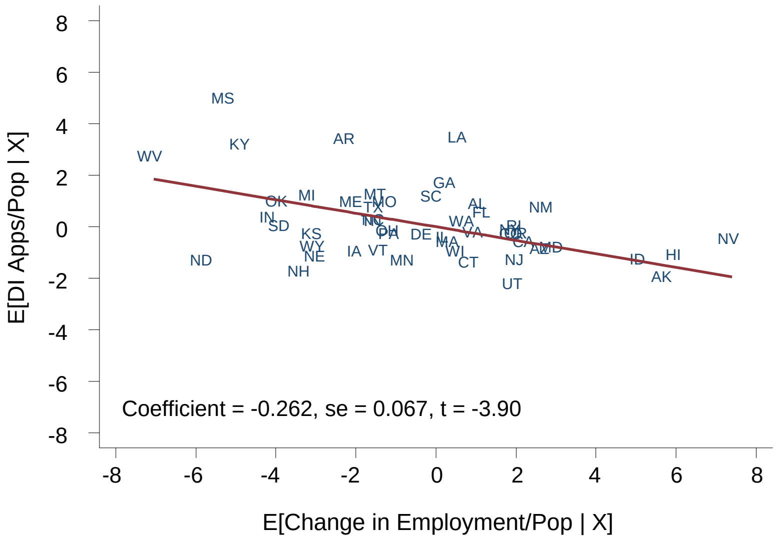
Source: Duggan and Imberman 2005

Employment Shocks and DI Applications: 1979-1984



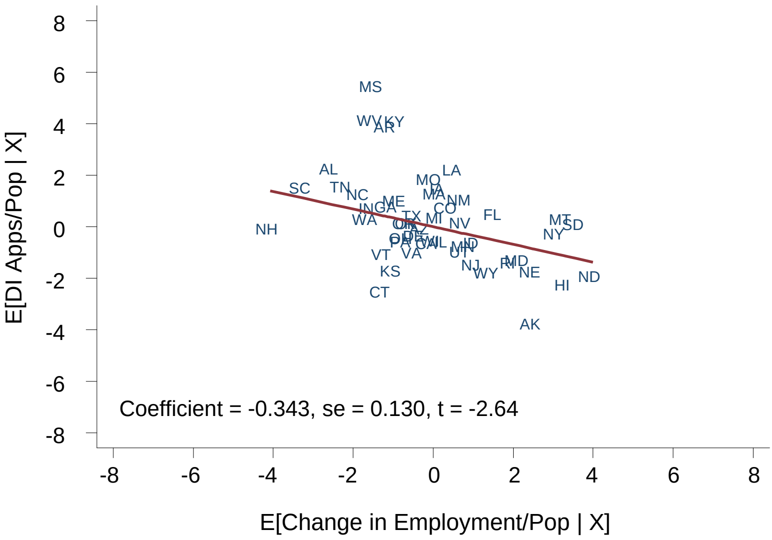
Source: Autor and Duggan 2003

Employment Shocks and DI Applications: 1984-1989



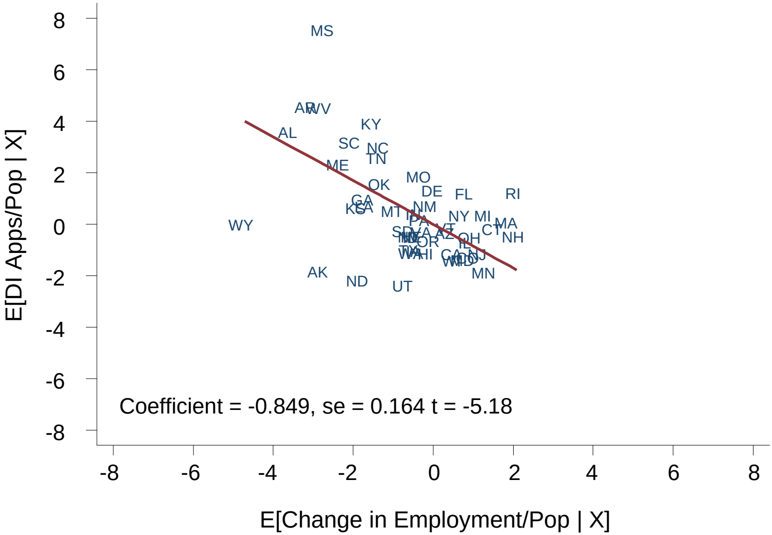
Source: Autor and Duggan 2003

Employment Shocks and DI Applications: 1989-1994



Source: Autor and Duggan 2003

Employment Shocks and DI Applications: 1993-1998



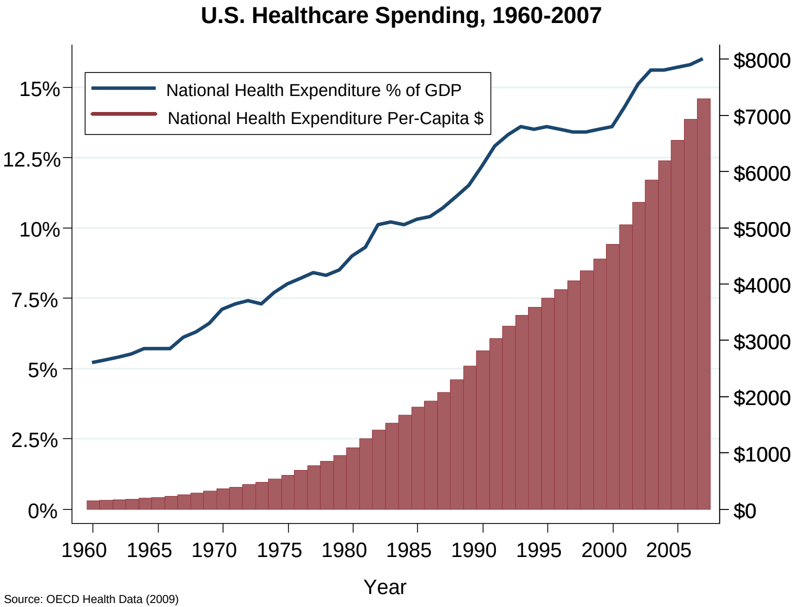
Source: Autor and Duggan 2003

Autor and Duggan 2003

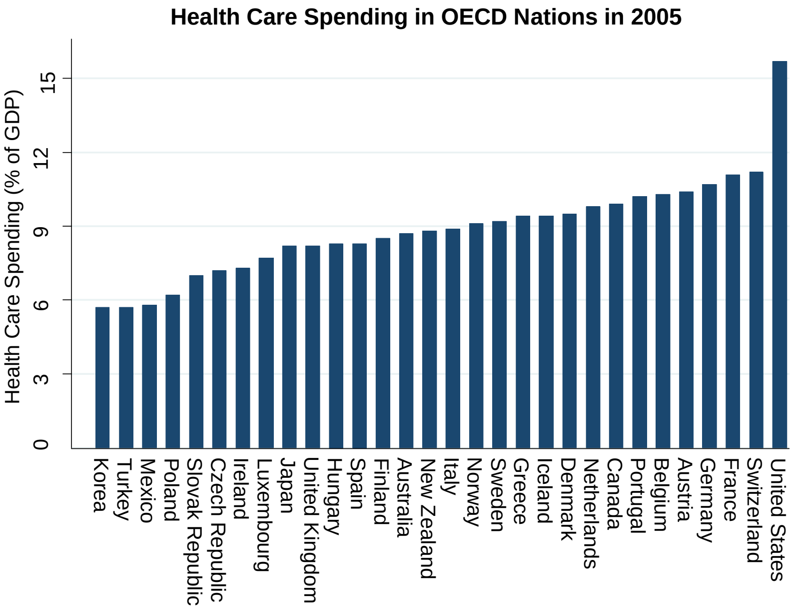
- Unemployment would be 0.65% higher if not for post-‘84 trends in DI participation
- Trace decline in LFP to the rise in DI over the past two decades via:
  - The 1984 inclusion of mental illness in DI eligibility
  - Rising wage inequality (combined with the progressively of system)
- Bottom line: DI applications are clearly sensitive to incentives
  - But evidence is insufficient to make welfare statements
  - Essential to decompose benefit effects into income and price elasticities to make normative judgment

Health Insurance

- Arrow (1963): seminal article that described special problems in providing healthcare using private markets
- We will touch upon a few issues in public sector intervention
- Health is an important field because of enormous size and rapid growth.
    - 17% of GDP
    - Annual growth rate of 3.4% (vs 1.4% growth in GDP)

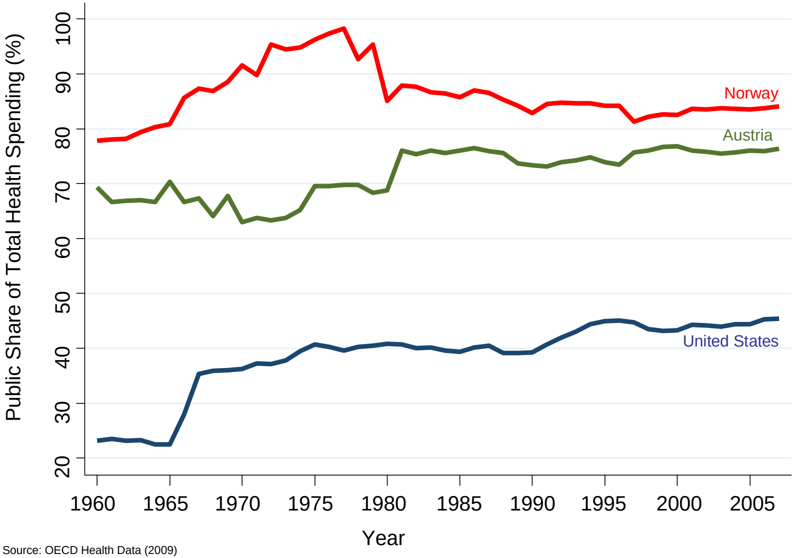


Source: OECD Health Data (2009)



Source: OECD Health Data (2009)

Public Health Share in OECD Nations, 1960-2007



Source: OECD Health Data (2009)

Americans' Source of Health Insurance Coverage, 2002

	People (millions)	% of Population
Total population	288.6	100.00%
Private	177.8	61.60%
• Employment-based	161	55.80%
• Individually purchased	16.8	5.80%
Public	83	28.80%
• Medicare	40.5	14.00%
• Medicaid	42.8	14.80%
• Veterans	6.9	2.40%
Uninsured	43.3	15.00%

Note: Numbers do not sum to 100% because some people have multiple coverage. Source: Gruber 2007

Growing Health Expenditures: Key Factors

1. **Fundamentals of supply and demand [market equilibrium]**

- Demand: Income effect → more demand (Hall and Jones 2006)
  - As you get richer, want to live longer, not consume more goods because marginal utility of consumption declines
  - More sushi dinners, not more sushi per dinner
- Supply: technological progress with more expensive methods
  - Two options for knee surgery: invasive, long recovery [old] vs. arthroscopic [new]. New technology more expensive.
  - LASIK surgery: expensive but allows you to completely eliminate need for glasses
  - Note difference relative to technological progress in other sectors: discovery of more expensive methods rather than reduction in costs of existing methods

Growing Health Expenditures: Key Factors

2. Price Distortions

- Demand: government tax subsidy for healthcare and insurance programs
  - Lower effective price for individuals $\rightarrow$ overconsumption
- Supply: fee-for-service payment schemes
  - Reimburse physicians for additional procedures $\rightarrow$ overproduction

Growing Health Expenditures: Key Factors

3. Regulatory Distortions

- Supply of healthcare: malpractice law
  - Fear of lawsuits $\rightarrow$ excess supply of healthcare by physicians
- Supply of physicians
  - Restrictions on number of physicians through medical school seats/licensing
  - American Medical Association acts like a union
  - Lower supply of physicians $\rightarrow$ higher wages and higher input costs

Market Failures and Government Interventions

1. Externalities/Internalities
   - Sin taxes (alcohol/cigarettes)
   - Rabin and O'Donoghue (2006): fat tax
2. Consumer myopia
   - Tax subsidies for health insurance
   - Samaritan's Dilemma: government provided insurance
3. Consumers lack information $\rightarrow$ suppliers choose level of consumption
   - Govt. provision of healthcare + fixed physician salaries
   - Regulation: licensing of doctors, FDA, legal system

Market Failures and Government Interventions

4. Heterogeneity of risk types → adverse selection in insurance market

5. Ex-ante risk uninsured: cannot contract before birth

6. Equity concerns: health inequality may directly enter social welfare function

    - Example: White infant mortality rate is 6 per 1000; black is 14 per 1000.

    - Black child born in DC has lower chance of reaching first birthday than one born in Jamaica.

- Solution: government provided health insurance/healthcare

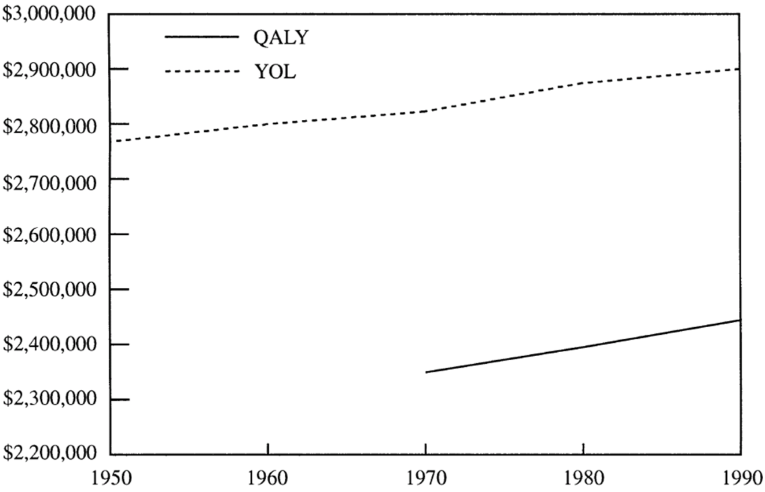
Measuring Health

- Before discussing optimal insurance, useful to define a measure of health consumption
- Higher medical expenditure not equivalent to more "health."
- Starting point: mortality.
- Need a monetary measure $\longrightarrow$ measure value of life.
- Literature estimates this using many methods (Aldy and Viscusi 2003)
  - Contingent valuation.
  - Wage premia for risky jobs.
  - Price of smoke detectors.
- Commonly used figure: \$100,000 per year of healthy life.

Cutler and Richardson 1997

- Propose a better definition of value of life that takes quality of life into account
- Measure QALY for several conditions using survey
  - What is your quality of life relative to that of a perfectly healthy person?
- Define a person's "health capital" as present value of expected QALYs times $100K
- This can be computed at various ages
- Can be used to assess which policies/interventions improve health capital the most

Health Capital at Birth, 1950-1990



Source: Cutler and Richardson 1997

Cutler and Richardson 1997

- Dramatic change in health capital over the past century from two channels
- Mortality rate declined by 66 percent
  - Largly due to improvements in infant mortality, treatment of cardiovascular disease.
- Improvements in morbidity as well, but some declines because people live longer
  - E.g. cancer more prevalent even though progress has been made in fighting cancer (Honore and Lleras-Muney 2007)
- Overall, health capital has increased by \$100K-\$200K from 1970-1990 (about 10%).
- Far outpaces growth in expected medical spending (growth of less than \$50K).

Optimal Govt. Intervention in Health Insurance

- Now consider optimal design of government health insurance policies
- Differences relative to other social insurance programs:
  1. Importance of provider side incentives.
  2. Interaction between private and public insurance (crowdout).
- Begin with a pure demand side model and then consider supply side.

Demand for Medical Care: Feldstein 1973

- Price of medical care is 1, total wealth of consumer is y
- $s =$ smooth index of disease severity
- $m =$ amount of medical care purchased
- $c(m) =$ patient's co-payment as a function of m
- $\pi =$ insurance premium

Feldstein 1973

- $x =$ non-medical consumption
- $H(s, m) =$ health as a fn. of disease state and medical care.
- Assume H is concave in m
- Let $U(x, H) =$ utility over the two goods

Feldstein 1973

- Insurer sets premium to cover costs in expectation:

$$\pi = \int [m(s) - c(m(s))] f(s) ds$$

- Individual chooses level of medical care by maximizing utility, taking π as given

$$\max_{m(s)} \int [U(y - \pi - c(m(s)), H(s, m(s))] f(s) ds$$

- At an interior solution, individual will set $\forall s :$

$$H_m = c'(m) \frac{U_X}{U_H}$$

Feldstein 1973: First Best Solution

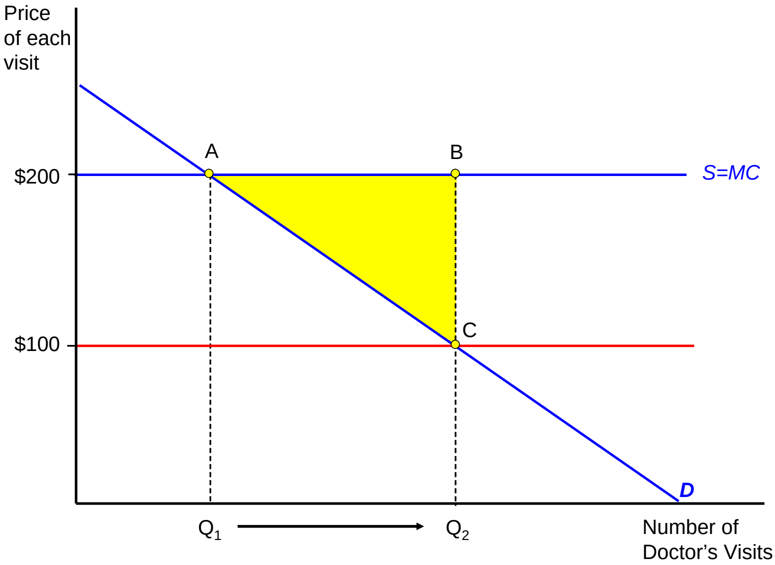
- Individual internalizes costs to insurer, so choose m based on $c'(m) = 1$:

$$H_m(m) = \frac{U_X}{U_H}$$

- Optimal copayment is zero in all states
- Note: this assumes that marginal utility of consumption is indepenent of health state
- In general case, optimal to set $MU^{sick} = MU^{healthy}$, in which case copayment may be desirable.

Feldstein 1973: Second Best

- In second best, individual only internalizes copayment
- Consumes more medical care, because $c'(m) < 1$ and H is concave
- Resulting deadweight loss from insurance is analogous to that caused by overconsumption of a good because of a subsidy.
- Optimal copay rate can be determined using tools analogous to that in optimal UI model
- Tradeoff between risk and moral hazard


Price
of each
visit
$200
$100
A
B
C
S=MC
D
Q1
Q2
Number of
Doctor's Visits

Empirical Evidence: Moral Hazard in Health Insurance

- Feldstein (1973): used cross-state variation to estimate an elasticity of demand for medical care w.r.t price of 0.5.
- Rich subsequent literature has yielded a variety of estimates.
- Manning et al (1987): gold standard estimate based on $136 million RAND experiment
  - Total sample: 6000.
  - Randomly assigned into 14 different ins. plans that varied in copay rate
  - Copay rate: was 0, 25, 50, or 95.
  - Tracked on average over 3 years, with full details on medical expenses.
  - Elasticity of about 0.1 for inpatient care, 0.2 for outpatient care

RAND Health Insurance Experiment

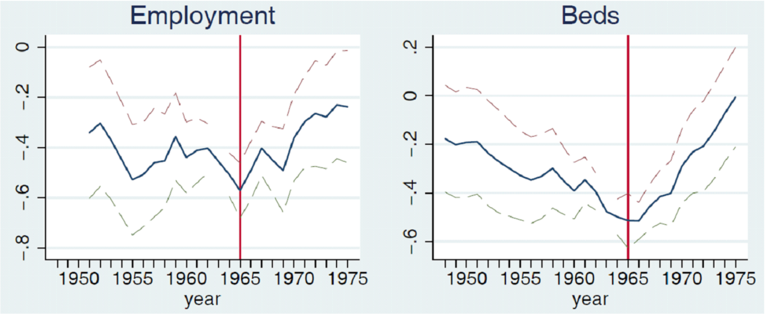
Plan	Face-to-Face Visits	Outpatient Expenses (1984 $)	Admis-sions	Inpatient Dollars (1984 $)	Prob. Any Medical (%)
Free	4.55	340	.128	409	86.8
	(.168)	(10.9)	(.0070)	(32.0)	(.817)
25 Percent	3.33	260	.105	373	78.8
	(.190)	(14.70)	(.0090)	(43.1)	(1.38)
50 Percent	3.03	224	.092	450	77.2
	(.221)	(16.8)	(.0116)	(139)	(2.26)
95 Percent	2.73	203	.099	315	67.7
	(.177)	(12.0)	(.0078)	(36.7)	(1.76)

Source: Manning et al. 1987

Finkelstein 2006

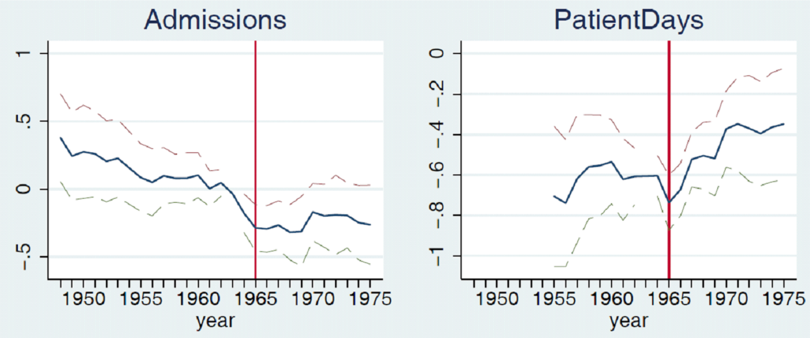
- General equilibrium effects may lead to much larger elasticities of consumption with respect to health insurance in equilibrium
    - Market-wide changes in demand alter hospitals' practice styles and technology
- Examines 1965 introduction of Medicare
- Identification strategy: geographic variation in ins. coverage prior to 1965
- In northeast, 50% of elderly were insured, in south, 12% were insured

$$\log(y_{ijt}) = \alpha_j * \mathbf{1}(county_j) + \delta_t * \mathbf{1}(Year_t) + \sum_{t=1948}^{t=1975} \lambda_t (Mcareimpact_z) * \mathbf{1}(Year_t) + X_{st}\beta + \varepsilon_{ijt} .$$



Source: Finkelstein 2006

$$\log(y_{ijt}) = \alpha_j * \mathbf{1}(county_j) + \delta_t * \mathbf{1}(Year_t) + \sum_{t=1948}^{t=1975} \lambda_t (Mcareimpact_z) * \mathbf{1}(Year_t) + X_{st}\beta + \varepsilon_{ijt} .$$



Source: Finkelstein 2006

Finkelstein 2006

- Impact of Medicare on hospital spending is six times larger than predicted by individual-level changes in RAND experiment
- Estimates imply that increased health insurance can explain half of increase in health spending between 1950 and 1990
- No direct normative implications: could be a liquidity or moral hazard effect.

Implications of Demand Side Model

- Optimal insurance structure: deductible coupled with lower copay as shocks become large
- Many policies look like this but not Medicare Part D
  - 0% of the drug costs up to $250
  - 75% of the costs for the next $2,250
  - 0% of the costs for the next $3,600
  - 95% of the costs above $5,100

Ellis and McGuire 1986

- Previous analysis assumed a passive doctor.
- In practice, physicians rather than patients likely to choose m
- When physicians choose level of m, physician compensation scheme determines efficiency of m
- High copayments for patients may not solve the problem
- Anecdotal evidence: dentists pulling out excess wisdom teeth

Ellis and McGuire 1986: Setup

- Goal: contrast efficiency of payment systems for physicians and analyze optimal system

- Payment for physician services is

$$P = \alpha + \beta c$$

$\alpha =$fixed cost payment for practice

$\beta =$payment for proportional costs (tests, nurses)

Ellis and McGuire: Compensation Schemes

- Various methods of payment (α, β):

1. **Fee-for-service** $[\alpha = 0, \beta > 1]$: No fixed payment for practice, but insurance company pays full cost of all visits to doctor + a surcharge.
2. **Salary** $[\alpha > 0, \beta = 1]$: practice costs paid for as well as marginal costs of treatment.
3. **Capitation** $[\alpha > 0, \beta = 0]$: varying by type and # of patient but not services rendered

Ellis and McGuire: Compensation Schemes

- General trend has been toward higher α, lower β
- Private market has shifted from FFS to HMO capitation schemes
- Medicare/Medicaid shifted in '80s to a prospective payment scheme
- Tradeoff: lower β provides incentives for doctors to provide less services. But they may provide too little!
- Lower costs, but complaints of lower quality of care

Ellis and McGuire: Optimal Payment Scheme

- To characterize optimal payment scheme, need to specify how physician chooses quality of care
- Physician's utility function:

$$U = \theta\pi + (1-\theta)q$$

$\pi =$profits earned by physician

$q =$quality of care $=$ benefit to patient.

- With payment scheme (α, β), profits are

$$\pi = \alpha + \beta c(q) - c(q)$$

Ellis and McGuire: Optimal Payment Scheme

- Doctors solve

$$\max_q \theta(\alpha + \beta c(q) - c(q)) + (1-\theta)q$$

- Society's problem is to maximize quality of care net of costs

$$\max_q q - c(q)$$

- Socially optimal quality level: q^* such that

$$c'(q^*) = 1$$

Ellis and McGuire: Optimal Payment Scheme

- The level of care q^D provided by doctor is such that:

$$\begin{aligned} \frac{dU}{dq} &= \theta(\beta-1)c'(q^D)-\theta+1=0 \\ \Rightarrow c'(q^D) &= \frac{1-\theta}{\theta(1-\beta)} \end{aligned}$$

- So, in order to get the doctor to choose the social optimum, need to set β such that $q^D = q^*$.

$$\begin{aligned} 1 &= c'(q^*) = c'(q^D) = \frac{1-\theta}{\theta(1-\beta)} \\ \Rightarrow & \quad \beta^* = 2-\frac{1}{\theta} \end{aligned}$$

Ellis and McGuire: Optimal Payment Scheme

$$\beta^* = 2 - \frac{1}{\theta}$$

- Optimal degree of incentive pay is increasing in θ.
- Intuition: if doctor is selfish (high θ), reimburse him for costs of provision so that he doesn't under-provide service to patients.
- But if he is benevolent, reduce the amount he gets paid for provision.
- He will naturally get benefits from taking care of them and will over-provide if he is paid for it too.
- HMOs desirable if healthcare providers are benevolent; FFS reimbursement if they are profit-seeking.

Ellis and McGuire Model: Limitations

- Ignores cream-skimming by doctors if they must bear costs
- Doctors assumed to be risk neutral
- Static model: ignores technological change and incentives to innovate
  - Finkelstein (2004): policies intended to change utilization of vaccines led to more innovation, some of which may have been unproductive
- Would be useful to derive an empirically implementable formula for β^*
  - Ex: use doctors' treatment of themselves or kids/relatives

Empirical Evidence on Supply Side Incentives

- Cutler (1995) examines mortality and readmission outcomes around 1983 Medicare reform.
  - Finds an effect on timing of death, but no effect in long run.
  - Suggests that physicians were practicing "flat of the curve" medicine.
  - Physicians may be benevolent enough that a capitation scheme is optimal.

Crowdout of Private Insurance

- So far have assumed a single insurer. In practice, both private and public ins. coexist.
- To what extent does crowdout of other insurance mechanisms diminish benefits of government intervention?
- Cutler and Gruber (1996): Medicaid crowdout
  - Medicaid expansions to pregnant mothers different across states
  - 50% of added Medicaid enrollment came from dropping private health ins. coverage through employer.
- Chetty and Saez (2008): optimal insurance with crowdout of private sector insurance contracts

Currie and Gruber 1995: Benefits of Public Insurance

Medicaid Eligibility Changes

(1)	(2)	(3)
A: Variation by State: Eligibility for Children		
Year	Missouri eligibility	Michigan eligibility
1982	12%	20%
2000	76%	34%
B. Variation by age: Eligibility in Washington D.C.		
Year	Age 14 eligibility	Age 0 eligibility
1982	18%	48%
2000	59%	56%

Source: Gruber 2007

Currie and Gruber 1995: Benefits of Public Insurance

- 30 pp increase in Medicaid eligibility among 15-44 year old moms has two effects
  - Greater utilization: early prenatal care visits rose by more than 50%.
  - Better health outcomes: infant mortality declined by 8.5% due to the expansions in Medicaid for pregnant women.
- Beneficial effects large because this is likely to be an underinsured, underserved population

Costs Per Life Saved of Various Regulations

Regulation concerning …	Year	Agency	Cost per life saved ($ millions)
• Childproof lighters	1993	CPSC	$0.10
• Food labeling	1993	FDA	0.4
• Reflective devices for heavy trucks	1999	NHTSA	0.9
Medicaid pregnancy expansions	1996	Currie and Gruber	1
• Children's sleepware flammability	1973	CPSC	2.2
• Rear/up/shoulder seatbelts in cars	1989	NHTSA	4.4
• Asbestos	1972	OSHA	5.5
Value of statistical life			7
• Benezene	1987	OSHA	22
• Asbestos ban	1989	EPA	78
• Cattle feed	1979	FDA	170
• Solid waste disposal facilities	1991	EPA	100,000

Source: Gruber 2007

Public Economics Lectures
Part 7: Public Goods and Externalities

Raj Chetty and Gregory A. Bruich

Harvard University
Fall 2009

Public Goods: Outline

1. Definitions and Baseline Model
2. Samuelson Rule
3. Lindahl Pricing
4. Social Choice: Median Voter Theorem
5. Public Goods with Endogenous Private Provision
6. Public Goods with Distortionary Taxation
7. Charity and Private Provision

Public vs. Private Goods

- **Private goods** benefit one individual h

$$\sum_h X_h \leq X$$

- **Public goods** benefit several individuals simultaneously

$$X_h \leq X \qquad \forall h$$

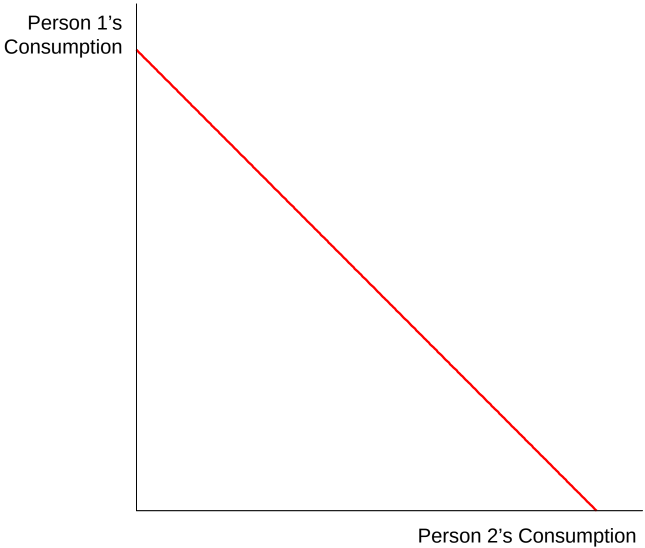
  - Ex: can of coke vs. teaching a class

- **Pure**: can accommodate any number of users.

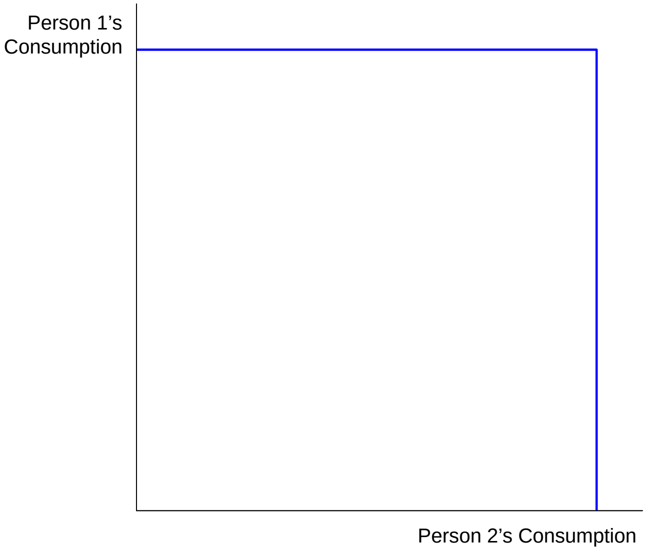
- **Impure:** subject to congestion

  - radio vs. roads

Private Good



Public Good



Public vs. Private Goods

- **Rival** vs. **non-rival.**
  - Pure are non-rival
- **Excludable** vs. **non-excludable**.
  - National Radio: impossible to exclude. Teaching: possible to exclude
- Most economic analysis focuses on pure public goods
- Public goods $\Rightarrow$ equilibrium outcome inefficient (large scale production externalities)

Public Goods Model: Setup

- Economy with H households, indexed by $h = 1, .., H$
- Two goods X and G
- X is always private, individual h consumes quantity X^h
- Denote by $X = \sum_h X^h$ the total quantity of good X in the economy
- Denote by G^h consumption of good G by h, with $G = \sum_h G^h$
- Utility of h is $U^h = U^h(X^h, G^h)$

Public Goods Model: Setup

- Social welfare = weighted sum of utilities, β^h weight on h
  - $\beta^h \geq 0$ and at least one $\beta^h > 0$
- Production possibility $F(X, G) = 0$
- Assume that U^h is increasing in X and G

First Best if G is Private

- To identify Pareto efficient outcomes, solve:

$$\max \sum_h \beta^h U^h(X^h, G^h)$$
$$\text{s.t. } F(\sum_h X^h, \sum_h G^h) \leq 0 \ [\lambda]$$

- Equivalent to $\max U^1$ s.t. $U^h \geq U_0^h$ for all $h \geq 0$ and $F \leq 0$. Lagrangian:

$$L = \sum \beta^h U^h - \lambda F$$

- First order conditions

$$[X^h] \ : \ \beta^h U_X^h = \lambda F_X$$
$$[G^h] \ : \ \beta^h U_G^h = \lambda F_G$$

First Best if G is Private

- Taking ratios of FOCs yields

$$\frac{U_G^h}{U_X^h} = \frac{F_G}{F_X}$$

- Set of Pareto efficient allocations is set of allocations that satisfy:

$$MRS_{GX}^h = MRT_{GX} \; \forall h$$

- Decentralized market equilibrium will implement such an allocation (1st Welfare Thm).

First Best if G is a Pure Public Good

- Let G denote level of PG, which everyone consumes
- Utility of h is $U^h = U^h(X^h, G)$
- Production possibility $F(X, G) = 0$ as before

First Best if G is a Pure Public Good

- To identify Pareto efficient outcomes,

$$\max \sum_h \beta^h U^h(X^h, G)$$
$$\text{s.t. } F(\sum_h X^h, \sum_h G^h) \leq 0 \; [\lambda]$$

- FOC's:

$$[X^h] \quad : \quad \beta^h U_X^h = \lambda F_X$$
$$[G] \quad : \quad \sum_h \beta^h U_G^h = \lambda F_G$$

- Using $\beta^h = \lambda F_X / U_X^h$ from f.o.c. for X^h we obtain:

$$\sum_h [\frac{U_G^h}{U_X^h}] = \frac{F_G}{F_X}$$

Samuelson (1954) Rule

- Condition for Pareto efficiency: **sum of** MRS is equal to MRT:

$$\sum_h MRS^h_{GX} = MRT_{GX}$$

- Intuition: an additional unit of G increases the utility of all households in the public good case
- With G a private good, an additional unit only increases one individual's utility

Samuelson (1954) Rule

- Excludability plays no role in the analysis.
  - Only relevant for determining feasible provision mechanisms
- Samuelson rule simple but difficult to implement in practice.
  - Govt needs to know preferences
  - Issue of how to finance the public good
- Samuelson analysis is a first-best benchmark
- How can optimal level of PG be implemented with available policy tools?

Model of Private Provision: Setup

- Private good X and a pure public good G.
- Price of each good is normalized to one (one-to-one transformation technology).
- Each household starts with an endowment Y^h of good X.
- Individual h contributes G^h to public good funding.
- Consumption of public good is $G = \sum_h G^h$ for everyone.
- Consumption of the private good is $X^h = Y^h - G^h$ for individual h.

Decentralized Private Provision Suboptimal

- Individual h solves

$$\max U^h(X^h, G^1 + .. + G^h + .. + G^H)$$
$$\text{s.t. } X^h + G^h = Y^h.$$

- Nash equilibrium outcome is $U_X^h = U_G^h$
- Samuelson Rule not satisfied
- Pareto improvement if each person invested $1/H$ more dollars in the public good:

$$\Delta W = -U_X^h(1/H) + U_G^h = U_G^h(1 - 1/H) > 0.$$

- Market outcome is inefficient; underprovision of G

Lindahl Equilibrium

- How to achieve Pareto efficiency through a decentralized mechanism?
- Suppose individual h has to pay a share τ^h of the public good and can pick a level of G
- Individual h chooses G to maximize

$$U^h(Y^h - \tau^h G, G)$$

- FOC: $\tau^h U_X^h = U_G^h$.
- Demand function of $G^h = G^h(\tau^h, Y^h)$

Lindahl Equilibria: Conditions

- A Lindahl Equilibrium satisfies the following two conditions:

  1. Public good must be fully financed.

$$\sum_h \tau^h = 1$$

  2. All individuals must demand same quantity of G.

- Lindahl equilibrium generically exists: H equations ($G^1 = ... = G^H$ and $\sum_h \tau^h = 1$) and H unknowns (τ^h)

Lindahl Equilibria: Key Properties (Foley 1970)

- Samuelson Rule applies and outcome is Pareto efficient:

$$\sum_h [\frac{U_G^h}{U_X^h}] = \sum_h \tau^h = 1$$

- With identical individuals, simply set tax $\tau = \frac{1}{H}$ and ask individuals to voluntarily contribute to G
- With heterogeneity, efficient outcome can be attained with public goods through prices that are individual-specific

Lindahl Pricing: Practical Constraints

1. Must be able to exclude a consumer from using the public good.
    - Does not work with non-excludable public good
2. Must know individual preferences to set personalized prices τ^h
    - Agents have no incentives to reveal their preferences

- Difference between Lindahl equilibria and standard equilibria:
    - No decentralized mechanism for deriving prices; no market forces that will generate the right price vector
- So how do we actually determine level of PG's in practice?
    - Voting on bundles of PG's and taxes
    - Does voting lead to the first best solution?

Voting Model: Setup

- Suppose that public good is financed by fixed taxes $\tau^h G$
- Individuals vote on G but not on τ^h
- Preferences over G given by $U^h(Y^h - \tau^h G, G)$
- Voting equilibrium: level G_{eq} of public that cannot be defeated in majority rule by any other alternative $\hat{G}$
- Condorcet Paradox: majority voting does not lead to a stable outcome
- Consider voting on public school spending by 3 parents (low, middle, and high income)

Condorcet Paradox

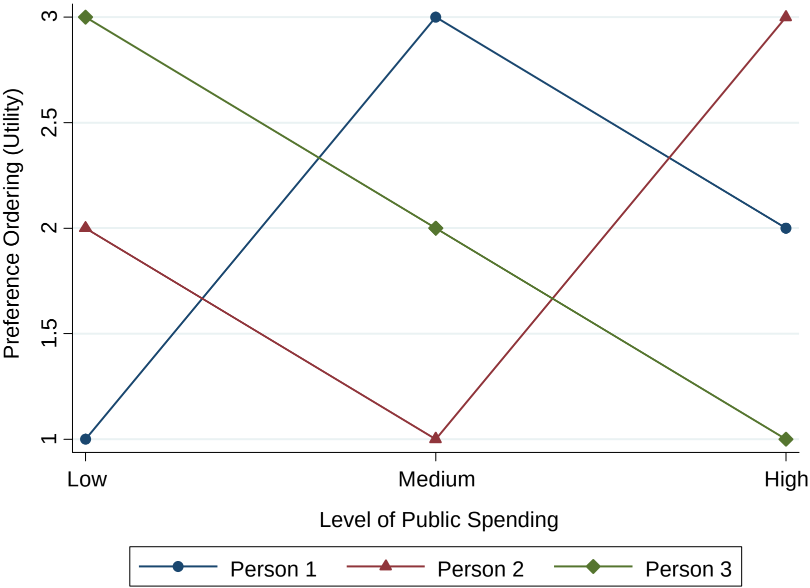
Preference Ordering	Individual		
	1	2	3
1st	H	M	L
2nd	M	L	H
3rd	L	H	M

Cycling in social ordering: H > M > L > H

Arrow (1951) and Single-Peaked Preferences

- Arrow's Impossibility Thm: Condorect Paradox is a general problem
- Only social choice rule that satisfies (a) Pareto Efficiency and (b) Independence of Irrelevant Alternatives is dictatorship.
- Subsequent work: restricts space of preferences to make progress
- Two assumptions that ensure existence of equilibrium:
  1. G unidimensional
  2. preferences over G are "single-peaked"

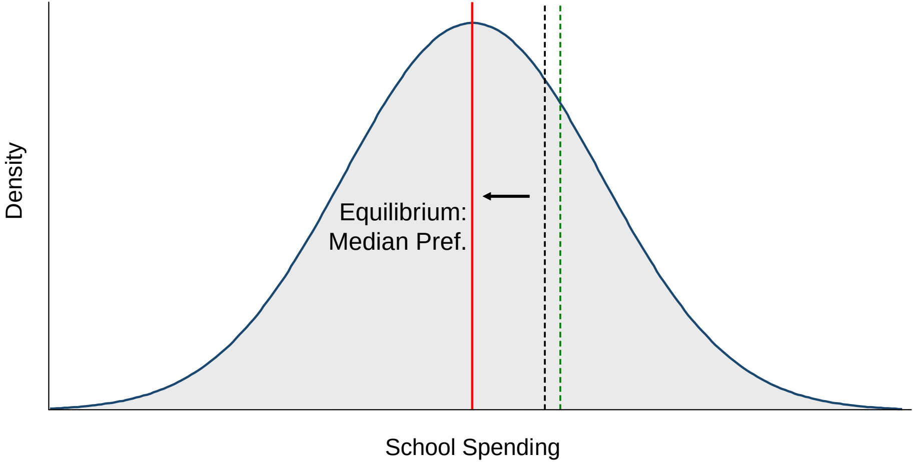
Single-Peaked Preferences



Median Voter Theorem

- With single-peaked preferences, majority voting rule produces a voting equilibrium (stable choice)
- Voting eq. is characterized by preferred level of voter whose preferred level of PG spending is at the median of the distribution
- Compute preferred spending for each individual, G^h
- Majority voting will select median of distribution of G^h

Median Voter Theorem

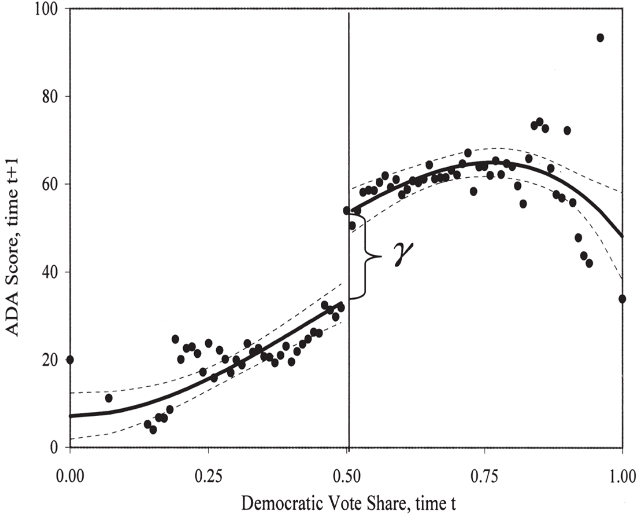


Median Voter Choice: Efficiency

- In general, median voter equilibrium is **not** Pareto efficient:
  - Suppose $\tau^h = 1/H$ for all h
  - Voting outcome: $MRS(G^{med}) = 1/H$.
  - Samuelson rule: $\sum_h MRS(G^h)/H = 1/H$
- Difference between median and mean determines degree of inefficiency
- Potential rationale for permitting lobbying to express intensity of preferences

Lee, Moretti, and Butler 2004

- In practice, citizens do not vote on every bill; elect representatives to do so.
- In a standard (Hotelling) model, median voter theorem predicts that candidates will implement median voter's preferences when elected
- Move toward center to win election
- Lee et al: does this happen in practice?
- Use "close" elections as experiments in an RD design



Source: Lee, Moretti, and Butler 2004

Lee, Moretti, and Butler 2004

- Evidence on Congressional voting sharply contradicts prediction of median voter theory
- Politicians' inability to credibly commit to a compromise dominates competition-induced convergence in policy.
- For example, a large exogenous increase in electoral strength for the Democratic party in a district does not result in shifting both parties' nominees to the left.
- Cannot rely on median voter logic to implement efficient choice even if mean and median are close
- Need to devise social choice mechanisms that account for commitment problems

Optimal Second Best Provision of PG's

- Suppose govt has decided to levy a tax and provide public goods based on some rule
- Two complications arise when trying to get to Samuelson First Best level:
  1. Interactions with private sector provision (crowdout).
     - Andreoni (2007): \$250B/yr in private contributions.
  2. Government cannot finance PGs through lump sum taxation
     - Must modify Samuelson rule to account for distortionary taxation

Public Goods with Endogenous Private Provision

- Interest in crowd-out began with Roberts (1984)
  - Expansion of govt services for poor since Great Depression accompanied by comparable decline in charitable giving for the poor.
  - Conclusion: government has grown tremendously without having any net impact on poverty or welfare
  - Evidence mainly based on time series impressions.
- But theory underlying this claim very sensible, as subsequent work showed

Bergstrom, Blume, and Varian (1986): Setup

- Individual h solves:

$$\max_{X^h, G^h} U^h(X_h, G_h + G_{-h})$$
$$\text{s.t. } X_h + G_h = Y_h$$

- FOC is $U_X^h = U_G^h$
- Nash equilibrium exists and is unique
  - G s.t. all individuals optimize given others' behavior
- Let G^* denote private equilibrium outcome

Bergstrom-Blume-Varian Model: Crowd-out

- Now suppose government introduces lump sum taxes t^h on each individual h
- Revenue used to finance expenditure on public good $T = \sum t^h$
- Individual's optimization problem is now:

$$\max U(X^h, G_h + G_{-h} + T)$$
$$\text{s.t. } X^h + G^h = Y^h - t^h$$

Bergstrom-Blume-Varian Model: Crowd-out

- Let $Z_h = G_h + t_h$ denote total contribution of individual h.
- Can rewrite this as:

$$\max U(X^h, Z_h + Z_{-h})$$
$$\text{s.t. } X^h + Z^h = Y^h$$

- This is isomorphic to original problem $\Rightarrow Z^* = G^*$
- Total public good provision is unchanged!
- Each person simply reduces voluntary provision by t_h

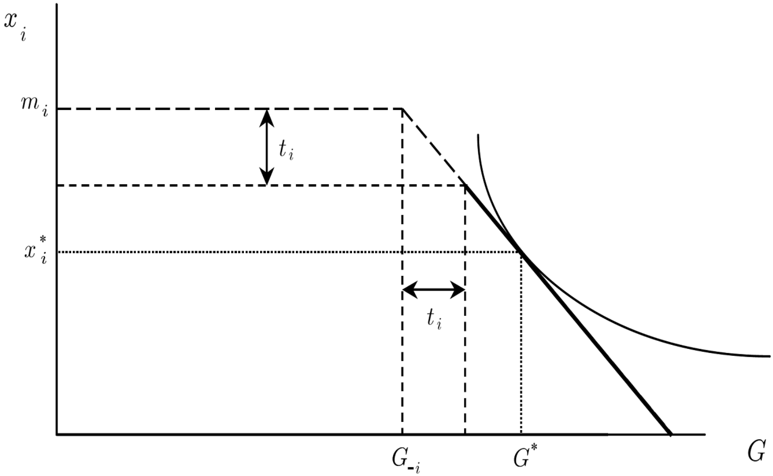


Figure 5: Complete Crowding Out

Source: Andreoni 2006

BBV Model: Additional Results

1. Total supply of the public good is indep. of the distribution of income among givers (Warr 1983)
   - Logic can be seen with two transfers.
   - Tax indiv. 1 to finance PG; then subsidize indiv. 2 and reduce PG expenditure.
   - Neither action has a real effect by crowdout result.

2. When preferences are identical and separable in x and G, all givers to a public good will have the same level of **private** consumption in eq. regardless of their incomes (BBV 1986).

$$\begin{aligned} u_x^1(x, G) &= u_G^1(x, G) = u_G^2(x, G) = u_x^2(x, G) \\ &\Rightarrow x_1 = x_2 \end{aligned}$$

3. As size of economy gets large, the proportion of individuals who give to the PG approaches zero (Andreoni 1988).

BBV Model: Key Assumptions

1. No corners: assumed the set of contributors are the same in both situations.
   - With corners, transfer neutrality breaks down: tax increase T results in no private contribution from individuals with $G^h < T$, but contributions increase on net.
2. Ignores direct utility from giving: $U(X^h, G^h, G)$.
   - Andreoni's (1990) "warm glow" model.
   - Stigler and Becker (1977) critique: should not simply modify preferences to explain patterns
3. Ignores prestige/signalling motives
   - Glazer and Konrad (1996)

Empirical Evidence on Crowd-Out

- Two empirical questions motivated by theory
  1. How large is the degree of crowd-out in practice?
  2. What are the income and price effects on charitable giving?
- Two strands of empirical literature
  1. Field evidence (observational studies)
  2. Lab experiments
- Traditionally, lab experiments have been more influential but recent field studies may change this
- Lab experiments may not capture important motives for giving: warm glow, prestige

Kingma 1989

- Studies individual contributions to public radio stations
  - Cross-sectional survey of individuals who listened to public radio.
  - 3,500 individuals and 63 different radio stations.
- Research Design: OLS regression of individual contributions on government support

$$D_i = \beta_0 + \beta_1 G_i + X_i \gamma + \epsilon_i$$

- D_i = individual contribution
- G_i = government support
- X_i = set of controls: individual income, individual education, age, price (tax bracket).

TABLE 1

SUMMARY STATISTICS FOR PUBLIC RADIO LISTENERS

	Contributors		Noncontributors	
	Mean	Standard Deviation	Mean	Standard Deviation
Contribution ($)	45	52	0	0
Age (years)	46	14	45	16
Income ($)	48,074	27,586	37,582	24,818
Education (years)	16.4	2.4	15.3	2.8
Listening (hours per week)	10.75	11.25	6.00	8.00

$$
\begin{aligned}
D = & \underset{(3.38)}{-65.036} + \underset{(6.49)}{.539}(\text{INCOME}) - \underset{(.94)}{.010}(R) - \underset{(3.02)}{.015}(G) \\
& - \underset{(.74)}{15.014}(\text{PRICE}) + \underset{(8.01)}{10.018}(\text{EDUCTN}) + \underset{(3.26)}{.288}(\text{AGE})
\end{aligned}
$$

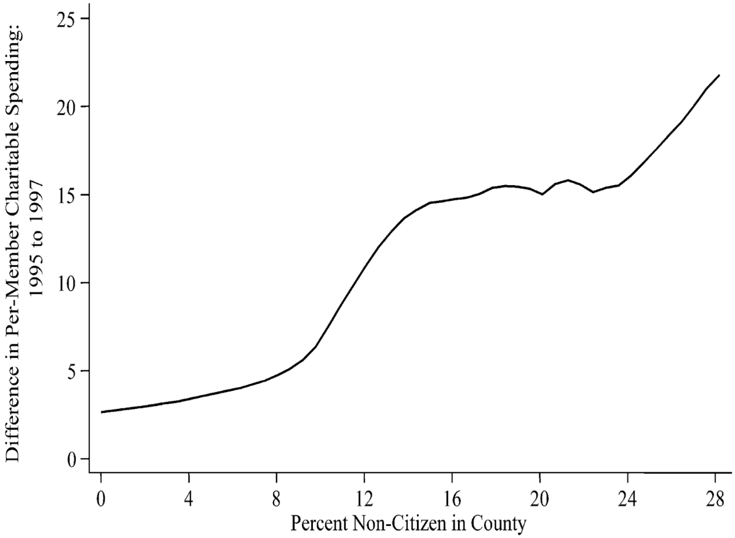
Source: Kingma 1989

Kingma 1989

- Main result: $\beta_1 = -0.015$ –> crowd-out rate of 20%
- Significantly negative but much less than 1-1 prediction of theory
- Problem: government support might depend on individual contributions.
  - E.g. non-contribution by individuals leads to govt provision
  - Creates a spurious negative correlation between govt support and individual contributions.
- We need an exogenous "shifter" that affects govt contribution without affecting individual contributions.
  - E.g.: legislated reform that bans govt support.

Hungerman 2005

- Studies crowdout of church-provided welfare (soup kitchens, etc.) by government welfare.
- Uses 1996 Clinton welfare reform act as an instrument for welfare spending.
- One aspect of reform: reduced/eliminated welfare for non-citizens.
- Motivates a diff-in-diff strategy: compare churches in high non-citizen areas with low non-citizen areas before/after 1996 reform.



Source: Hungerman 2005

Hungerman 2005

- Estimates imply that total church expenditures in a state go up by 40 cents when welfare spending is cut by $1.
- Exogenous variation make these estimates much more credible

Andreoni and Payne 2003

- Government spending crowds-out private donations through two channels: willingness to donate + fundraising
- Use tax return data on arts and social service organizations
- Panel study: includes organization and year fixed effects

Andreoni and Payne 2003

- OLS still yields "wrong signed" estimates:
  - More government spending $\longrightarrow$ more fundraising
  - Endogeneity still a problem: hurricane $\longrightarrow$ more dollars for Red Cross and Federal aid
- Use the following instruments
  1. Total state-level transfer to non-profits (state budget)
  2. Representative on senate/house appropriations committee
  3. NIH fundings to univs in state (relieves funding for other purposes)

Table 3—Relationship Between Fund-raising and Government Grants, 2SLS

Dependent variable in second stage Fund-raising expenditures	Arts organizations (1)	(2)	Social service organizations (3)	(4)
Government funding ($1,000s)	**−264.70**	**−142.93**	**−53.75**	−19.35
	(113.77)	(64.26)	(20.97)	(13.55)
Program revenue ($1,000s)	**32.91**	**26.53**	6.90	**9.75**
	(11.77)	(9.05)	(5.64)	(4.16)
Dues and fees ($1,000s)	***158.57***	**181.99**		
	(86.41)	(67.75)		
Investment securities	**4.44**	**3.78**	−0.95	−0.91
(beginning of year, $1,000s)	(1.83)	(1.38)	(0.90)	(0.88)
Assistance to individuals ($1,000s)			**53.56**	***26.30***
			(23.72)	(15.87)
R^2 on second stage	0.6144	0.7911	0.8915	0.9141
Results from first stage				
Instrument set	NIH grants to universities in state one-year lag	Federal transfers to nonprofits one-year lag	NIH grants to universities in state one-year lag	Federal transfers to nonprofits one-year lag
F-test on instruments	4.72	2.55	5.75	3.92
(p-value)	(0.00)	(0.01)	(0.00)	(0.00)
Overidentification test	2.35	14.08	1.34	12.51
(degrees of freedom)	(3)	(7)	(4)	(8)
(p-value)	(0.50)	(0.05)	(0.85)	(0.19)
Hausman test	19.89	11.53	22.03	4.47
(p-value)	(0.00)	(0.00)	(0.00)	(0.03)
Fixed effects	Organization and year	Organization and year	Organization and year	Organization and year
Number of observations	2,417	2,417	4,954	4,954
Number of organizations	233	233	534	534

Source: Andreoni and Payne 2003

Andreoni and Payne 2003

- \$1000 increase in government grant leads to \$250 reduction in private fundraising
  - Suggests that crowdout could be non-trivial if fundraising is a powerful source of generating private contributions
- Subsequent study by Andreoni and Payne (2008) confirms that it is
  - Using similar strategy and a larger panel, find that \$1 more of government grant to a charity leads to 56 cents less private contributions
  - 70 percent (\$0.40) due to the fundraising channel
  - Suggests that individuals are relatively passive actors

Marwell and Ames 1981

- Early lab experiments testing free-rider behavior.
- Groups of 5 subjects, each given 10 tokens.
- Can invest tokens in either an individual or group account.
  - Individual: 1 token = $1 for me; Group: 1 token = 50 cents for everyone
- Nash equilibrium is 100% individual but Pareto efficient outcome is 100% group.
- Compute fraction invested in group account under various treatments

Table 2

Summary of results: Experiments 1–11.

Experiment	Mean % of resources invested
1. Basic experiment	42%
2. Skewed resources and/or interest	53%
Experiments 1 and 2, combined	51%
3. Provision point	51%
4. Small groups with provision point (except those with sufficient interest to provide the good themselves)	60%
5. Experienced subjects	47%
6. High stakes	
Experienced interviewers	35%
All interviews	28%
7. Feedback, no changing initial investment	46%
8. Feedback, could change investment in individual account	50%
9. Feedback, could change investment in individual account — college students	49%
10. Manipulated feedback	
Low	43%
Medium	50%
High	44%
11. Non-divisibility	
Divisible (control)	43%
Non-divisible	84%
12. Economics graduate students	20%

Source: Maxwell and Ames 1981

Marwell and Ames 1981

- Finding: 40 to 60% of tokens were still invested in the public good.
- Experiment run on various groups of high school and college students.
- Only one group free-rode a lot: 1st year econ graduate students (20% donation rate).
  - "Economists Free Ride, Does Anyone Else?"
  - Marglin: thinking like an economist undermines community

Andreoni 1988

- Isaac, McCue, and Plott (1985): when the game is repeated with same set of players, public good contribution levels fall over time.
- Andreoni (1988): is this b/c of learning or strategic behavior?
- Game to distinguish these two hypotheses:
    - 10 iterations of Marwell-Ames game
- Two different samples:
    - Group A: play with strangers
    - Group B: play with partners (stable groups)
- Strategic hypothesis predicts strangers free ride more

Table 2

Percent of subjects free riding.[a]

	Round										
	1	2	3	4	5	6	7	8	9	10	All
Partners	16.6	13.3	20.0	23.3	33.3	30.0	40.0	40.0	40.0	70.0	34.3
Strangers	15.0	12.5	15.0	15.0	15.0	17.5	22.5	25.0	30.0	42.5	20.5
Difference	1.6	0.8	5.0	8.3	18.3	12.5	17.5	15.0	10.0	27.5	13.8

[a] $n_P = 30$; $n_S = 40$.

Source: Andreoni 1988

Andreoni 1993

- Uses lab experiment to directly test crowdout hypothesis with "government" provision
- Payoffs $U_h = (7 - G_h)^{1-\alpha} G^{\alpha}$
- Two groups: no-tax and tax
- No-tax group can choose $G_h = 0, 1, 2, .., 7$
- Tax group automatically gets 2 tokens allocated to G and can choose $G_h = 0, 1, 2, .., 5$
- Each game repeated twenty times
- Nash equilibrium in **no-tax** game is $G_h = 3$ but Pareto efficient outcome is $G_h = 6$

	YOUR INVESTMENT							
TOTAL INVESTMENT BY THE OTHER TWO GROUP MEMBERS	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
0	0	1	3	6	9	10	11	10
1	1	4	8	11	14	15	15	14
2	5	9	14	18	20	21	20	17
3	12	17	22	26	28	28	25	22
4	21	28	33	36	37	35	32	27
5	34	40	45	48	47	44	39	32
6	49	56	60	61	59	54	47	38
7	68	74	77	76	72	64	55	44
8	90	95	96	93	86	76	64	51
9	115	118	117	111	102	89	74	58
10	143	144	140	131	119	103	85	66
11	175	173	166	153	137	118	97	75
12	210	205	193	177	157	134	109	84
13	248	239	223	203	178	151	122	93
14	290	276	256	230	201	169	136	103

FIGURE 1. THE NO-TAX PAYOFF MATRIX

Source: Andreoni 1993

	YOUR INVESTMENT						
		0	1	2	3	4	5
TOTAL INVESTMENT BY THE OTHER TWO GROUP MEMBERS	0	33	36	37	35	32	27
	1	45	48	47	44	39	32
	2	60	61	59	54	47	38
	3	77	76	72	64	55	44
	4	96	93	86	76	64	51
	5	117	111	102	89	74	58
	6	140	131	119	103	85	66
	7	166	153	137	118	97	75
	8	193	177	157	134	109	84
	9	223	203	178	151	122	93
	10	256	230	201	169	136	103

FIGURE 2. THE TAX PAYOFF MATRIX

Source: Andreoni 1993

TABLE 1–AVERAGE REAL CONTRIBUTIONS TO THE PUBLIC GOOD PER ROUND FOR THE NO-TAX AND TAX CONDITIONS

Round	No-tax condition			Tax condition			Overall difference
	Trial 1	Trial 2	Average	Trial 1	Trial 2	Average	
1	2.94	2.83	2.89	4.05	3.66	3.86	0.97
2	2.66	2.55	2.61	3.72	3.72	3.72	1.12
3	3.33	2.61	2.97	3.38	3.72	3.55	0.58
4	3.11	2.38	2.75	2.77	2.77	2.77	0.02
5	2.72	3.33	3.03	4.27	3.61	3.94	0.91
6	2.72	2.72	2.72	3.44	3.38	3.41	0.69
7	2.88	3.05	2.97	4.11	2.94	3.53	0.56
8	2.72	2.66	2.69	3.22	2.88	3.05	0.36
9	2.88	2.55	2.72	4.05	3.55	3.80	1.09
10	3.22	3.11	3.17	3.77	3.11	3.44	0.27
11	2.88	2.83	2.86	3.61	3.22	3.42	0.56
12	2.61	2.55	2.58	3.00	3.00	3.00	0.42
13	2.77	2.83	2.80	3.77	3.33	3.55	0.75
14	2.61	2.66	2.64	3.44	2.83	3.14	0.50
15	2.77	2.61	2.69	3.27	2.94	3.11	0.42
16	3.05	2.72	2.89	2.94	2.88	2.91	0.02
17	3.27	2.77	3.02	3.50	3.61	3.56	0.54
18	2.44	3.05	2.75	3.72	2.72	3.22	0.48
19	3.05	2.38	2.72	3.38	2.83	3.11	0.39
20	2.11	2.38	2.25	3.16	2.83	3.00	0.75
Average:	2.84	2.73	2.78	3.53	3.18	3.35	0.57

Source: Andreoni 1993

Andreoni 1993

- Public good levels are significantly higher in the tax case.
- However, crowd-out is substantial: 71.5% on average.
  - Compare with empirical studies that find 30% crowding out.
- Crowd-out increases in final rounds.
- Considered an upper bound of degree of crowd out
  - Missing warm glow, social pressure, lack of salience.

Crowdout: Summary of Evidence

- Rate of crowdout is probably 30 cents on the dollar on average, but probably highly heterogeneous.
- Non-trivial but far from BBV/Roberts prediction.
- Key factors are probably warm glow and salience
- Suggests that carefully targeted govt programs can still have considerable net impact.

Financing PGs with Distortionary Taxation

- Second problem in implementing the Samuelson Rule is that the government cannot use lump sum taxation in practice because of redistributional concerns
- For this section, ignore private crowdout problem
- Instead, consider goods where individuals are at corners, such as roads or defense

Pigou's Conjecture (1947)

- Total costs of providing public good are higher than its production costs when it is financed by distortionary taxation

  1. At the optimum: the MB of the public good should be equal to the MC of production plus the marginal deadweight burden of taxation
  2. The optimal level of public goods with distortionary taxation is **lower** relative to a 1st best where govt can use lump sum taxation

- Subsequent formal analysis (Atkinson and Stern 1974) showed this is true, but with a few caveats

PGs with Distortionary Taxes: Setup

- Large number of identical individuals
- Utility over private consumption (c), labor (l), and PG (G)

$$U(c, l, G) = c - l^{k+1}/(k+1) + v(G)$$

- Prices of c and G are both 1 (MRT = 1)
- Individuals do not contribute b/c the contribution of one individual has a negligible effect on G.

PGs with Distortionary Taxes: Setup

- Two policy instruments: lump sum tax R and linear tax on labor τ:

$$c = wl(1-\tau) - R$$

- Individual chooses l to maximize

$$wl(1-\tau) - R - l^{k+1} + v(G)$$

where G is viewed as fixed (individual is small).

- Implies that

$$w(1-\tau) = l^k$$
$$\Rightarrow \quad l = w^e(1-\tau)^e$$

where $e = 1/k$ is the elasticity of labor supply with respect to the net-of-tax rate $1-\tau$

- Public good level equals tax revenue: $G = wl\tau + R$

PGs with Distortionary Taxes: 1st Best

- When lump sum tax instrument is available, govt maximizes

$$\begin{aligned} W &= wl(1-\tau) - R - l^{k+1}/(k+1) + v(G) \\ &= wl(1-\tau) - R - l^{k+1}/(k+1) + v(wl\tau + R) \end{aligned}$$

where $l = [w(1-\tau)]^e$ is chosen optimally by the individual

- FOC in R implies that $v\prime(G) = 1$ (Samuelson rule)

- Public good provided up to the point where sum of MRS $v'(G)/1$ equal MRT 1

PGs with Distortionary Taxes: 1st Best

- In first best, optimal linear tax is $\tau^* = 0$

$$\begin{aligned}\frac{\partial W}{\partial \tau} &= -wl + wlv'(G) + w\tau v'(G)\partial l/\partial \tau \\ \frac{\partial W}{\partial \tau} &= w\tau v'(G)\partial l/\partial \tau\end{aligned}$$

- Therefore $\frac{\partial W}{\partial \tau}(\tau^*) = 0 \Rightarrow \tau^* = 0$

PGs with Distortionary Taxes: 2nd Best

- In second best, lump sum tax unavailable ($R = 0$).
- Govt chooses τ to maximize:

$$W = wl(1-\tau) - l^{k+1}/(k+1) + v(wl\tau)$$

- Using the envelope theorem yields f.o.c. for τ:

$$0 = \partial W/\partial \tau = -wl + v'(G)(wl - w\tau \partial l/\partial(1-\tau))$$
$$\Rightarrow \quad 1 = v'(G)[1 - \frac{\tau}{1-\tau} e]$$

- Added term $\frac{\tau}{1-\tau} e$ in formula relative to Samuelson rule
  - $v'(G^{SB}) > v'(G^{FB}) = 1$ which implies $G^{SB} < G^{FB}$ because $v(G)$ is concave
  - Higher threshold (MCPF > 1): depends on e.

Heterogeneity: Gaube 2000

- In a setting with heterogeneity in prefs for PG (v_h), proves that
  1. Without redistributive prefs., $G^{SB} < G^{FB}$.
  2. With redistributive tastes, could have $G^{SB} > G^{FB}$.
- Intuition: if public parks benefit mostly low income households, over-provide parks to enhance redistribution
  - First-best level of redistribution cannot be achieved using standard tax instruments.
  - By providing park instead of welfare, redistribute income without distorting incentive to work.
- Example of theory of the second best (Lipsey and Lancaster 1956)

Kreiner and Verdelin 2009

- Consider a general model with non-linear income taxes (including lump sum)
- Q: What threshold should be used for PG's in this setting?
- A: Depends on whether non-linear tax system is reoptimized when PG's are funded
- If yes, then Samuelson rule correct again
  - E.g. if public good benefits all equally, then simply raise lump sum tax and distributional problem is unchanged
  - More generally, changes in optimal non-linear tax system will have second-order effects on welfare $\rightarrow$ can be ignored.
- Illustrates danger of Ramsey analyses

Subsides for Private Provision of PGs: Charitable Giving

- Alternative to distortionary taxes is subsiding private provision.
- E.g. in the U.S., charitable contributions are tax deductible ($20 bil tax expenditure).
- Theoretical questions:
  - Should we have such a subsidy?
  - How large should such a subsidy be?
  - What are the key determinants of the optimal subsidy?
- Empirical questions:
  - How sensitive is charitable giving w.r.t. tax subsidies?
  - Where does the money end up going (social value)?

Subsidies for Charity: Setup

- Warm-glow model. Individual maximizes

$$U(c, g) \text{ s.t. } c + g = y - \tau(y - \theta g)$$

where θ measures degree of deductibility of charitable insurance.

- Price of giving \$1 to charity is $\$1 - \theta\tau$.

- Define price and income elasticities:

$$\beta = \frac{1-\theta\tau}{g} \cdot \frac{\partial g}{\partial(1-\theta\tau)} = \text{price elasticity of charitable giving}$$

$$\gamma = \frac{y}{g} \cdot \frac{\partial g}{\partial y} = \text{income elasticity of charitable giving}$$

Subsidies for Charity: Setup

- First consider case where govt uses tax revenue to fund same PG as individual.
    - Here marginal value of PG and charity are identical.
- Let P denote total funding for public good:

$$P = \tau(y - \theta g) + g = k + (1 - \theta\tau)g$$

- Question is whether to use tax subsidy and get indiv to contribute or just fund through tax revenue

Optimal Subsidies for Charity

- Result 1: If $\beta < -1$ then deduction of charities unambiguously desirable

- What is gained in additional contributions is larger than tax revenue loss.

$$
\begin{aligned}
P &= k + (1 - \theta\tau)g \\
\frac{dP}{d\theta} &= -\tau g + \frac{dg}{d\theta}\frac{1-\theta\tau}{g}g \\
&= -\tau g - \frac{dg}{d(1-\theta\tau)}\frac{1-\theta\tau}{g}\tau g \\
&= -\tau g - \beta\tau g \\
&= -\tau g(1+\beta)
\end{aligned}
$$

- Therefore $\beta < -1$ implies $\frac{dP}{d\theta} > 0$.

- Clearly desirable to subsidize at least up to the point where $\beta(\theta) = 1$.

Optimal Subsidies for Charity (Saez 2004)

- Now consider case where marginal values of private charity and PG differ
  - Marginal value of public spending $= 1$
  - Marginal value of private charity $= \lambda(G)$
- Multiplier $\lambda(G) \in [0, 1]$ measures external effect of charitable contributions on social welfare

Optimal Subsidies for Charity (Saez 2004)

- With flat social welfare weights, optimal tax rate t_G for charitable good G satisfies
$$\frac{\lambda + t_G}{1 + t_G} = \frac{1}{\beta}$$
- Generalizes Ramsey inverse-elasticity rule by allowing $\lambda > 0$
- Analogous to Sandmo 1975 correction for externalities
- If $\lambda = 1$ (PG equivalent to charity), t_G should be set so that $\beta(t_G) = 1$, as above

Optimal Subsidies for Charity (Saez 2004)

- Key elements for optimal tax or subsidy of charitable contributions:
  - Who benefits from charitable contributions (λ)?
  - Are charitable contributions responsive to the subsidy (β)?
  - In a many-person model with heterogeneity, need social welfare weights of those contributing.
    - What are the incomes of those contributing?

Empirical Evidence

- Existing studies have estimated β and γ - income and price elasticities.
- General specification:

$$\log(g) = \alpha + \beta \log(1 - \tau) + \gamma \log y + \epsilon$$

- Early work (Feldstein and Taylor 1976, Clotfelter 1985): cross-sectional regressions with controls.
- Results: $\gamma = 0.8$, $\beta = -1.3$.
- But results confounded: effectively comparing rich and poor

Empirical Evidence: Randolph (1995)

- Uses ten year tax return panel (1979-1988) and fits DD-type models.
- Finds short-term elasticities: 1.2; long-term elasticities: 0.6.
- Income effects are larger in the long-term than in the short-term.

Externalities: Outline

1. Definition and Basic Model
2. Correcting Externalities
3. Prices vs. Quantities (Weitzman 1974)
4. 2nd Best Taxation with Externalities (Sandmo 1975)
5. Empirical Applications

Definition

- An externality arises whenever the utility or production possibility of an agent depends **directly** on the actions of another agent.
- Important distinction between "pecuniary" vs. "non-pecuniary" externalities
    - Consuming an apple vs. consuming loud music
    - Not a technological distinction; depends on market in place
    - Coasian view: can convert all externalities into pecuniary externalities with appropriate markets, property rights.
- Only non-pecuniary externalities justify policy intervention

Externalities: Main Questions

1. Theoretical: what is the best way to correct externalities and move closer to the social optimum?

2. Empirical: how to measure the size of externalities?

    - Key difference: cannot use standard revealed-preference methods

Model of Externalities

- Firms produce x cars using $c(x)$ units of the numeraire y.
- Generates x units of pollution: $P(x) = x$.
- Consumers have wealth Z and quasilinear utility:

$$u(x) + y - d \cdot P(x)$$

where $d =$ marginal damage (MD) of pollution

- Social welfare is

$$W = u(x) + Z - c(x) - d \cdot x$$

- Let p denote the market price of cars.

Model of Externalities: Equilibrium

- Firms max profits:

$$\max px - c(x)$$

- Consumers max utility, taking level of pollution as fixed:

$$\max u(x) + Z - px$$

- Demand satisfies

$$u'(x^D) = p$$

- Supply satisfies

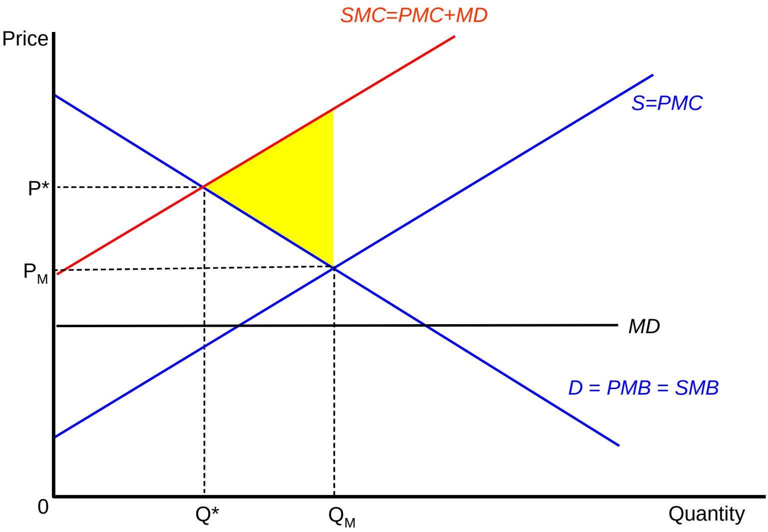
$$c'(x^S) = p$$

- PMB equals PMC in equilibrium:

$$u'(x^D) = c'(x^S)$$

- But this is not Pareto efficient

Negative Production Externalities: Pollution



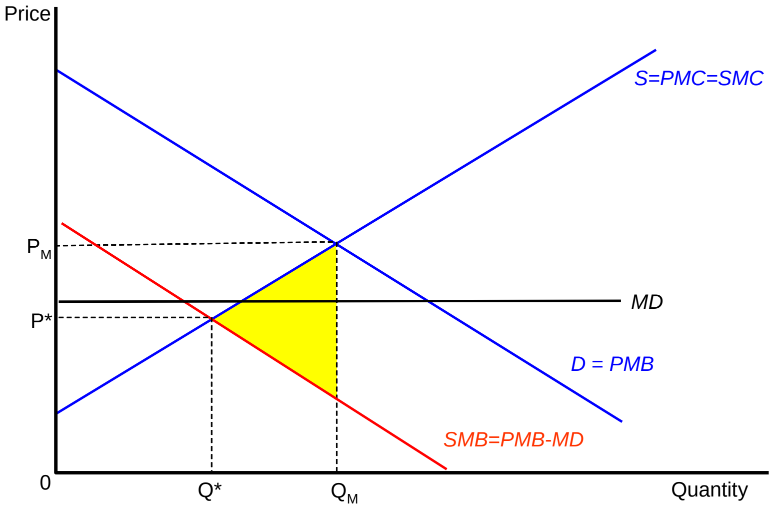
Model of Externalities: Deadweight Loss

- Perturbation argument: can increase social welfare by reducing production by Δx:

$$\begin{aligned} dW &= u'(x)\Delta x - c'(x)\Delta x - d \cdot \Delta x \\ &= -d \cdot \Delta x > 0 \text{ if } \Delta x < 0 \end{aligned}$$

- First Welfare Theorem does not hold
- Analogous result for consumption externalities

Negative Consumption Externalities



Remedies for Externalities

1. Coasian bargaining solution
2. Pigouvian corrective taxation
3. Regulation
4. Permits (cap-and-trade)

Coasian Solution

- Externalities emerge because property rights are not well defined.
- Establish property rights to create markets for pollution.
- Consider example of pollution in a river.
  - If consumer owns river, in competitive equilibrium, firms pay d for every unit of pollution emitted.
  - Marginal cost of production is now $c'(x) + d$, leading to 1st best.
- Symmetric solution when firm owns river.
- Assignment of property rights affects distribution but not efficiency

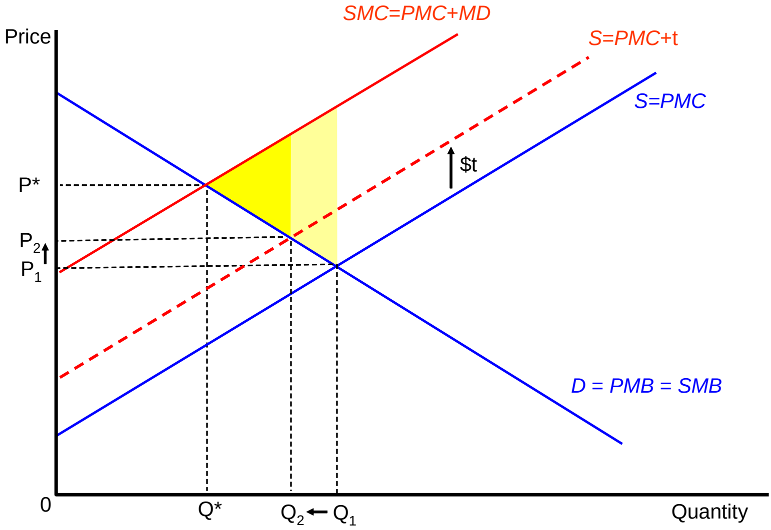
Coasian Solution: Limitations

1. Cost of bargaining
   - Ex: air pollution – would require millions of agents to coordinate and bargain
   - To reduce transactions costs, need an association to represent agents
   - This "association" is the government
2. Asymmetric information: competitive equilibrium can break down
   - Often hard to identify precise source of damage
   - E.g. atmospheric pollution very diffuse, marginal damages unclear

Pigouvian Taxation

- Impose tax $t = MD(Q^*)$
- Restores Pareto efficiency and maximizes social welfare
- Practical limitations:
  - Must know marginal damage function to set t
  - Difficult to measure the marginal damage in practice

Pigouvian Tax



Regulation: Command and Control

- Must reduce pollution to set level or face legal sanctions.
- Same outcome as Pigouvian taxation: move people to x_2

Advantages:

1. Ease of enforcement
2. Salience, political expedience

Disadvantages:

1. Dynamics: no incentive to innovate
2. Allocative inefficiency with heterogeneity in cost of pollution reduction

Permits: Cap-and-Trade

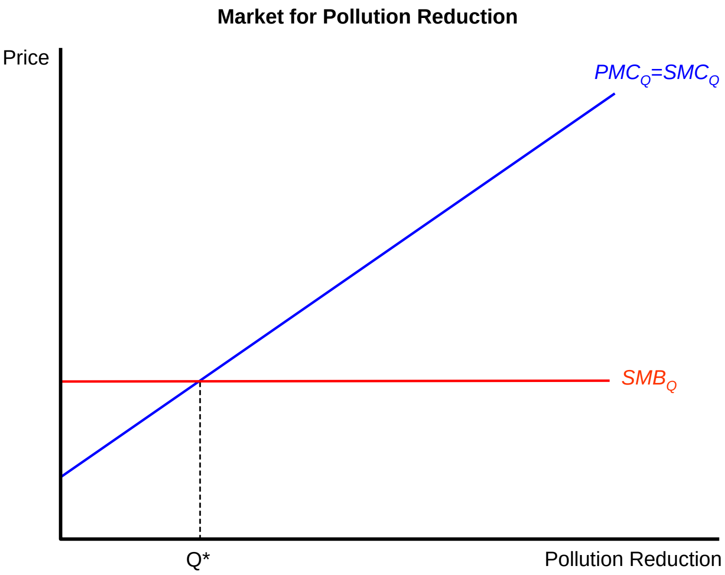
- Cap total amount of pollution and allow firms to trade permits to pollute
- Address disadvantages of regulation using an auction-based permit system.
- Hybrid of regulation and Coasian solution.
- In eq., firms with highest MC of reducing pollution will buy permits; those that can easily reduce pollution will do so.
- If total number of permits is set to achieve the social optimum, both allocative and productive efficiency will be achieved.
- Also have dynamic incentives to innovate because each firm is bearing a marginal cost of pollution.

Weitzman 1974: Prices vs. Quantities

- Price mechanism (taxes) identical to quantity mechanism (permits) in simple model above. How to choose?
- Weitzman (1974): with uncertainty re. shape of MB and MC curves, price and quantity no longer equivalent.
- Now the standard method of choosing between regulation and taxes

Weitzman 1974: Market for Pollution Reduction

- Let q denote pollution **reduction** starting from private market eq., where $q = 0$.
- Let $B(Q)$ denote social benefits of pollution reduction
- Let $C(Q)$ denote social costs.
- In simple model above:
  - MB of pollution reduction is constant, $B'(Q) = d$.
  - MC given by loss in surplus from producing one less car: $u'(x) - c'(x)$.
  - More generally, MC should be interpreted as cost of reducing pollution through cheapest method (e.g. cleaner plants)


Market for Pollution Reduction
Price
PMC_Q=SMC_Q
SMB_Q
Q*
Pollution Reduction

Weitzman Model: Policy without Uncertainty

- In eq'm, PMB of pollution reduction is 0 $\Rightarrow$ level of pollution reduction is $Q = 0$.
- Social optimum:
$$\max B(Q) - C(Q)$$
- First order condition:
$$C'(Q^*) = B'(Q^*)$$
- With no uncertainty, can obtain optimum with either quantity or price policy.
  - Quantity: require amount Q^*.
  - Price: set price for pollution reduction of $p^* = C'(Q^*)$.

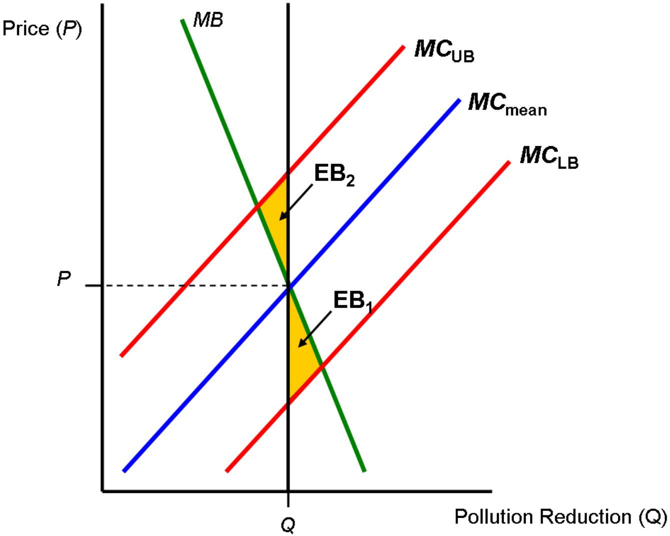
Weitzman: Optimal Policy with Uncertainty

- Now suppose that there is uncertainty about the marginal costs of reducing pollution.
- Cost is now $C(Q, \theta)$ with θ unknown.
- Marginal cost lies between MC_{LB} and MC_{UB}, with mean value given by MC_{mean}.
- Objective: maximize **expected** social welfare
- Formally, choose one of two options: p or Q directly:

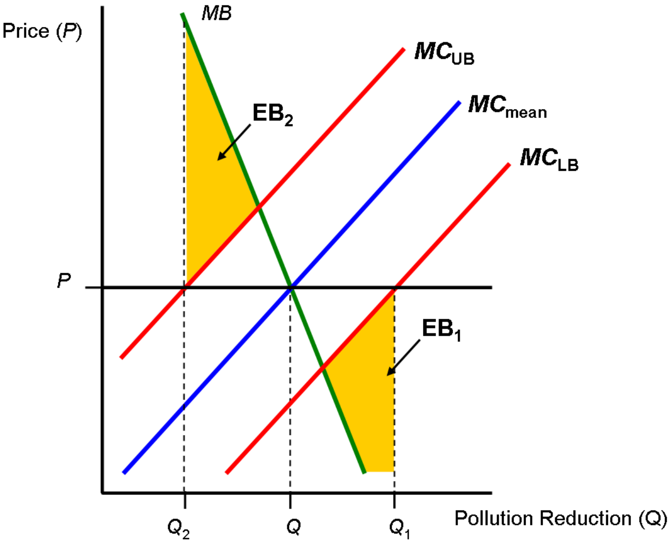
$$\max\{E_\theta B(Q) - C(Q, \theta), E_\theta B(Q(p)) - C(Q(p), \theta)\}$$

- Choice depends on steepness of marginal benefit curve.

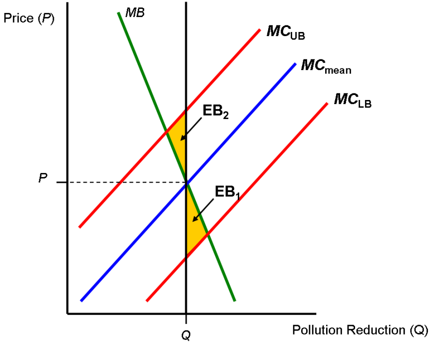
MB steep, Quantity regulation



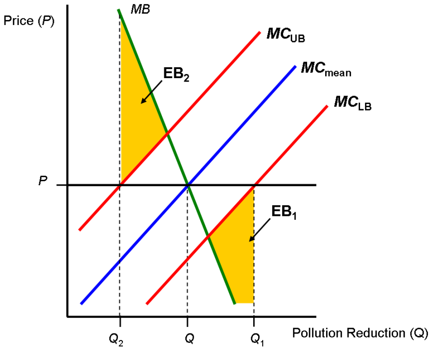
MB Steep, Price Regulation



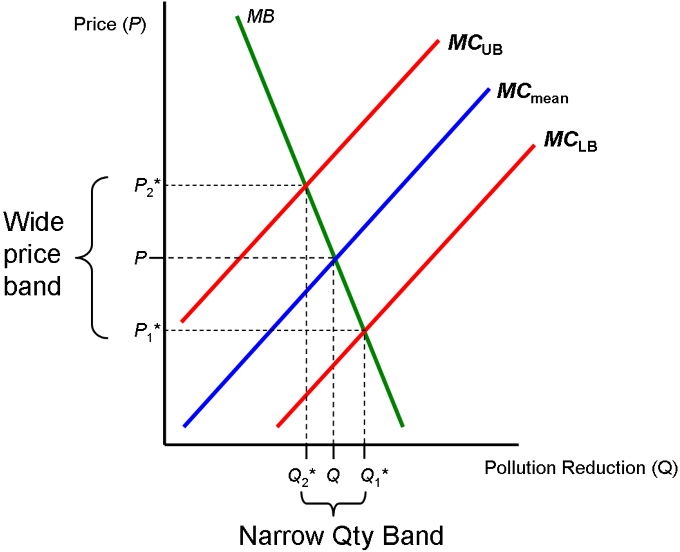
Quantity Regulation



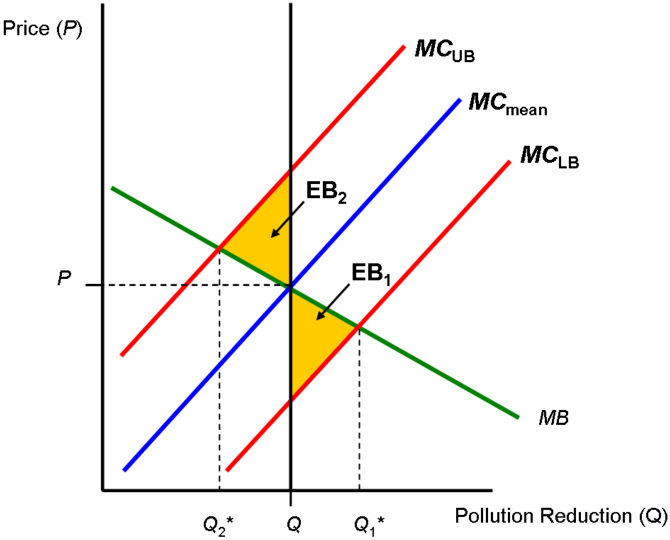
Price Regulation



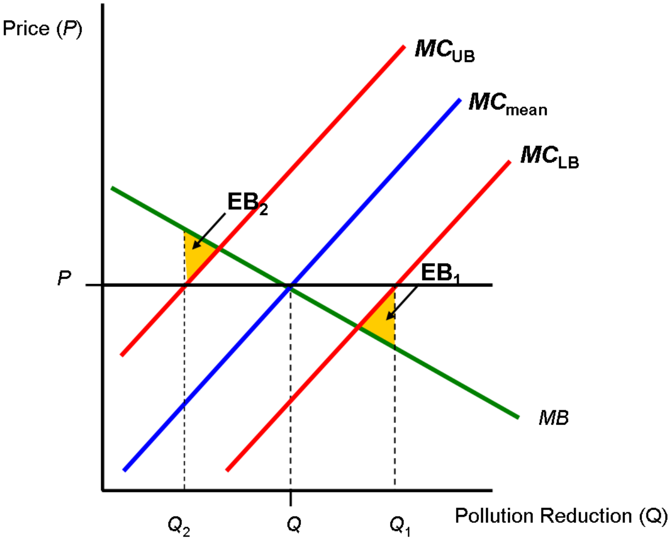
Price Band vs. Quantity Band with Steep MB



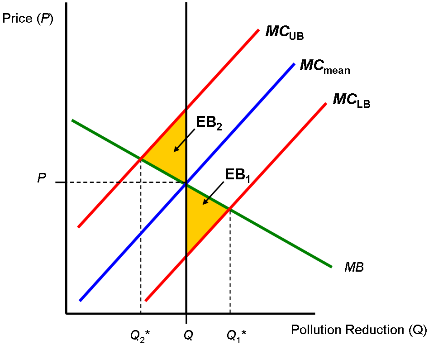
MB Flat, Quantity Regulation



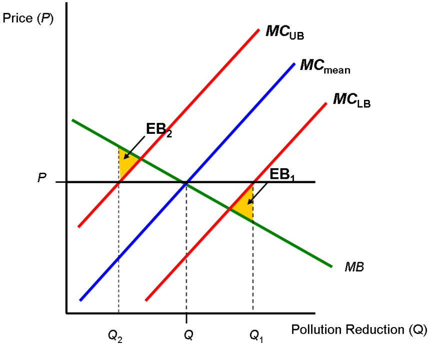
MB Flat, Price Regulation



Quantity regulation



Price Regulation



Weitzman: Uncertainty about Benefits

- Now suppose that there is uncertainty about the marginal **benefits** of reducing pollution but that the costs are known.
- Price and quantity policies are again equivalent.
- For a given p, the government knows the Q that will result exactly since $p = C'(Q)$.
- More generally, uncertainty matters only when it is about the cost/benefit schedule for the agent who chooses level of pollution reduction.
  - If consumer chooses level of pollution reduction, then only uncertainty about marginal benefits matters

Optimal Second-Best Taxation with Externalities

- In general, cannot restore 1st best b/c externality is one of many deviations from first best.
- Most important other deviation: govt also uses distortionary taxes to finance public goods and redistribute income.
- Sandmo (1975): optimal tax policy with externalities and a revenue requirement.
- Combination of Ramsey and Pigou problems

Sandmo 1975: Setup

- Denote by $d(x_N)$ the externality cost of consumption of good N
- Let w be the wage rate and $q_i = p_i + \tau_i$ denote post-tax prices.
- Let Z denote non wage income.
- Producer prices fixed; all pre tax prices normalized to 1.
- Individuals have utility functions of the following form:

$$u(x_1, .., x_N, l) - d(x_N)$$

- Utility is maximized subject to:

$$q_1 x_1 + .. + q_N x_N \leq wl + Z$$

Sandmo 1975: Setup

- Individual maximization program

$$L = u(x_1, .., x_N, l) + \lambda(wl + Z - (q_1 x_1 + .. + q_N x_N))$$

- Maximization yields indirect utility $v(q)$.
- Government maximization program:

$$\begin{aligned} \max_q W(q) &= v(q) - d(q) \\ \text{s.t. } \sum \tau_i x_i &\geq R \end{aligned}$$

- Analogous to Ramsey tax problem, but here SWF differs from private sector objective

Sandmo 1975

- Let θ = marginal social welfare gain from \$1 of a lump sum tax and λ = marginal value of relaxing agent's budget constraint
- τ_{ip} = optimal Pigouvian tax rate (when $R = 0$)
  - $\tau_{ip} = 0$ for goods 1 to $N - 1$ and $\tau_{ip} = d'(x_N)$ for good N
- τ_{ir} = optimal Ramsey tax rate (when $d(x_n) = 0$)
- Let τ_i denote optimal tax rate in Sandmo model

Sandmo 1975: Additivity Result

- Main result: can express optimal tax rate as Ramsey rate plus Pigouvian correction.
- Consider case where Slutsky matrix is diagonal (zero cross-price elasticities)
- Then optimal tax on good i, τ_i satisfies

$$\begin{aligned} \frac{\tau_i - \tau_{ip}}{1+\tau_i} &= (\theta/\lambda)/\epsilon^c_{ii} \\ \Rightarrow \tau_i &= \frac{\theta x^c_i}{\lambda} / \frac{dx^c_i}{dp_i} + \tau_{ip} \\ &= \tau_{ip} + \tau_{ir} \end{aligned}$$

Sandmo 1975: Additivity Result

- Useful analytic representation but not an explicit formula for the optimal tax rate
  - Ramsey tax will affect level of cons, which affects optimal Pigouvian tax
  - Conversely, Pigouvian tax will affect optimal Ramsey tax rate
- Qualitative lesson: no justification to tax goods that are complementary to those that produce negative externalities.
  - Just tax fuel, not cars
- Optimal policy is always to directly tax source of the externality
  - Cornaglia and Adda (2003) example of tax on number of cigarettes vs. cotinine levels

Double Dividend Debate

- Claim: gas tax has two "dividends"
  1. discourages pollution, raising social welfare
  2. allows govt. to reduce other distortionary taxes, improving efficiency.
- True if we are at a corner where revenue req. is below level what is generated by optimal Pigouvian taxes.
- More realistic case: already at a Ramsey-tax interior optimum.
- Suppose we discover that production of computers generates negative externality

Double Dividend Debate

- Is there a double dividend from taxing computers?
  - No. Already at Ramsey optimum $\rightarrow$ no efficiency gain from raising taxes on PC's and reducing taxes on other goods
  - Only get single dividend of improving environment
- Obtain double dividend only if taxes on polluting good were initially too low from a Ramsey perspective.
- General lesson: separate externality and optimal second-best tax problems.
- Measure externalities and identify optimal corrective taxes without worrying about other aspects of tax system.

Externalities: Empirical Measurement

- Two approaches
  - Indirect market-based methods
  - Contingent valuation

Edlin and Karaca-Mandic 2006

- Accident externalities from driving automobiles.
- If I drive, I increase probability you will get into an accident $\rightarrow$ externality cost imposed on you
- How to estimate this externality cost and appropriate Pigouvian tax on driving?
- Examine relationship between traffic density and per-capita insurance costs and premiums.
- Look at slope to infer size of externality cost.
- Identification assumption: variation in traffic density at state level not correlated with other determinants of premiums (e.g. types of cars, etc.).

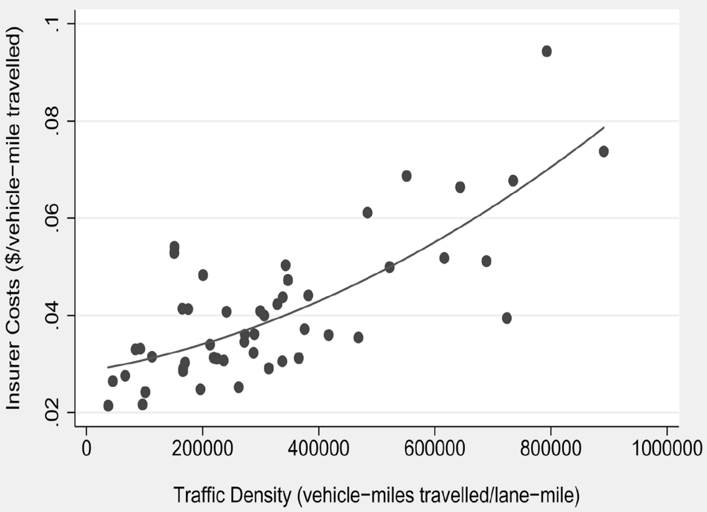


FIG. 1.—Traffic density and insurance costs (1996 dollars)

Source: Edlin and Karaca-Mandic 2006

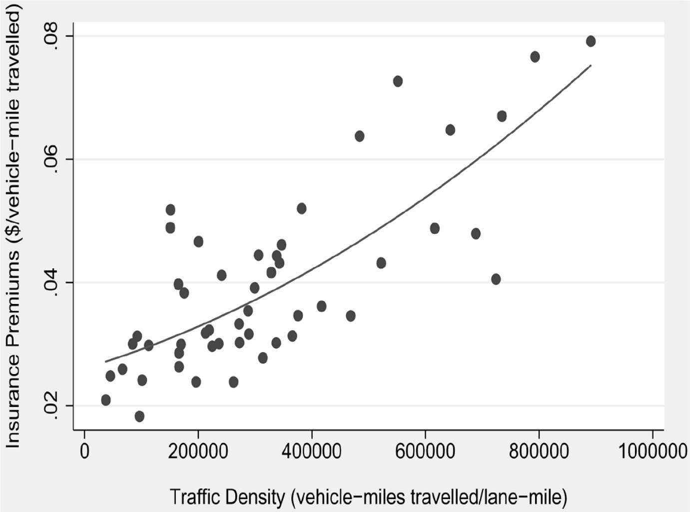


FIG. 2.—Traffic density and insurance premiums (1996 dollars)

Source: Edlin and Karaca-Mandic 2006

TABLE 2
LINEAR INSURANCE RATE MODEL, 1987–95

REGRESSOR	DEPENDENT VARIABLE: INSURER COSTS PER VEHICLE, $\tilde{r}$			DEPENDENT VARIABLE: INSURANCE PREMIUMS PER VEHICLE, r	
	OLS (1995 Only) (1)	OLS (2)	IV (3)	OLS (4)	IV (5)
Traffic density, D	.00042**	.00058**	.0019**	.00036**	.0014**
	(.00009)	(.00029)	(.0009)	(.00018)	(.00067)
State dummy variables	No	Yes	Yes	Yes	Yes
Time dummy variables	No	Yes	Yes	Yes	Yes
Malt alcohol beverages per capita	8.80*	−2.04	.43	.79	2.80
	(4.54)	(5.63)	(5.87)	(2.44)	(2.99)
Real gross product per capita	6,535.20*	5,373.50	2,224.50	2,463.41	−113.00
	(3,779.90)	(3,985.50)	(4,866.50)	(2,388.40)	(3,245.50)
Hospital cost	.16**	−.30**	−.40**	.02	−.05
	(.08)	(.12)	(.15)	(.05)	(.07)
% young male population	30.99	−4.98	−.75	8.18	11.64
	(27.83)	(14.52)	(14.92)	(8.13)	(9.45)
Precipitation	1.90**	.10	.06	−.49*	−.53*
	(.92)	(.36)	(.37)	(.29)	(.33)
Snowfall	.32	.01	−.07	−.12	−.19
	(.41)	(.22)	(.23)	(.12)	(.14)
No-fault	95.02**	150.11**	175.07**	95.87**	116.29**
	(32.08)	(17.04)	(28.06)	(8.80)	(18.40)
Add-on	−1.35	210.06**	251.52**	139.60**	173.52**
	(37.59)	(51.66)	(58.46)	(39.48)	(43.74)
R^2	.73	.92	.91	.97	.96

Source: Edlin and Karaca-Mandic 2006

Edlin and Karaca-Mandic 2006

- Conclude that traffic density substantially increases insurer costs, and that relationship is convex
  - Increase in traffic density from average driver has external cost of $2,000 per year in California
  - Comparable figure in $10 per year in North Dakota
- Suggests that insurance premiums should be doubled in CA to achieve social optimum

Brookshire et al. 1982

- Infer willingness to pay for clean air using effect of pollution on property prices (capitalization)

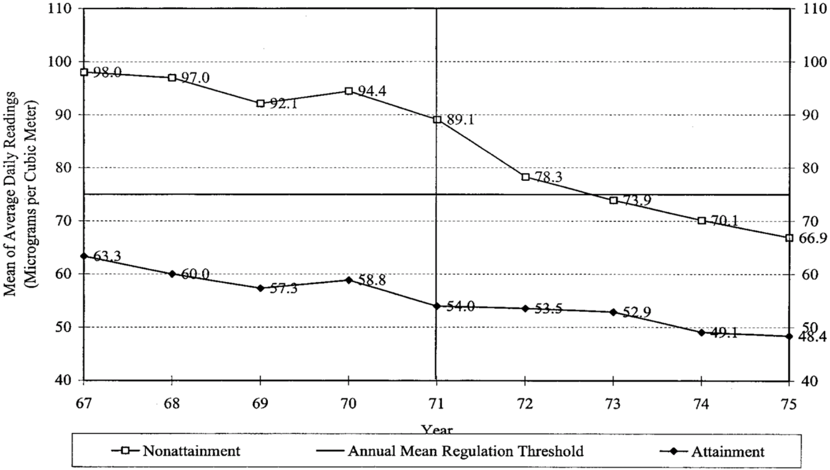
- Compare prices of houses in polluted vs non-polluted areas.

$$P_i = \alpha + Pollution_i + X_i\beta + \epsilon_i$$

- Problems
  - Omitted variable bias: polluted neighborhoods worse on many dimensions
  - Sorting: people with allergies avoid polluted areas

Chay and Greenstone 2005

- Also study home prices but use Clean Air Act as an exogenous change in pollution.
- Clean Air Act: imposed ceilings on pollution levels by county in mid 1970s.
- High pollution counties experience sharp reductions in pollution levels relative to low pollution counties



Source: Chay and Greenstone 2005

TABLE 4
ESTIMATES OF THE IMPACT OF MID-DECADE TSPS NONATTAINMENT ON 1970–80 CHANGES IN TSPS POLLUTION AND LOG HOUSING VALUES

	(1)	(2)	(3)	(4)
	A. Mean TSPs Changes			
TSPs nonattainment in 1975 or 1976	−9.96 (1.78)	−10.41 (1.90)	−9.57 (1.94)	−9.40 (2.02)
F-statistic TSPs nonattainment*	31.3 (1)	29.9 (1)	24.4 (1)	21.5 (1)
R^2	.04	.10	.19	.20
	B. Log Housing Changes			
TSPs nonattainment in 1975 or 1976	.036 (.012)	.022 (.009)	.026 (.008)	.019 (.008)
F-statistic TSPs nonattainment*	8.5 (1)	6.2 (1)	9.3 (1)	6.4 (1)
R^2	.01	.56	.66	.73
County Data Book covariates	no	yes	yes	yes
Flexible form of county covariates	no	no	yes	yes
Region fixed effects	no	no	no	yes
Sample size	988	983	983	983

Source: Chay and Greenstone 2005

Chay and Greenstone 2005

- Conclusion: 1% increase in pollution $\rightarrow$ 0.25% decline in house values
- Clean air act increased house values by $45 bil (5%) in treated counties
- Conceptual concern with short-run market-based methods: people may not be fully aware of changes in pollution

Glaeser and Luttmer 2003

- Quantify efficiency costs of pure command-and-control solutions instead of price/tradeable permit mechanisms
- Study allocation of apartments under rent control
- Traditionally assume that with price controls, still have allocative efficiency.
- But regulation will generally lead to allocative inefficiency that generates first-order welfare losses.
    - For small price caps, allocation inefficiency dwarfs undersupply inefficiency.

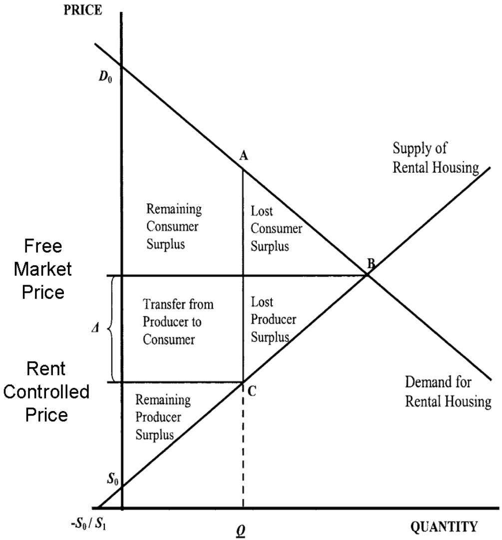


Figure 1. Classical Analysis of Welfare Losses from Rent Control

Source: Glaeser and Luttmer 2003

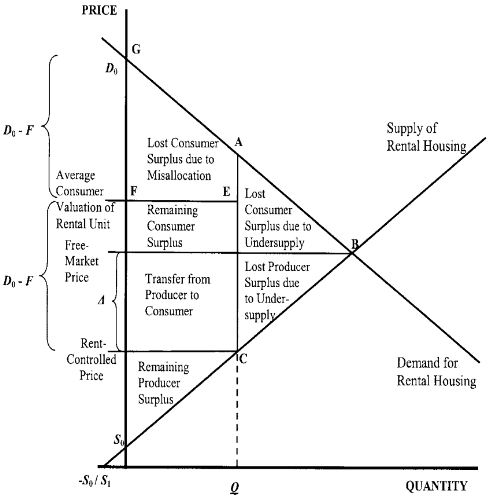


Figure 2. The Welfare Losses from Rent Control When Apartments are Randomly Allocated Across Consumers

Source: Glaeser and Luttmer 2003

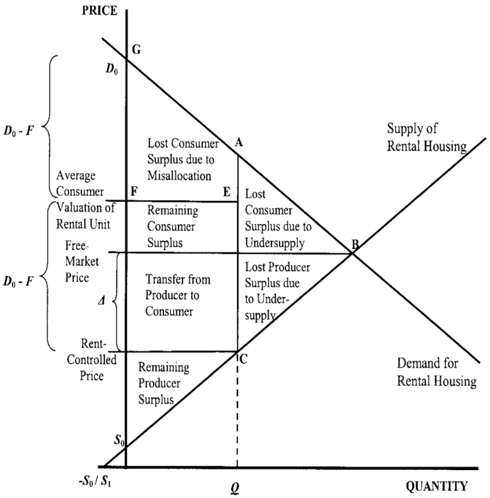


FIGURE 2. THE WELFARE LOSSES FROM RENT CONTROL WHEN APARTMENTS ARE RANDOMLY ALLOCATED ACROSS CONSUMERS

Source: Glaeser and Luttmer 2003

Glaser and Luttmer 2003

- Quantify welfare losses from misallocation by comparing consumption patterns in rent-controlled (NYC) and free-market places across demographic groups.
- Predict apartment size using number in family, income, education, age, etc. using 105 large MSAs
- Test if actual apartment allocations in NYC match predictions
- Identifying assumption: preferences stable across MSAs
- Check: placebo tests using Chicago and Hartford

TABLE 3—ACTUAL AND EFFICIENT ALLOCATION

Actual apartment size (number of rooms):	Allocation of New York City households across apartments (fraction of households) — Efficient apartment size (number of rooms): 1	2	3	4	5	Marginal
1	0.0688	0.0206	0.0000	0.0000	0.0000	0.0894
2	0.0204	0.0766	0.0375	0.0000	0.0000	0.1345
3	0.0002	0.0373	0.2667	0.0465	0.0001	0.3508
4	0.0000	0.0000	0.0466	0.2126	0.0243	0.2835
5	0.0000	0.0000	0.0000	0.0243	0.1175	0.1418
Marginal	0.0894	0.1345	0.3508	0.2835	0.1418	1.0000

Note: The table shows the joint distribution of the actual and efficient allocation of households to apartments in the baseline treatment group (10,000 households in New York City).

Source: Glaeser and Luttmer 2003

Contingent Valuation

- For some outcomes, it is impossible to have a market value
  - Ex: protecting endangered species
- Common solution: “contingent valuation” surveys
  - How much would you be willing to pay to avoid extinction of whales?

Diamond and Hausman 1994

- Describe problems with contingent valuation using surveys
- No resource cost to respondents
- Lack of consistency in responses
  - Framing Effects: whales then seals vs. seals then whales
  - WTP to clean one lake = WTP to clean 5 lakes
- Diamond and Hausman: let experts decide based on a budget voted on by individuals for the environment instead of relying on valuation

Behavioral Economics Applications: Internalities

- Sin taxes intended to correct "internalities."
- Internal costs of smoking cigarettes dwarf the external costs.
- Suggests that conventional Pigouvian taxation should be small (relative to actual taxes observed on e.g. cigarettes and alcohol).
- Q: Does addictive nature of cigarettes motivate taxation?
    - A: Highly sensitive to positive model of addiction
    - Challenge: difficult to determine which model is right empirically

Becker and Murphy 1988

- Show that addictive goods can be modeled in perfectly rational framework.
- Dynamic model with habit formation.
- Current consumption of the addictive good decreases long-run utility but increases marginal utility of consumption tomorrow:
    - If discount rate high enough, rationally choose to become addicted.
- Implication: no reason for special taxes on these goods; set taxes according to Ramsey rules.

Gruber and Koszegi 2004

- Hyperbolic discounting preferences for smokers

$$U_0 = u(c_0) + \beta(\sum_{t\geq 1} \gamma^t u(c_t)) \text{ with } \beta < 1.$$
$$U_1 = u(c_1) + \beta(\sum_{t\geq 2} \gamma^t u(c_t))$$

- Planner maximizes U_0 with $\beta = 1$ (true utility).
- Individuals overconsume c: fail to take full account of harm to future selves.
- Taxes reduce demand for each self; can partly correct the internality.
- Calibration implies corrective tax should be very large.

Bernheim and Rangel 2004

- Model of “cue-triggered” addiction. Two selves:
  - Cognitive self with rational preferences
  - Visceral brain triggered by random cues in which addictive good is consumed at any cost.
- Probability of trigger increases with past consumption levels.
- Ideal policy: only allow rational consumption, eliminate consumption in hot mode.
- Corrective taxation may not be desirable: only distorts consumption in rational state, not visceral state.
- Better solution: regulated dispensation – must place orders one period in advance

O'Donoghue and Rabin 2006

- Studies optimal sin taxes in a model with two types of consumers: rational and those who overconsume (e.g., because of self-control problems)
- Can be thought of as a hybrid of Becker and Gruber-Koszegi models
- Key result: irrationality among a few consumers leads to substantial role for corrective taxation/subsides.
  - For rational individuals, excess burden due to taxation is second-order (Harberger triangle).
  - For irrational individuals, welfare gains from correction of internality is first-order (Harberger trapezoid).
  - Therefore always optimal to have a positive tax; calibrations suggest fairly large corrective taxes