

Public Economics Lectures

Part 1: Introduction

Raj Chetty and Gregory A. Bruich

Harvard University
Fall 2012

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 → 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, labor laws, min education laws, environment
- 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
 - Romney: replacing Medicare with decentralized private insurance will improve health outcomes and reduce costs
 - Obama: Romney's proposal will worsen health outcomes and raise costs
- Only one of these views can be correct
 - Injecting science into these debates has great practical value

Motivation 2: Academic Interest

- Public economics is typically the end point for many other subfields
- Macro, development, labor, and corporate finance questions often ultimately motivated by a public economics question
 - 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 ensure that you work on relevant topics

Motivation 3: Methodology

- Public economics is at the frontier of a methodological transformation in applied microeconomics
- Data-driven approach to answering important policy questions
 - Combines a broad set of skills: applied theory, applied econometrics, simulation methods
 - Useful skill set for many applied fields in economics
- Topics in the course reflect a broad set of methodological themes

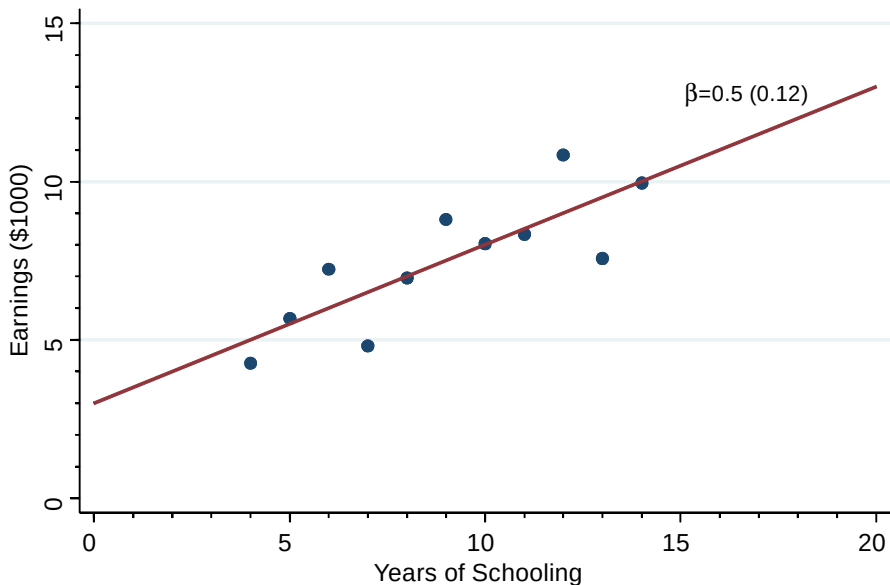
Theme 1: Connecting Theory to Data

- 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?
- Traditional approach: theoretical models and numerical simulations
 - Theory often makes weak predictions: optimal tax rate between 0 and 100%
 - Numerical simulations rely on strong assumptions
- Recent work derives robust formulas that can be implemented using well-identified empirical estimates

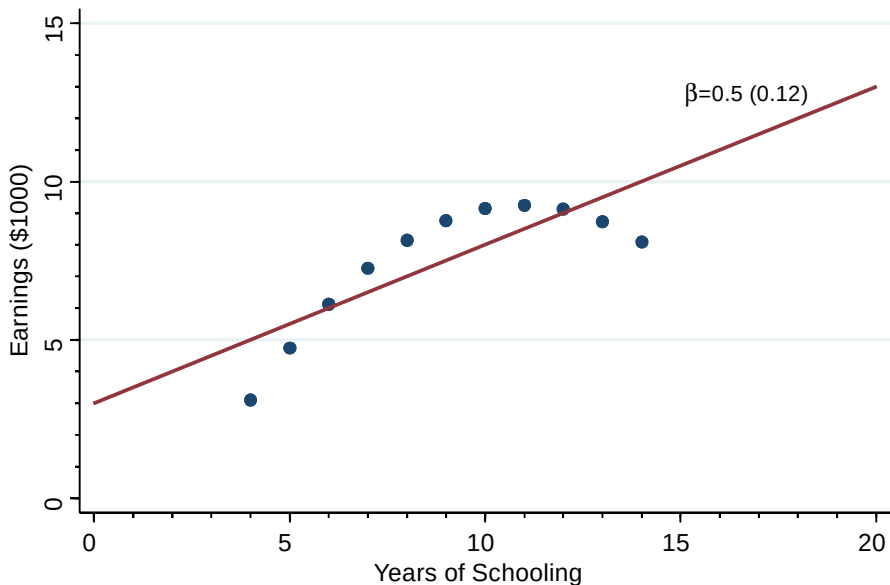
Theme 2: Quasi-Experimental Empirical Methods

- Research in public economics exploits a variety of quasi-experimental research designs to identify parameters of interest
 - Event studies, regression discontinuity, synthetic control
- Good way to learn practical lessons in applied econometrics
 - What is “identification by functional form” and why is it undesirable?
 - Is the LATE or ATE of greater interest in your problem?
 - When is propensity score reweighting credible?
 - When do weak instrument problems arise and how can they be fixed?
- Emphasis on non-parametric graphical techniques rather than parametric regression models

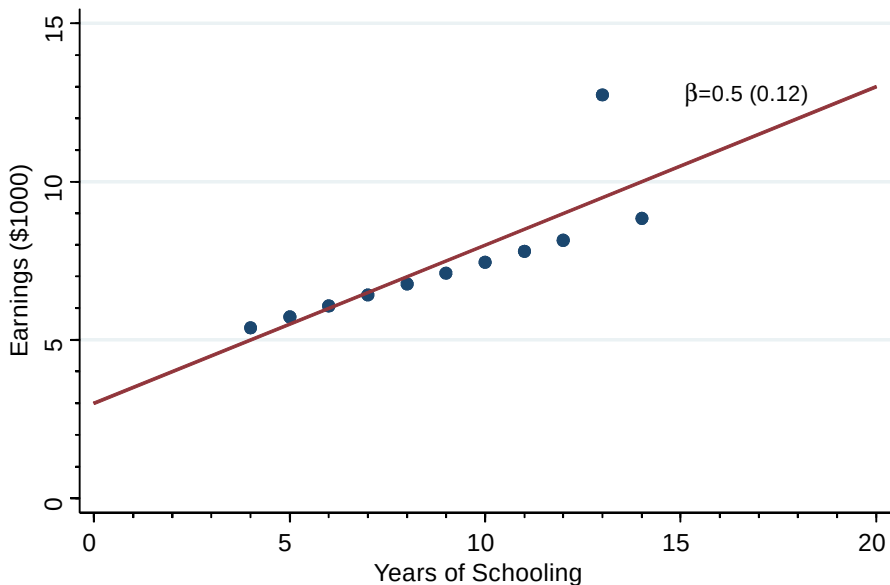
Anscombe (1973): Dataset 1



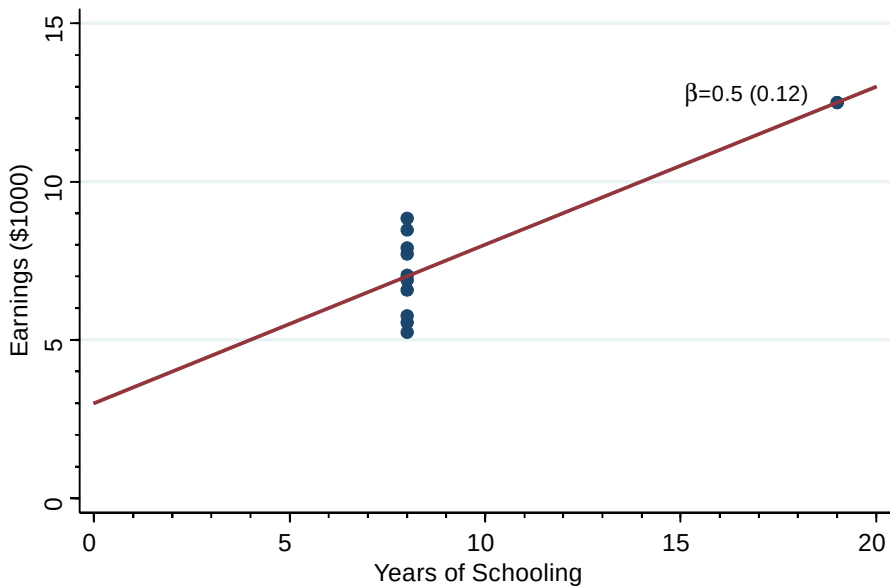
Anscombe (1973): Dataset 2



Anscombe (1973): Dataset 3



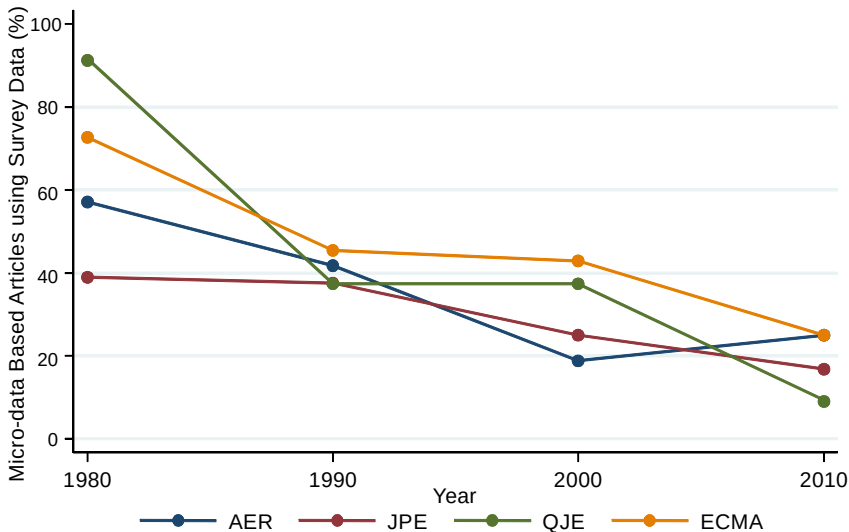
Anscombe (1973): Dataset 4



Theme 3: “Big Data”

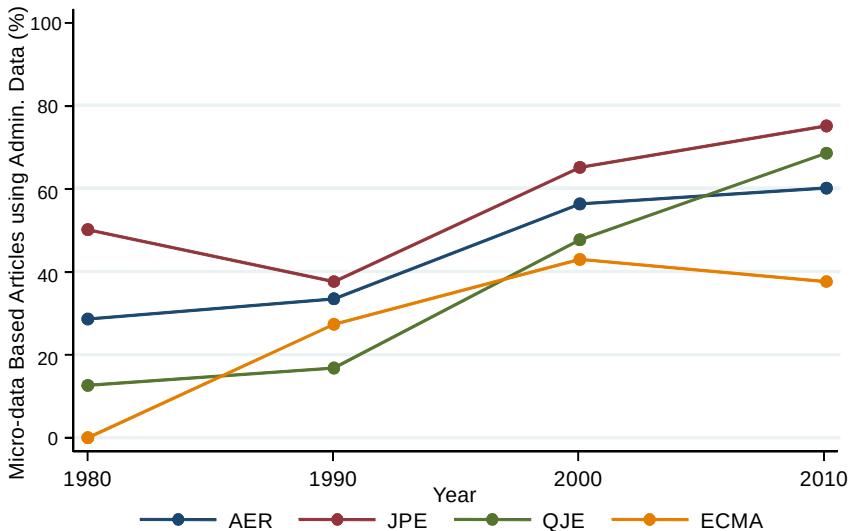
- Compelling implementation of quasi-experimental methods requires a lot of data
- Recent availability of very large datasets has transformed research in applied microeconomics
 - Scanner data on consumer purchases
 - Tax and social security records
 - School district databases

Use of Pre-Existing Survey Data in Publications in Leading Journals, 1980-2010



Note: "Pre-existing survey" datasets refer to micro surveys such as the CPS or SIPP and do not include surveys designed by researchers for their study. Sample excludes studies whose primary data source is from developing countries.

Use of Administrative Data in Publications in Leading Journals, 1980-2010



Note: "Administrative" datasets refer to any dataset that was collected without directly surveying individuals (e.g., scanner data, stock prices, school district records, social security records). Sample excludes studies whose primary data source is from developing countries.

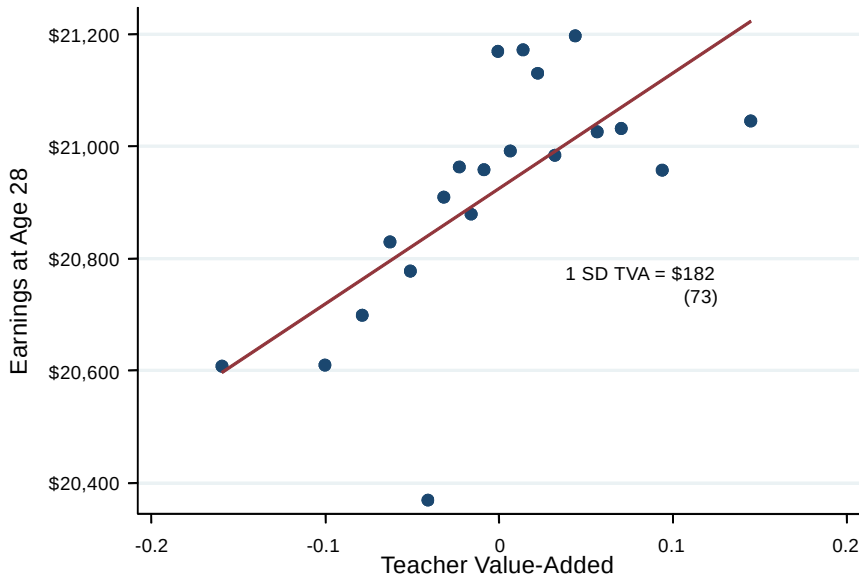
United States Tax Data

- 7 billion tax records covering full pop. from 1996 to today
- Includes a rich set of information on individuals
 - Earnings from W-2's (covers non-filers)
 - Employer ID
 - College attendance
 - Retirement savings, charitable contributions
 - Housing and mortgage interest
 - Geographical location
 - Birth, death, marriage, children, family structure
- Analogous corporate databank contains information for 5 million firms per year, linked to workers

What are the Benefits of Administrative Data?

- ① Higher quality information: virtually no missing data or attrition
 - Current Population Survey non-response rate now 31% for income
- ② Longitudinal tracking over long periods
 - Match rates of 95% over 20+ years in studies of long-term impacts of early childhood education [Chetty et al. 2011, Chetty, Friedman, Rockoff 2012]

Earnings at Age 28 vs. Teacher Value-Added in Elementary School

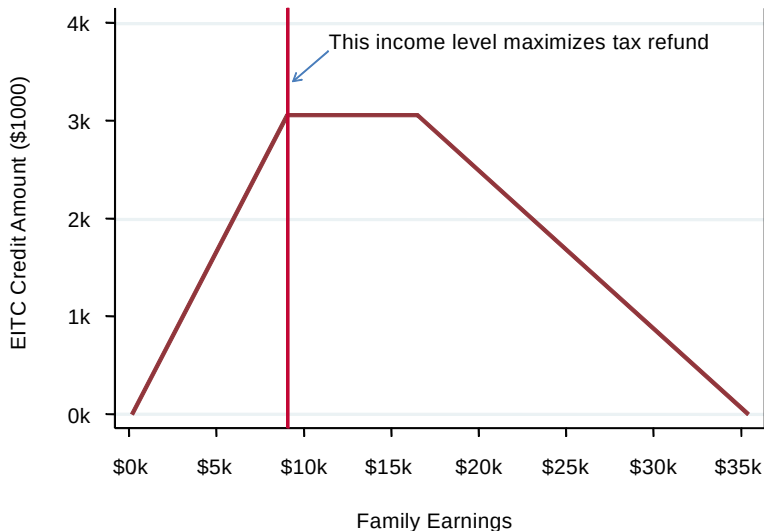


Source: Chetty, Friedman, Rockoff 2012

What are the Benefits of Administrative Data?

- ① Higher quality information: virtually no missing data or attrition
 - Current Population Survey non-response rate now 31% for income
- ② Longitudinal tracking over long periods
 - Match rates of 95% over 20+ years in studies of long-term impacts of early childhood education [Chetty et al. 2011, Chetty, Friedman, Rockoff 2012]
- ③ Very large sample sizes: 2,000 times the size of the CPS
 - Can develop new research designs

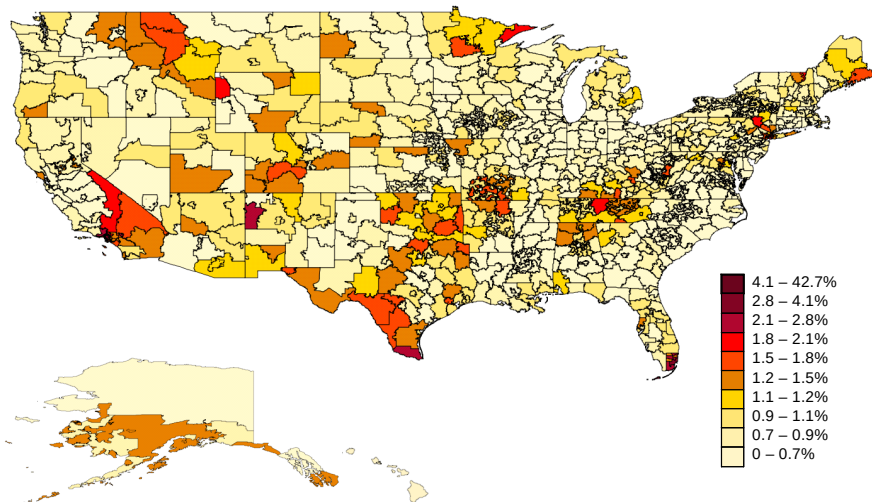
Earned Income Tax Credit Schedule for Single Earners with One Child in 2008



U.S. Income Distributions for EITC-Eligible Individuals with Children in 2008

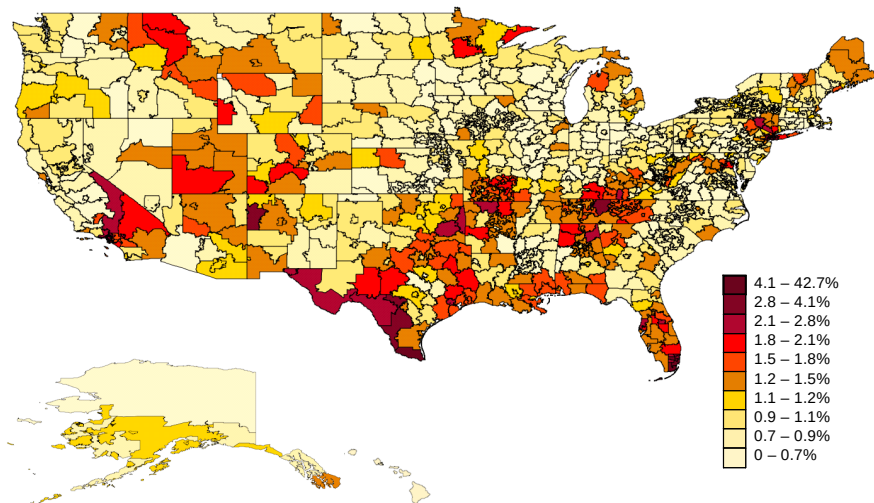


Fraction of Tax Filers Who Report Income that Maximizes EITC Refund in 1996



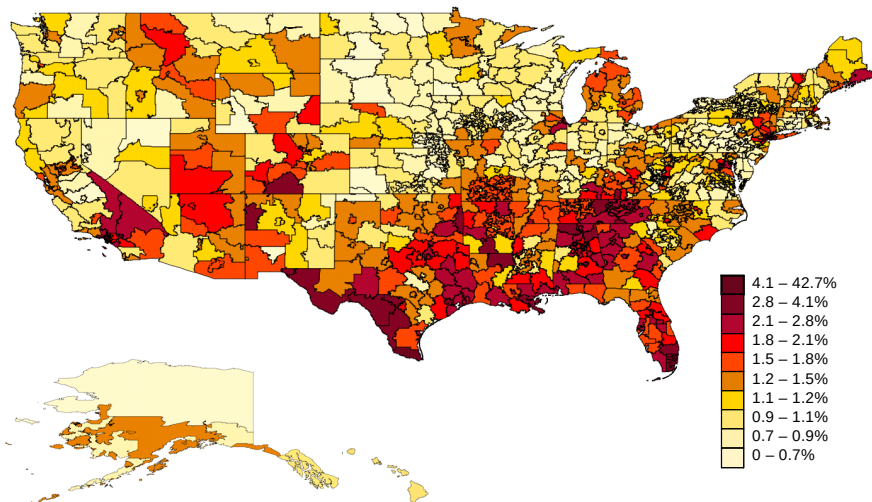
Source: Chetty, Friedman, Saez 2012

Fraction of Tax Filers Who Report Income that Maximizes EITC Refund in 1999



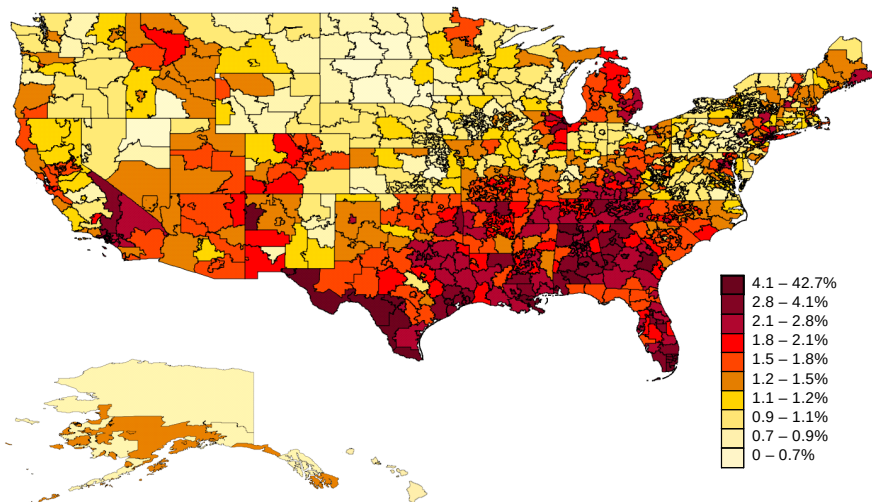
Source: Chetty, Friedman, Saez 2012

Fraction of Tax Filers Who Report Income that Maximizes EITC Refund in 2002



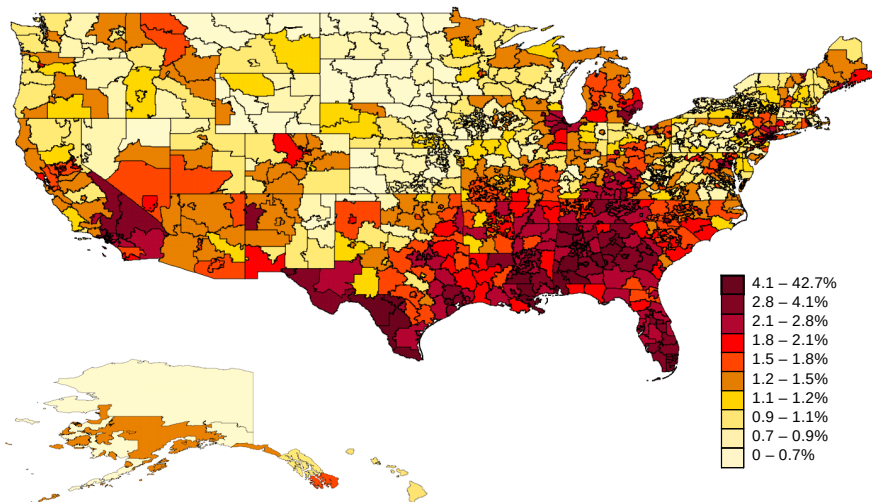
Source: Chetty, Friedman, Saez 2012

Fraction of Tax Filers Who Report Income that Maximizes EITC Refund in 2005



Source: Chetty, Friedman, Saez 2012

Fraction of Tax Filers Who Report Income that Maximizes EITC Refund in 2008

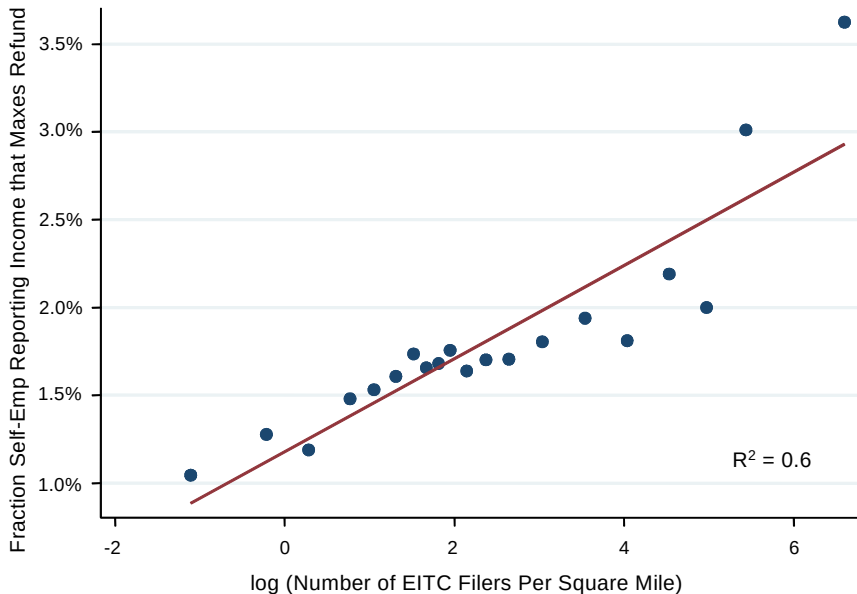


Source: Chetty, Friedman, Saez 2012

Theme 4: Behavioral Models

- Recent work in public economics draws on insights from psychology and economics literature
 - Strong evidence that individuals fail to optimize
- Raises new policy questions
- Suggests new policy instruments
 - E.g. information, social incentives

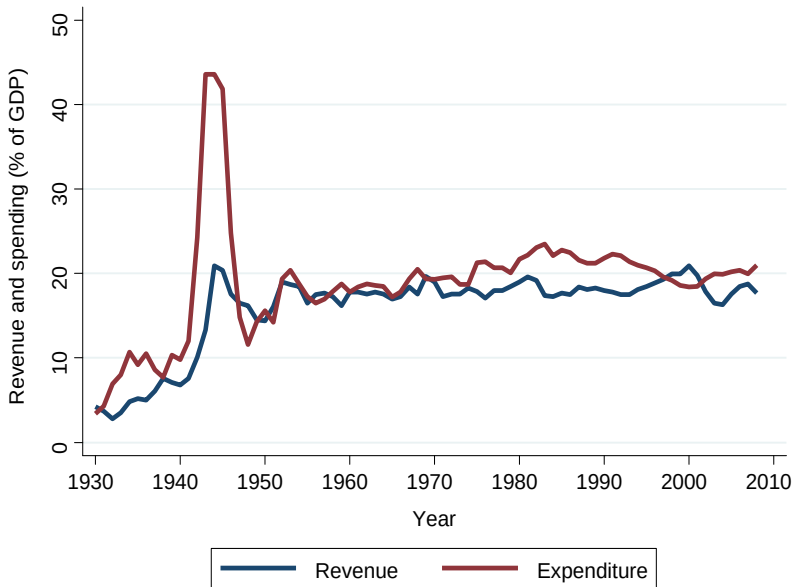
Correlation Between Response to EITC and EITC Filer Density by ZIP Code



Background Facts: Size and Growth of Government

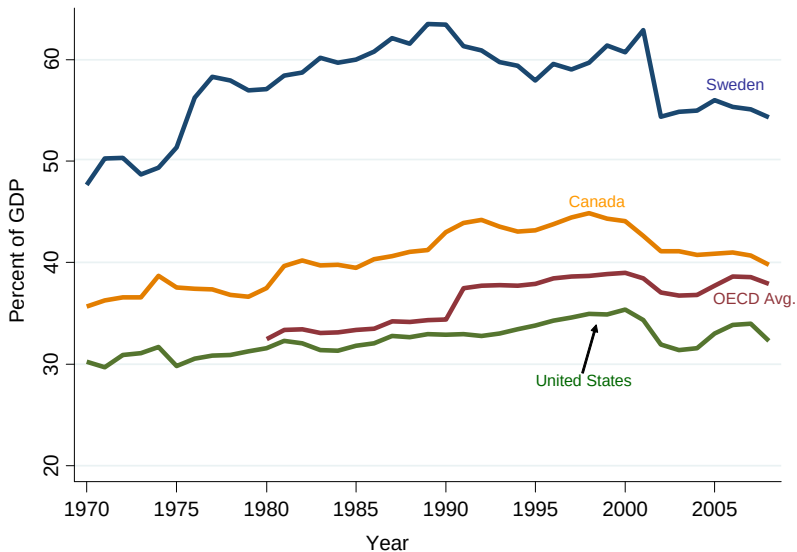
- Government expenditures = $1/3$ GDP in the U.S.
- Higher than 50% of GDP in some European countries
- Decentralization is a key feature of U.S. govt
 - One third of spending (10% of GDP) is done at state-local level (e.g. schools)
 - Two thirds (20% of GDP) is federal

Federal Government Revenue and Expenditure 1930-2009



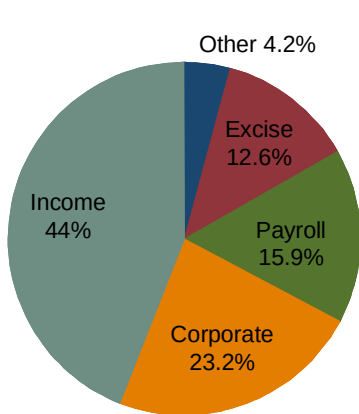
Source: Office of Management and Budget, Historical Tables, FY 2011

Total Government Spending by Country, 1970-2007

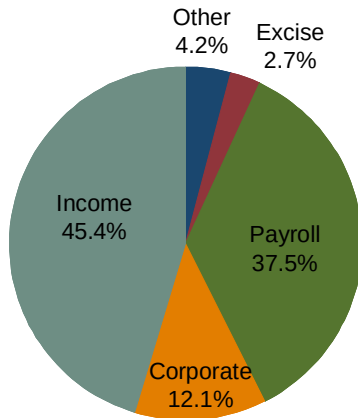


Source: OECD Economic Outlook (2009)

Federal Revenues (% of total revenue)



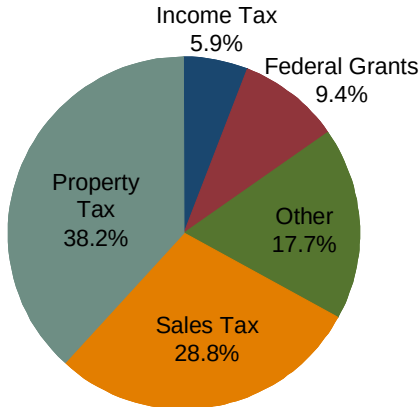
1960



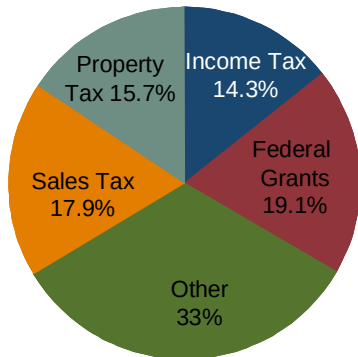
2008

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

State/Local Revenues (% of total revenue)



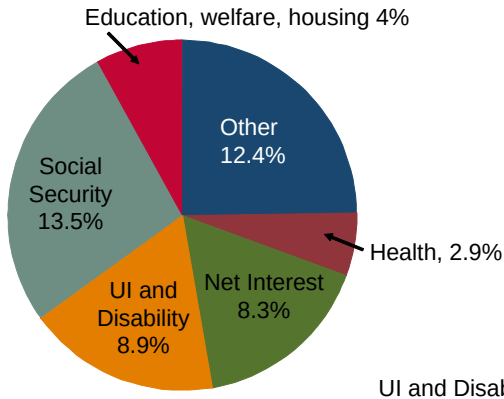
1960



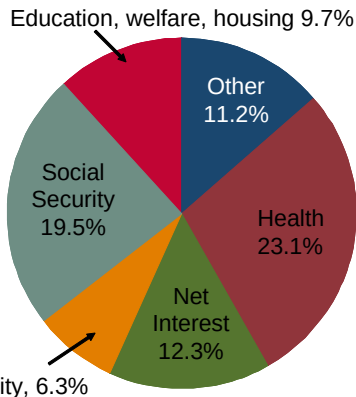
2007

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

Federal Spending (% of total spending)



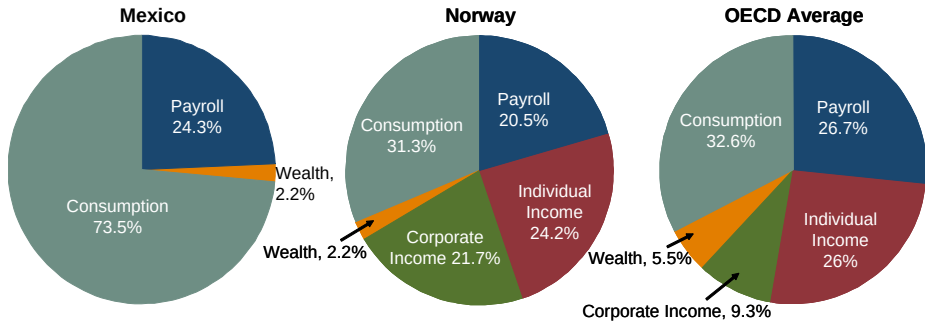
1960



2001

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

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

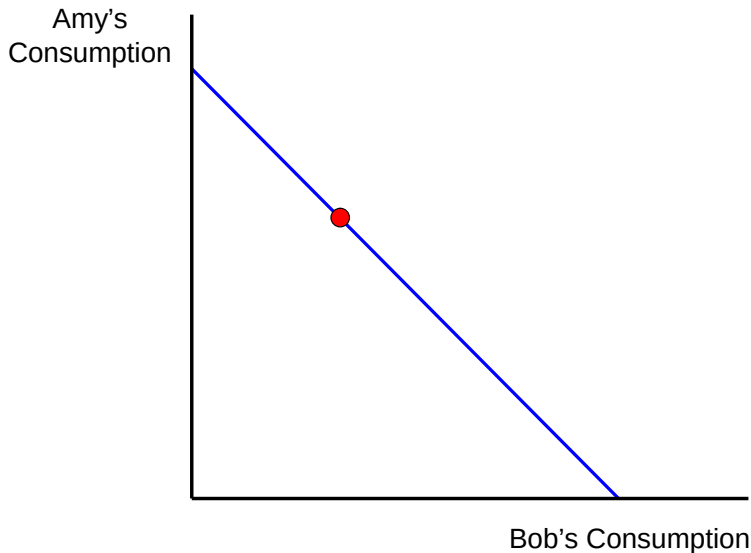


Source: OECD 2002

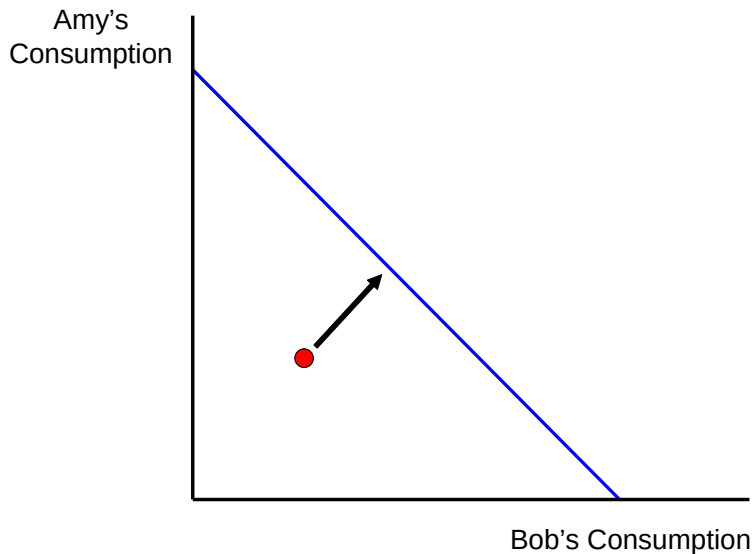
Government Intervention in the Economy

- Organizing 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 redistributive concerns?

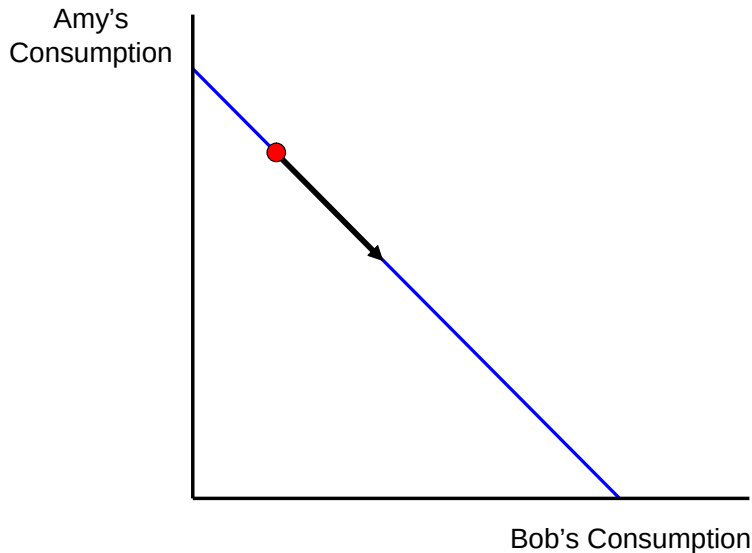
Efficient Private Market Allocation of Goods



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 → 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

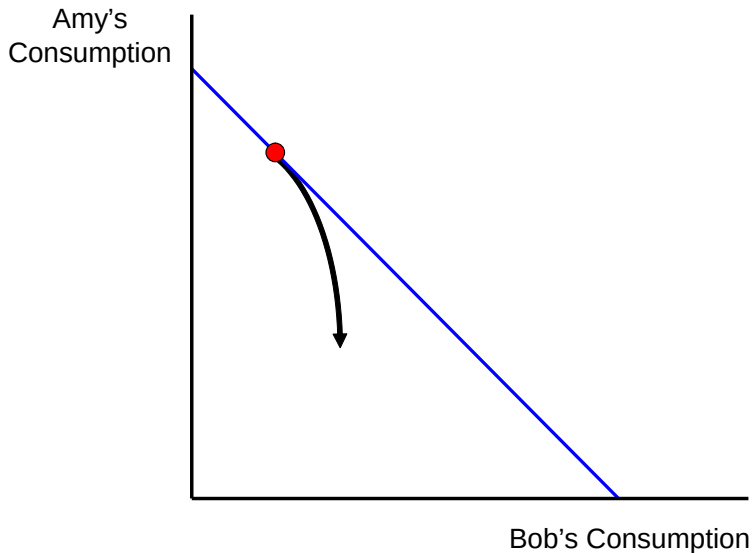
- If agents do not optimize, government intervention (e.g. by forcing saving via social security) may be desirable
- This is an “individual” failure rather than a market failure
- Conceptual challenge: how to avoid paternalism critique
 - Why does govt. know better what's desirable for you (e.g. wearing a seatbelt, not smoking, saving more)
- Difficult but central issues for optimal policy design

- 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 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 distort incentives (behavioral responses in private sector)
- In practice, there are sharp tradeoffs between costs and benefits of government intervention

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 Corporate Taxation
- 5 Social Insurance
- 6 Public Goods and Externalities

Public Economics Lectures

Part 2: Incidence of Taxation

Raj Chetty and Gregory A. Bruich

Harvard University
Fall 2012

- 1 Definition and Introduction
- 2 Partial Equilibrium Incidence
- 3 Partial Equilibrium Incidence with Salience Effects
- 4 Partial Equilibrium Incidence: Empirical Applications
- 5 General Equilibrium Incidence
- 6 Capitalization
- 7 Mandated Benefits

References on Tax Incidence

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

- 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 → 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 $\rightarrow$ 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)

- 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

- 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 $\rightarrow$ 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 or 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 $\varepsilon_D = \frac{\partial D}{\partial q} \frac{q}{D(q)} = \frac{\partial \log D}{\partial \log q}$ 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

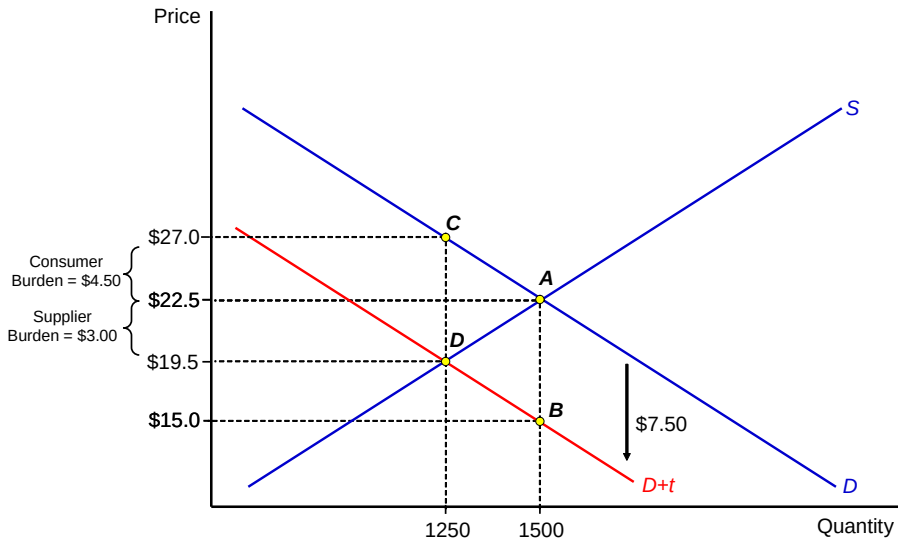
- Equilibrium condition

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

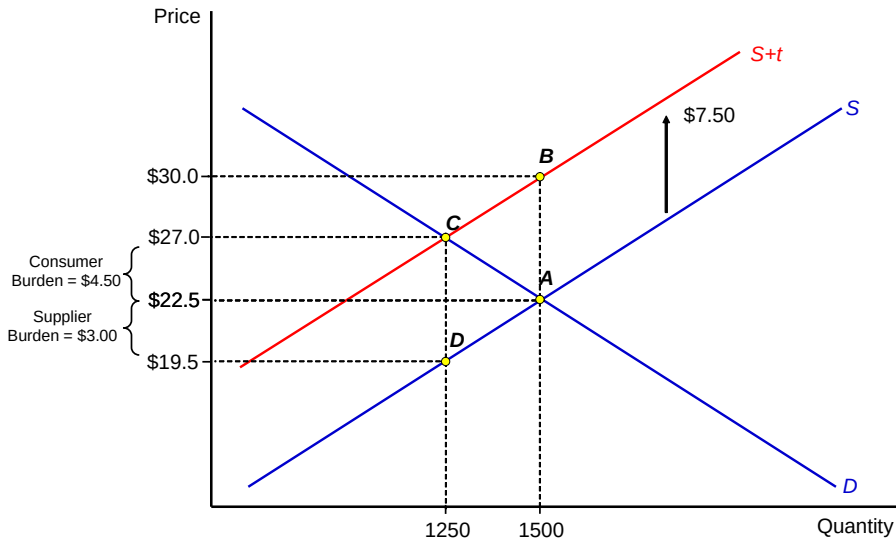
defines an equation $p(t)$

- 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

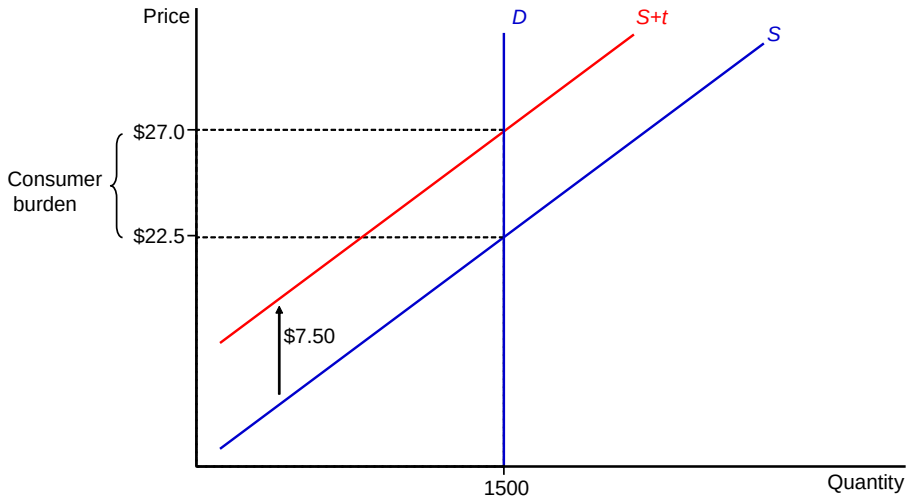
Tax Levied on Consumers



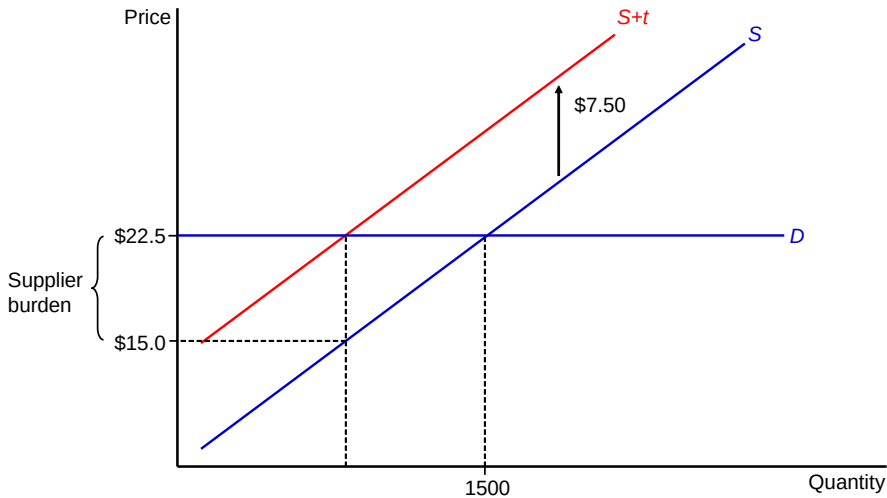
Tax Levied on Producers



Perfectly Inelastic Demand



Perfectly Elastic Demand



Formula for Tax Incidence

- Implicitly differentiate equilibrium condition

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

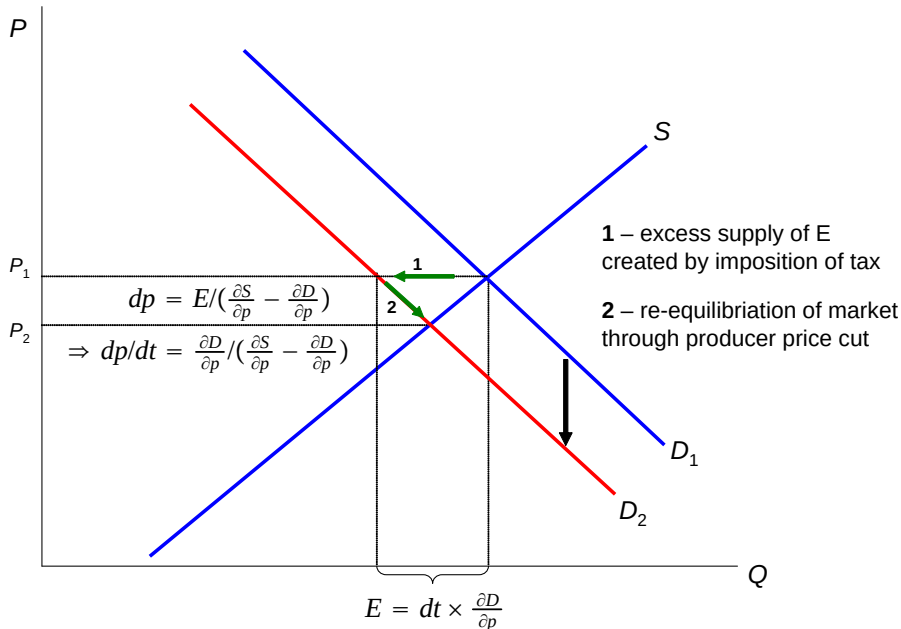
to obtain:

$$\begin{aligned}\frac{dp}{dt} &= \frac{\partial D}{\partial p} \frac{1}{\left(\frac{\partial S}{\partial p} - \frac{\partial D}{\partial p}\right)} \\ \Rightarrow \frac{dp}{dt} &= \frac{\varepsilon_D}{\varepsilon_S - \varepsilon_D}\end{aligned}$$

- 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

- 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 $q = (1 + \tau)p$
- Let demand for good x be denoted by $x(p, \tau)$

- 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:

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

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 $\varepsilon_{x,1+\tau}$

$$\theta = \varepsilon_{x,1+\tau} / \varepsilon_{x,p}$$

- 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)

Documenting Salience Mechanism

- Concern with posting tax inclusive prices: may have influenced behavior through various channels besides salience
- Common problem in field experiments termed “Hawthorne effects”
- Difficult to rule out all mechanisms, but helpful to present evidence that mechanism of interest is very powerful

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)

- 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

Period	TREATMENT STORE		
	<u>Control Categories</u>	<u>Treated Categories</u>	<u>Difference</u>
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	<u>Control Categories</u>	<u>Treated Categories</u>	<u>Difference</u>
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)

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

Source: Chetty, Looney, and Kroft (2009)

Effect of Posting Tax-Inclusive Prices: Mean Quantity Sold

TREATMENT STORE			
Period	<u>Control Categories</u>	<u>Treated Categories</u>	<u>Difference</u>
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)

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

DDD Estimate
-2.20
(0.58)

Source: Chetty, Looney, and Kroft (2009)

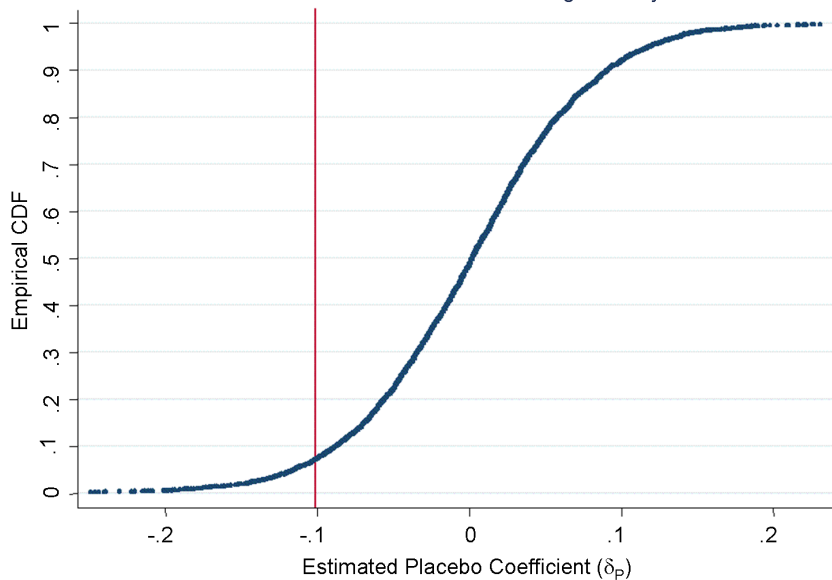
Statistical Significance and Permutation Tests

- Standard error computations always require specific assumptions about error structure
 - Ex: allow for correlation in purchases across products within a store-week-category cell
 - Standard parametric approach is to cluster standard errors by store-week-category
- But appropriate level of clustering is often unclear
 - Should we also allow for correlation across stores or categories?

Non-Parametric Permutation Tests

- Useful technique for inference with correlated errors: permutation (Fisher's exact) test
 - Pretend intervention occurred in each of the other cells (store, week, category) of the sample and recompute DDD estimate
 - Calculate where actual treatment effect lies in empirical CDF of placebo treatment effects

Figure 1
Distribution of Placebo Estimates: Log Quantity



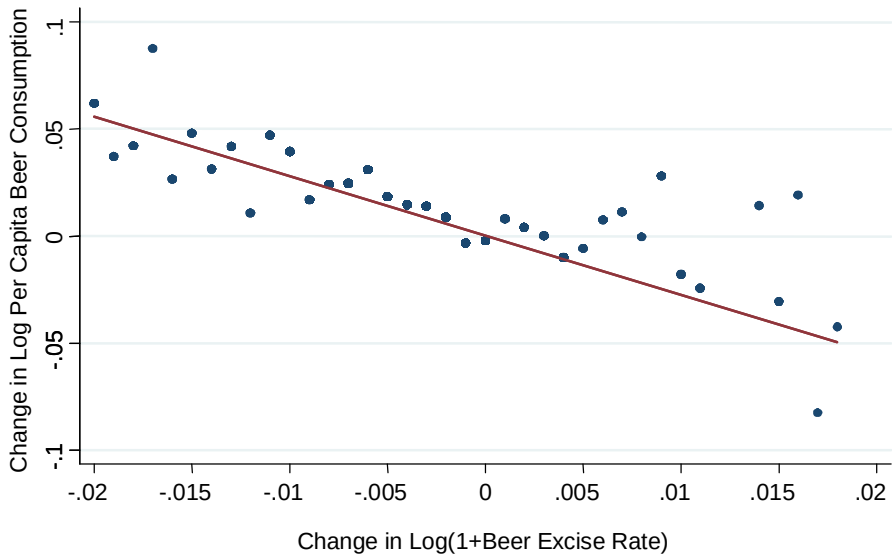
Source: Chetty, Looney, and Kroft (2009)

Non-Parametric Permutation Tests

- Key assumption underlying permutation test: treatment is truly random
 - Probability of treatment cannot vary across cells
- True by construction in experiment
- But may not hold in settings where policy changes are endogenous
 - Ex: unemployment benefits increased during recessionary periods
 - Nevertheless, often a useful benchmark

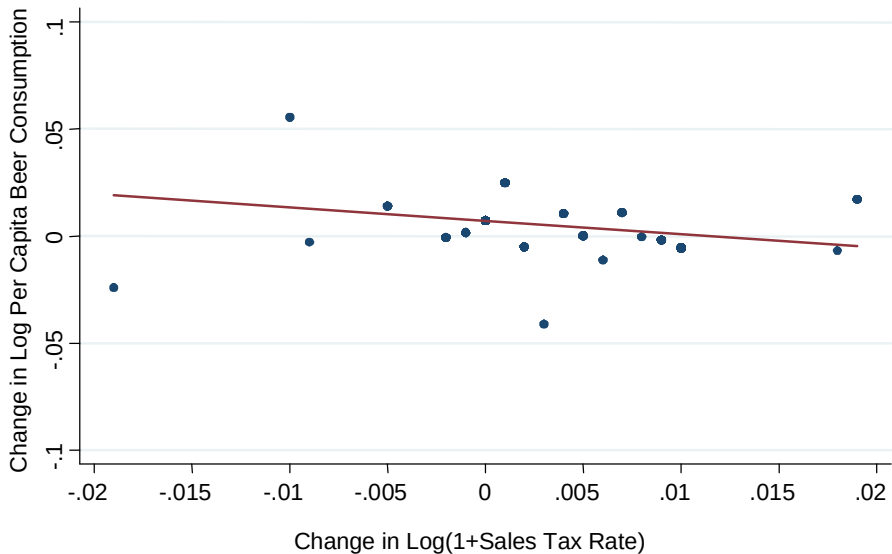
- Compare effects of price changes and tax changes
- Alcohol subject to two state-level taxes in the U.S.:
 - Excise tax: included in price
 - Sales tax: added at register, not shown in posted price
- Exploiting state-level changes in these two taxes, estimate θ
 - Addresses concern that experiment may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

Per Capita Beer Consumption and State Beer Excise Taxes



Source: Chetty, Looney, and Kroft (2009)

Per Capita Beer Consumption and State Sales Taxes



Source: Chetty, Looney, and Kroft (2009)

Effect of Excise and Sales Taxes on Beer Consumption

Dependent Variable: $\Delta \text{Log}(\text{per capita beer consumption})$

	Baseline	Bus Cyc, Alc Regs.	3-Year Diffs	Food Exempt
	(1)	(2)	(3)	(4)
$\Delta \text{Log}(1+\text{Excise Tax Rate})$	-0.87 (0.17) ^{***}	-0.89 (0.17) ^{***}	-1.11 (0.46) ^{**}	-0.91 (0.22) ^{***}
$\Delta \text{Log}(1+\text{Sales Tax Rate})$	-0.20 (0.30)	-0.02 (0.30)	-0.00 (0.32)	-0.14 (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:

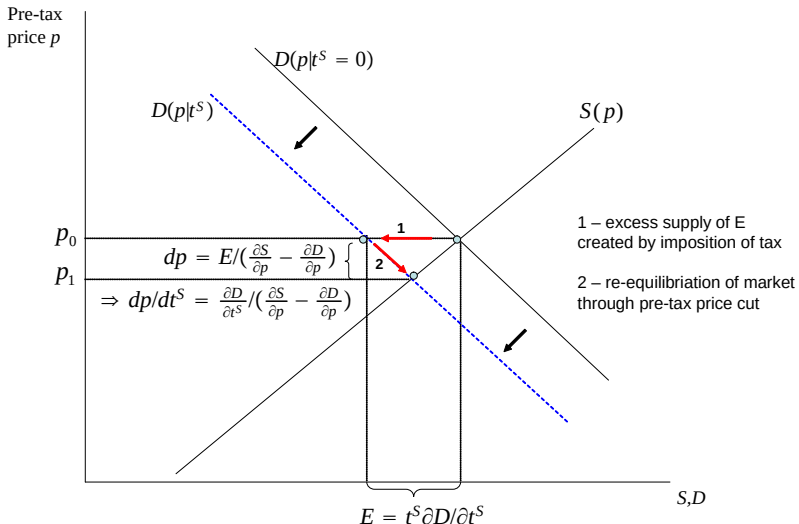
$$(p + t)x(p, t, Z) + y(p, t, Z) = Z$$

- 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 Saliency Effects



Source: Chetty, Looney, and Kroft (2009)

Tax Incidence with Salience Effects: Formula

- Incidence on producers of increasing t is

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

- Incidence on producers attenuated by θ
 - 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

Tax Incidence: Empirical Applications

- ① [Evans, Ringel, and Stech 1999]: Cigarette excise taxes
- ② [Hastings and Washington 2010]: Food stamps
- ③ [Rothstein 2010]: 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: tradeoff between quality of research design and importance/novelty

Developing Empirical Research Designs

- All of the empirical studies we will consider here start by formulating clear research designs
- Why develop an explicit design rather than simply use all available variation in tax rates?
- Consider estimating effect of a treatment (e.g., tax) T on outcome y

$$y_i = \alpha + \beta T_i + \varepsilon_i$$

- Treatment is assigned based on a “selection” model

$$T_i = \alpha_T + \beta_T X_i + \eta_i$$

- Treatment may be non-random: $\text{cov}(X_i, \varepsilon_i) \neq 0$, $\text{cov}(\eta_i, \varepsilon_i) \neq 0$

Developing Empirical Research Designs

- Traditional approach to accounting for confounding factors or selection: control for observables X_i when estimating treatment effect

$$y_i = \alpha + \beta T_i + \gamma X_i + \varepsilon_i$$

- Can be done using OLS regression, matching, propensity-score reweighting, etc.
- Problem with these approaches: don't know source of variation in T_i
 - Must be some reason that one person got treated and another did not even if they are perfectly matched on observables (e.g., twins)
 - η_i must be correlated with T_i to have variation in $T_i|X_i$
 - But that same unobserved factor could also affect outcome: no way to know if $\text{cov}(\eta_i, \varepsilon_i) = 0$
 - Cannot be sure that estimate of treatment effect β is consistent

Developing Empirical Research Designs

- A “research design” is a source of variation in η_i that is credibly unrelated to ε_i
 - Ex: a reform that affects people above age 65 but not below
 - People at age 64 and 65 likely to have similar outcomes $\rightarrow$
 $cov(\eta_i, \varepsilon_i) = 0$
- General lesson: “controlling” for confounding factors using regression or reweighting will rarely give you convincing estimates
- However, reweighting can be a useful technique to obtain better control groups when paired with a quasi-experimental research design
 - More on this below (Dinardo, Fortin, Lemieux 1996)

- Question: How do cigarette tax increases affect prices?
 - Do they take money from cigarette companies or smokers?
- Partial equilibrium is a plausible approximation for cigarettes, so use that framework here

Cigarette Taxation: Background

- Cigarettes taxed at both federal and state levels in U.S.
- Total revenue of about \$35 billion per year, similar to estate taxation
- Federal tax increased from \$0.39 to \$1.01 per pack in 2009
- Variation among states: from 30 cents per pack in VA to \$4.35 in NY in 2012
- Controversial commodity due to health and paternalism concerns

- Since 1975, more than 200 state tax changes → natural experiments to investigate tax incidence
- Exploit these state-level changes in excise tax rates using simple diff-in-diff research designs
- Idea: Suppose federal govt. implements a tax change. Compare cigarette prices before and after the change

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

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

Difference-in-Difference

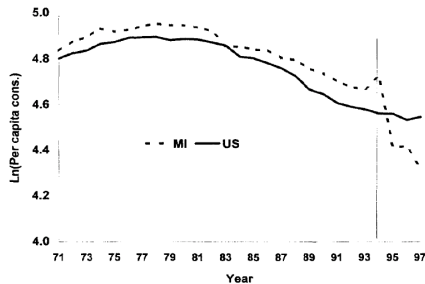
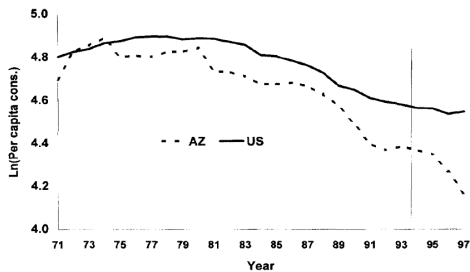
- But what if price fluctuates because of climatic conditions or trends in demand?

→ First difference (and time series) estimate biased

- Can relax ID assumption 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 for DD: “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 Trends Assumption

- Can use placebo DD to test parallel trends assumption
 - Analogous to permutation test: pretend reform occurred at other points and replicate estimate
- If DD in other periods is not zero, then $DD_{t=1}$ likely biased
 - Useful to plot long time series of outcomes for treatment and control
 - Pattern should be parallel lines, with sharp change just after reform
 - Rest of U.S. a good control for MI in example above but not AZ

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 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}
 - Ex: states with higher taxes may have more anti-tobacco campaigns

- Include time and state dummies to solve 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 $\hat{T}_{jt}$ 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 estimates by pooling several changes
- Disadvantage: fixed effects is a black-box regression, more difficult to check trends non-parametrically as with a single change
 - Combine with graphical, non-parametric evidence around certain policy changes
 - Also useful to scatter residuals $\hat{P}_{jt}$ vs. $\hat{T}_{jt}$
- Same parallel trends identification assumption as *DD*
 - Potential violation: policy reforms may respond to trends in outcomes
 - Ex: tobacco prices falling $\rightarrow$ state decides to raise tax rate

- 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		Net retail price in Tennessee, 1970–1994	
	Nominal (1)	Real (2)	Nominal (3)	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

- 100% pass through implies supply elasticity of $\varepsilon_S = \infty$ at state level
 - Theory suggests that pass through would be lower 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: $\varepsilon_D = -0.42$
→ 10% increase in price induces a 4.2% reduction in consumption
- How to compute price elasticity of demand when using variation arising from tax changes?
- Tax passed 1-1 onto consumers, so we can substitute $\Delta P = \Delta T$ here
- Then compute ε_D from $\hat{\beta} = (\Delta Q / Q) / \Delta T$ from regression coefficient of log demand on cigarette tax:

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

with P (price) and Q (quantity) are sample means

IV Estimation of Price Elasticities

- How to estimate price elasticity of demand when tax and prices do not move together 1-1?
 - 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:

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

- 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 coeff. is ratio of reduced-form to first-stage coeff.:

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

- 2SLS rescales reduced-form to account for $\Delta P / \Delta T \neq 1$

- *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 ($T_{j,t-1}, T_{j,t-2}, \dots$).
- Are identification assumptions still valid here? Tradeoff between LR and validity of identification assumptions

- 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 levied on cigarette companies lead to poor paying more for same goods, with no impact on companies!

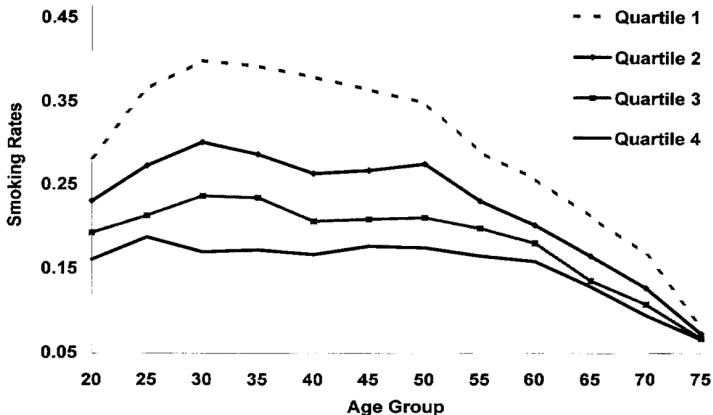


FIGURE 8. *Smoking Rates by Income Quartiles and Age—1992–1993 CPS TUS*

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

- 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** (0.007)	-0.2337** (0.010)	-0.1439** (0.015)	-0.1944** (0.014)
Benefit Household*Week3	-0.2907** (0.008)	-0.3167** (0.010)	-0.2027** (0.017)	-0.2853** (0.015)
Benefit Household*Week4	-0.3352** (0.008)	-0.3768** (0.011)	-0.2293** (0.017)	-0.3208** (0.015)
Week2	-0.0186** (0.003)	-0.0104** (0.004)	-0.0310** (0.005)	-0.0223** (0.007)
Week3	-0.0062* (0.003)	-0.0071 (0.004)	-0.0025 (0.005)	-0.0132 (0.007)
Week4	-0.0061* (0.003)	-0.0147** (0.004)	-0.0019 (0.005)	0.0087 (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)

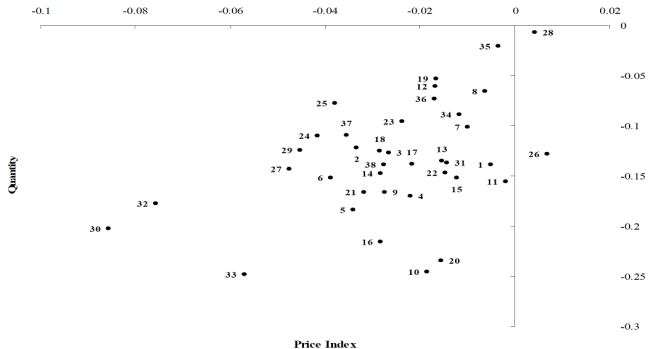
Table 8: Change in Log(Price Index) and Log (Price Index of The First Week of The Month) across different Stores

	NV High Poverty				NV Low Poverty			
	All	Store 1	Store 2	Store3	All	Store 1	Store 2	Store3
<i>Log.(Price Index)</i>								
Week2	-0.0195** (0.002)	-0.0172** (0.004)	-0.0212** (0.003)	-0.0204** (0.004)	-0.0012 (0.003)	-0.0011 (0.005)	-0.0038 (0.005)	0.0014 (0.005)
Week3	-0.0243** (0.002)	-0.0138** (0.004)	-0.0268** (0.003)	-0.0334** (0.004)	0.0015 (0.003)	0.0021 (0.005)	-0.0005 (0.005)	0.0027 (0.005)
Week4	-0.0266** (0.002)	-0.0165** (0.005)	-0.0293** (0.003)	-0.0349** (0.004)	-0.0046 (0.003)	-0.0075 (0.005)	-0.0036 (0.005)	-0.0027 (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 Cream; 25. Frozen Dessert; 26. 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.

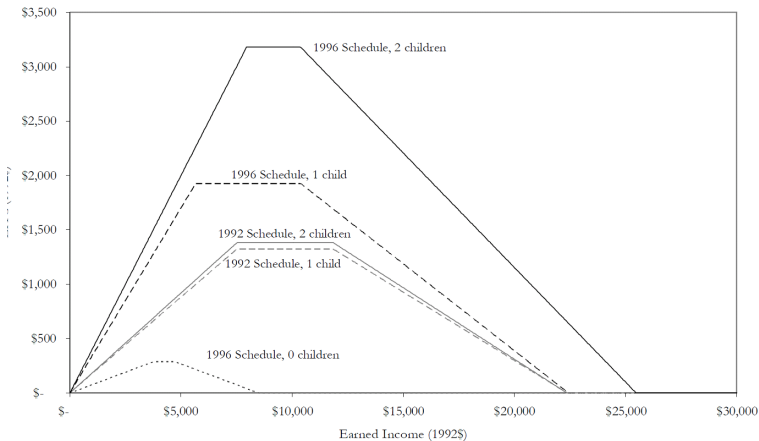
Hastings and Washington: Results

- 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
- Interesting theoretical implication: subsidies 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 a subsidy such as the EITC

- 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 $\rightarrow$ 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?

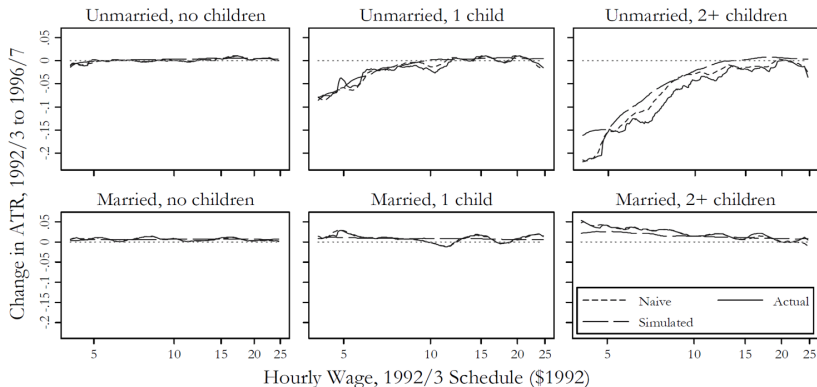
- Rothstein considers a model of the labor market with three types of agents
 - ① Employers
 - ② EITC-eligible workers
 - ③ EITC-ineligible workers
- Extends standard partial eq incidence model to allow for differentiated labor supply and different tax rates across demographic groups
- Heterogeneity 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.

- Two main challenges to identification:
 - ① EITC 1992-1994 expansion when nation coming out of recession
→ 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.
→ 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

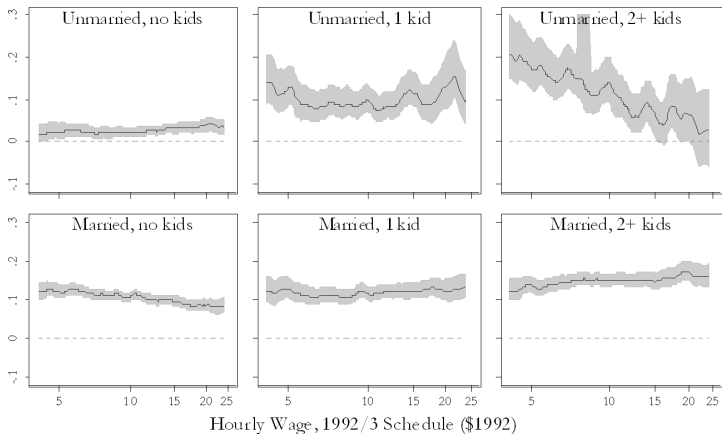
- Ideal test: measure how wage of a given individual changes when EITC is introduced relative to a similar but ineligible individual
- Problem: data is CPS repeated cross-sections. Cannot track “same individual.”
- Moreover, wage rigidities may prevent cuts for existing employees.
- Solution: reweighting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

DFL Reweighting

- Generalizes propensity score reweighting
- Used to examine changes in distributions 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.

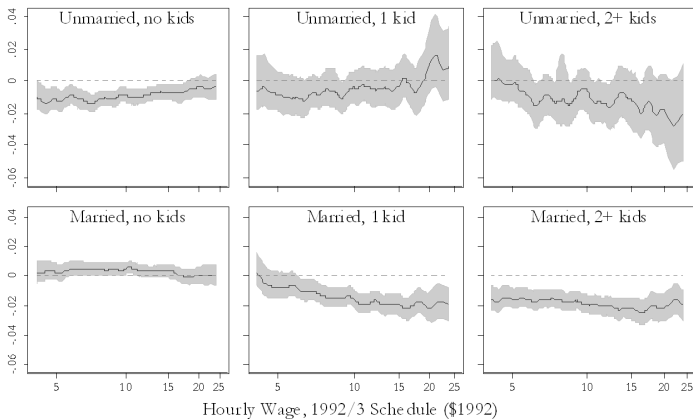
- Example: 100 short, 100 tall pre-reform and 200 short, 100 tall post-reform
- Then put $\frac{2}{3}$ weight on tall and $\frac{1}{3}$ on short when calculating wage distribution after reform
- Compare reweighted post-reform distribution to pre-reform distribution to assess effect of expansion on wages
- 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



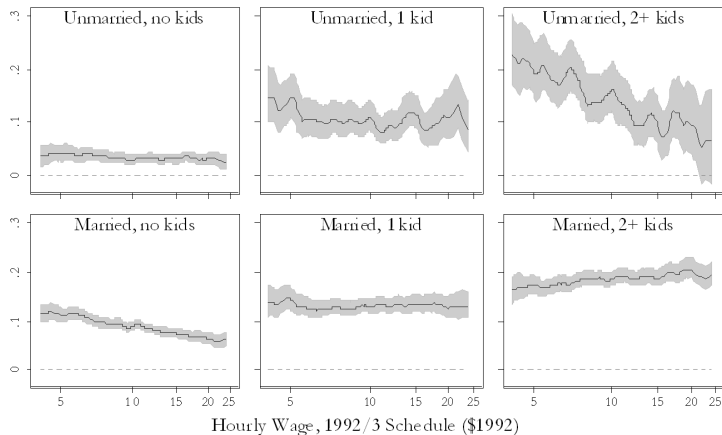
Source: Rothstein 2008.

Intensive margin: Change in weekly hours conditional on working



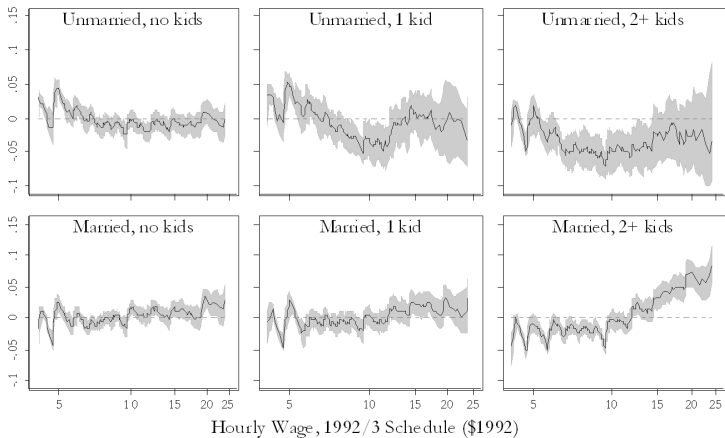
Source: Rothstein 2008.

Extensive margin: Change in employment-population ratio



Source: Rothstein 2008.

Wage changes



Source: Rothstein 2008.

- 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 (0.25)	-0.74 (0.25)	-1.06 (0.23)	-1.06 (0.23)
Change in avg. ATR (across groups)	σ^2	-0.24 (0.32)	-1.02 (1.26)	0.00 (1.02)	-0.26 (2.02)
Change in population size (avg. across groups)	$-\sigma$	-0.10 (0.21)	-0.16 (0.37)	-0.20 (0.31)	-0.12 (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)	σ	-1.26 (0.39)	-5.14 (1.75)	2.45 (1.33)	4.78 (2.28)
Change in population size (avg. across groups)	-1	-0.38 (0.29)	-0.52 (0.63)	-0.22 (0.26)	-1.14 (0.70)

Source: Rothstein 2008.

- 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

- 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 Short run vs. long run effects; important due to evidence of nominal wage rigidities.
- ④ 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.

General Equilibrium Analysis

- Now move beyond two-good partial equilibrium model to analyze impacts on all prices
- Typical goal: 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 GE 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**

- Characterize impacts over time or across generations
- 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 $r = (1 - d\tau)p_2F_{2K}$
- Linearize the 10 eq'ns around initial equilibrium to compute the effect of $d\tau$ on all 10 variables ($dw, dr, dL_1, \dots$)
- 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

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 $\rightarrow$ 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 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 about \$80 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

- Findings very sensitive to structural assumptions
 - Ex: assumption of perfect competition
 - Key behavioral elasticities and functional form assumptions
- Modern econometric methods conceptually not suitable for GE problems
 - The whole point is “spillover effects” (contamination)
- Need a new empirical paradigm to deal with these problems

- 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

- 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:

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

- Find $\beta = 0.89$ (0.07)

- 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

- 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

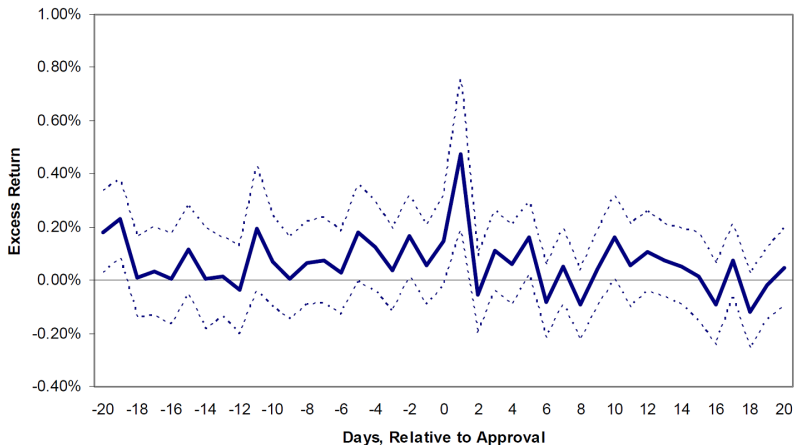
Capitalization: Empirical Applications

- Classic application: Cutler (1988) on effect of Tax Reform Act of 1986 on corporations
- Focus on two more recent studies here
 - ① [Friedman 2008] Effect of Medicare Part D on drug companies
 - ② [Linden and Rockoff 2008] Effect of a sex offender moving into neighborhood on home values

- 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?
- Basic research design: event study
- Plot excess (market adjusted) returns for drug companies around FDA approval of drugs

- Event studies are a powerful research design when treatments are staggered in time across individuals
- Use a group of treated individuals as counterfactuals for each other to account for time series trends
- Good for identifying sharp, short-run effects but not longer-term impacts
- Methodology
 - 1 Define "event time" as calendar time minus date of treatment for each treated obs.
 - 2 Plot means/medians, etc. of outcome variable by event time

Excess Returns Around Drug Approval Date

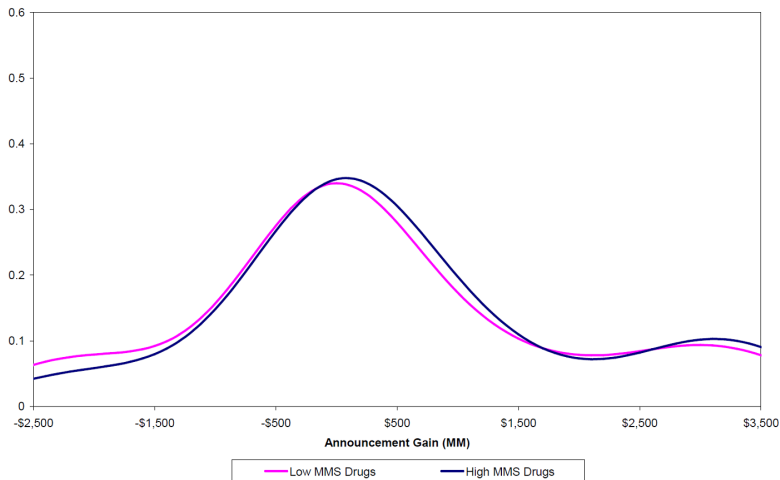


Source: Friedman 2008

- Test 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:

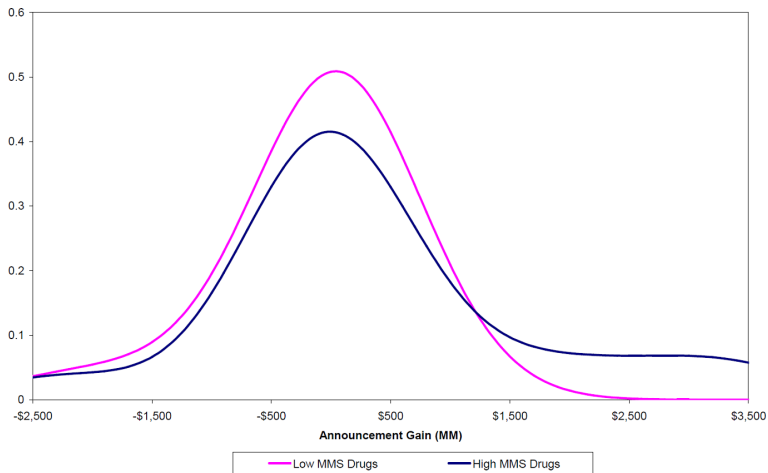
$$Excess_i = \alpha + \beta MMS_i + \gamma Post2003_t + \lambda Post2003_t \cdot MMS_i$$

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



Source: Friedman 2008

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



Source: Friedman 2008

- 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

- 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
- Research design: examine 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



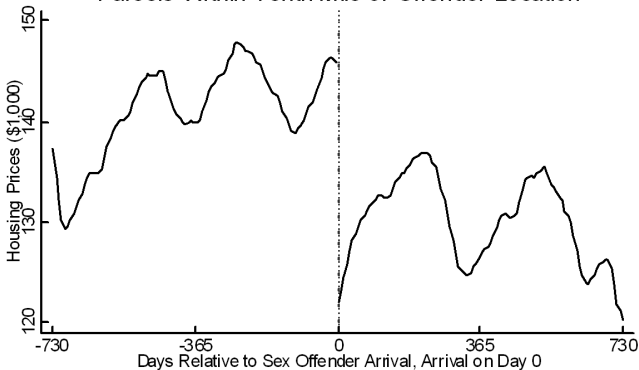
Source: Linden and Rockoff 2008.

Illustration of Identification Strategy



Source: Linden and Rockoff 2008.

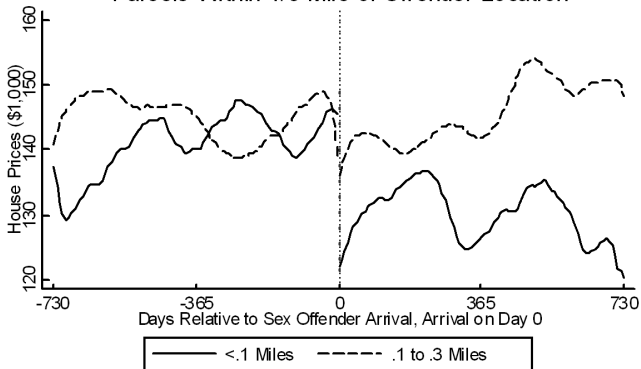
Figure 3a: Price Trends Before and After Offenders' Arrivals
Parcels Within Tenth Mile of Offender Location



Note: Results from local polynomial regressions (bandwidth=90 days) of sale price on days before/after offender arrival.

Source: Linden and Rockoff 2008.

Figure 3b: Price Trends Before and After Offenders' Arrivals
Parcels Within 1/3 Mile of Offender Location



Note: Results from local polynomial regressions (bandwidth=90 days) of sale price on days before/after offender arrival.

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?

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

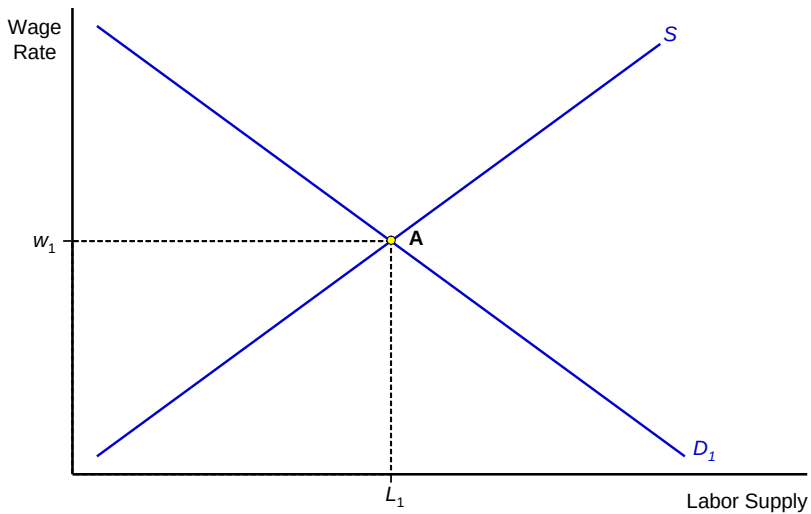
- Labor demand (D) and labor supply (S) are functions of the wage, w
- Initial equilibrium:

$$D(w_0) = S(w_0)$$

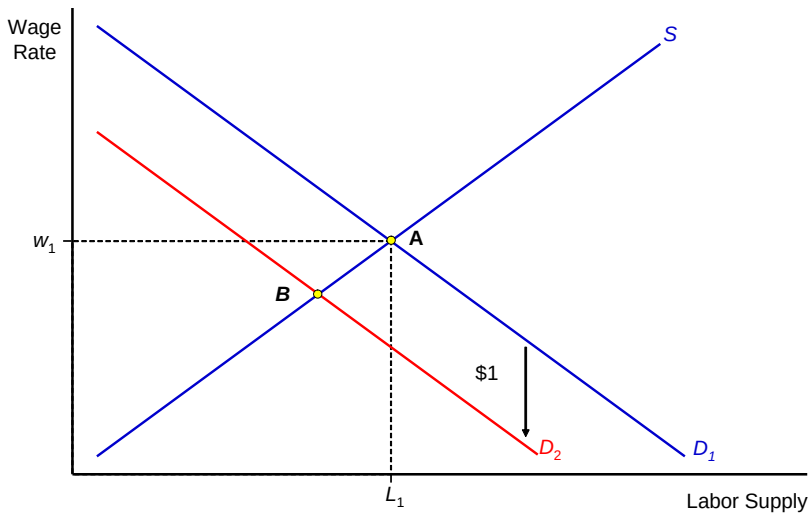
- Now, govt mandates employers provide a benefit with cost t
- Workers value the benefit at αt dollars
- Typically $0 < \alpha < 1$ but $\alpha > 1$ possible with market failures
- Labor cost is now $w + t$, effective wage $w + \alpha t$
- New equilibrium:

$$D(w + t) = S(w + \alpha t)$$

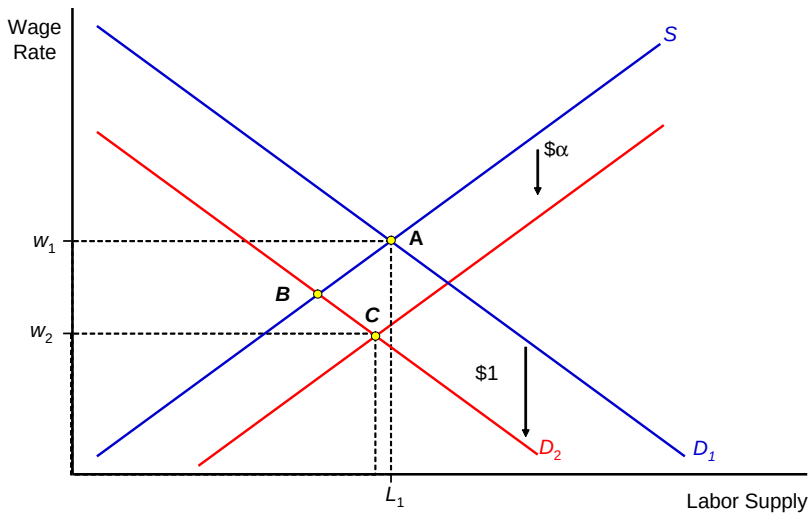
Mandated Benefit



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

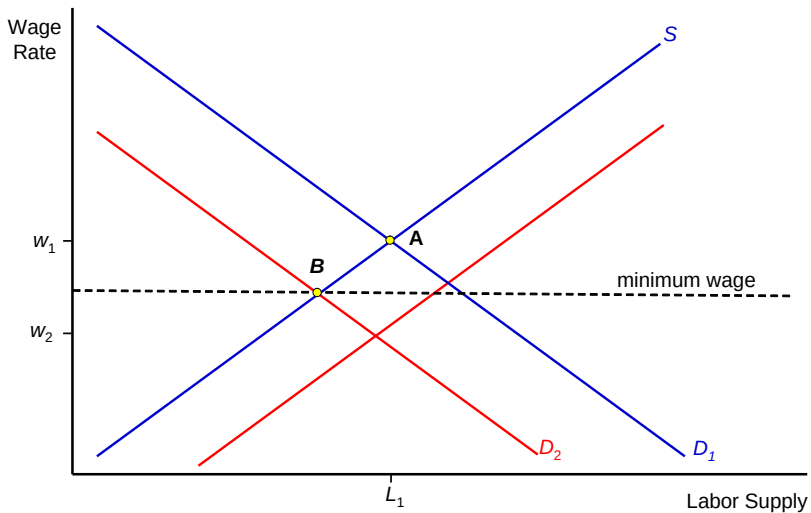
Mandated Benefits: Empirical Applications

- Focus here on Acemoglu and Angrist (2001) study of Americans with Disabilities Act
 - Other applications
- ① Gruber and Krueger (1991) on workers compensation mandates
 - ② Gruber (1994) on mandated maternity benefits
 - ③ Kolstad and Kowalski (2012) on health insurance mandates

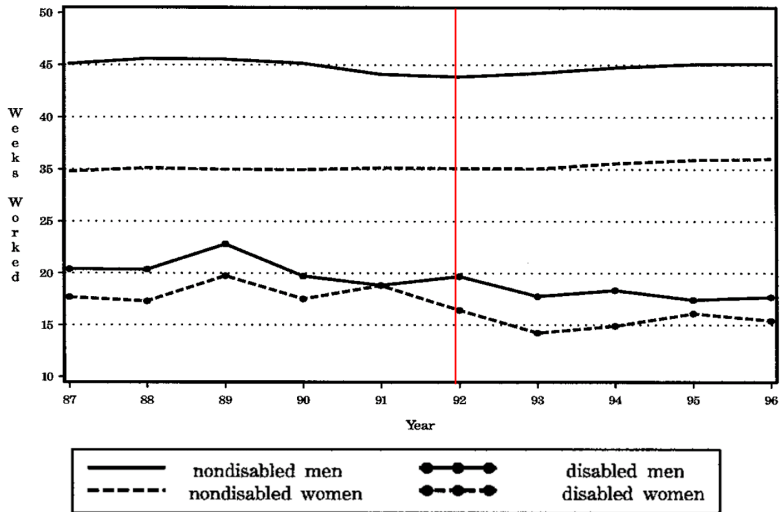
Acemoglu and Angrist (2001)

- 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 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 (.71)		-1.08 (1.07)		-.052 (.044)		-.055 (.052)	
Disability × 1989	2.00 (.71)		.67 (1.06)		-.062 (.043)		.050 (.050)	
Disability × 1990	-.79 (.71)		-1.33 (1.05)		-.035 (.044)		-.111 (.050)	
Disability × 1991	-.53 (.70)	-.70 (.78)	.08 (1.04)	1.09 (1.16)	-.020 (.043)	.046 (.049)	.105 (.050)	.191 (.055)
Disability × 1992	.57 (.69)	.40 (.95)	-2.81 (1.02)	-1.56 (1.40)	-.00 (.043)	.078 (.059)	-.010 (.050)	.100 (.067)
Disability × 1993	-1.44 (.69)	-1.61 (1.14)	-4.37 (1.02)	-2.89 (1.69)	-.04 (.044)	.050 (.072)	-.058 (.050)	.074 (.082)
Disability × 1994	-1.63 (.70)	-1.80 (1.35)	-5.00 (1.01)	-3.29 (1.98)	-.143 (.044)	-.042 (.084)	-.043 (.050)	.112 (.095)
Disability × 1995	-2.93 (.72)	-3.09 (1.56)	-3.93 (1.01)	-1.99 (2.29)	-.098 (.046)	.014 (.098)	.005 (.049)	.184 (.110)
Disability × 1996	-2.68 (.70)	-2.84 (1.77)	-4.41 (1.03)	-2.23 (2.62)	-.158 (.044)	-.034 (.110)	-.111 (.050)	.090 (.125)
Disability × linear trend		-.004 (.22)		-.23 (.33)		-.012 (.014)		-.023 (.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 2012

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

- Incidence: 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

- Auerbach (1985) handbook chapter
- Chetty, Looney, Kroft (AER 2009)
- Chetty (Annual Review 2009)
- Mas-Colell, Whinston, Green Chapter 3 for background on price theory concepts

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 falls by more than \$1 because the tax distorts behavior
- How to implement policies that minimize these efficiency costs?
 - Start with positive analysis of how to measure efficiency cost of a given tax system

Marshallian Surplus: Assumptions

- Simplest analysis of efficiency costs: Marshallian surplus
- Two assumptions:
 - 1 Quasilinear utility: no income effects, money metric
 - 2 Competitive production

Partial Equilibrium Model: Setup

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

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

- 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

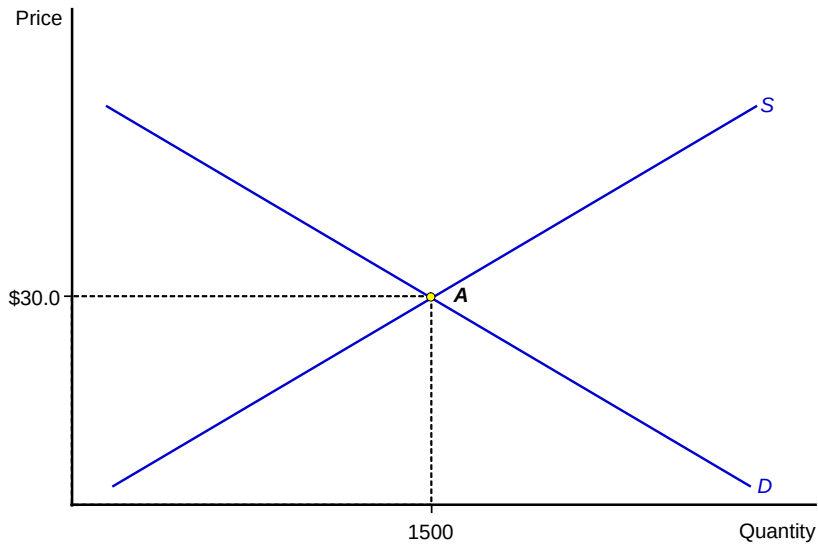
$$p = c'(S(p))$$

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

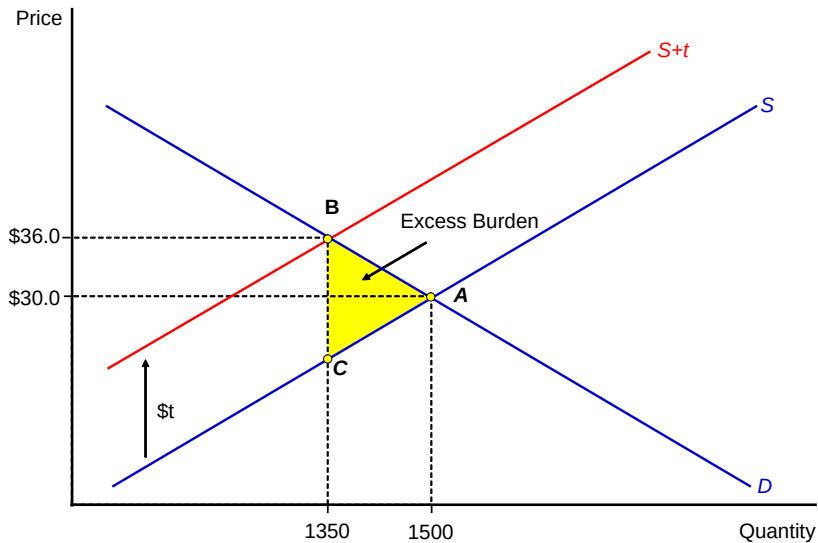
$$Q(\tau) = D(p + \tau) = S(p)$$

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

Excess Burden of Taxation



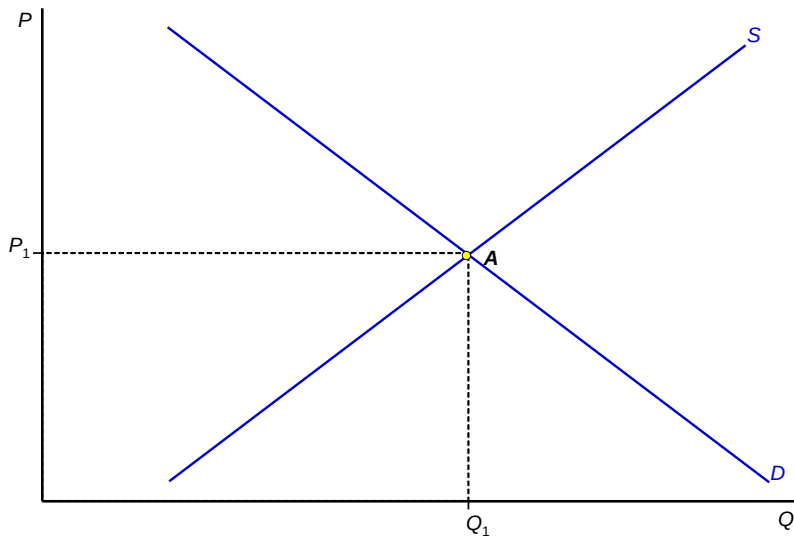
Excess Burden of Taxation



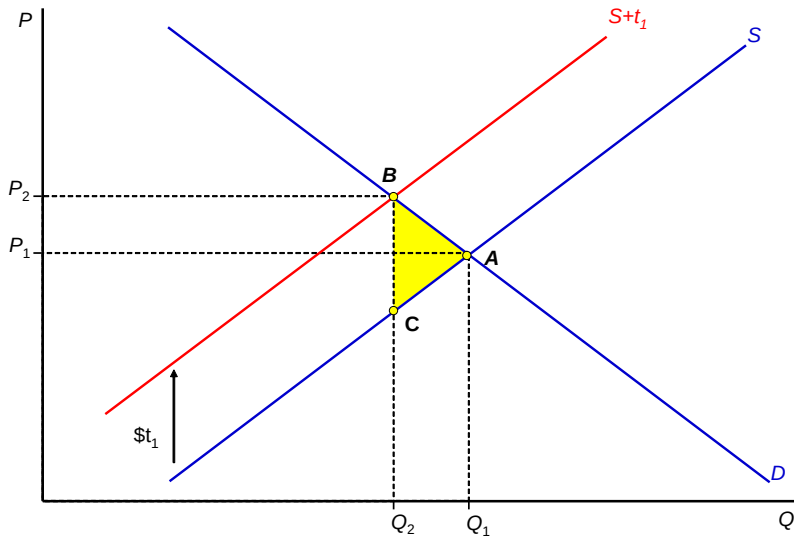
Efficiency Cost: Qualitative Properties

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

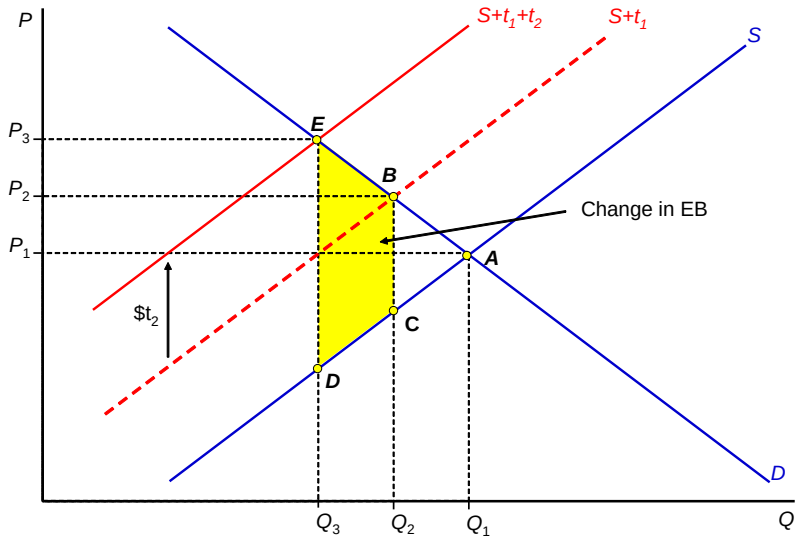
EB Increases with Square of Tax Rate



EB Increases with Square of Tax Rate

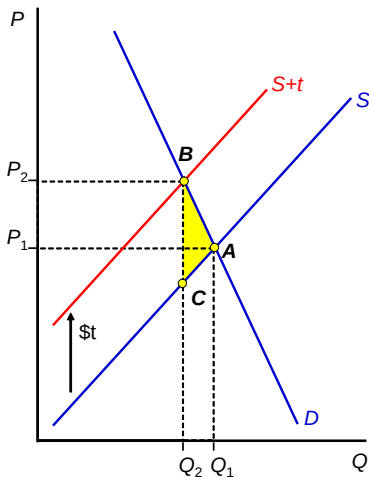


EB Increases with Square of Tax Rate

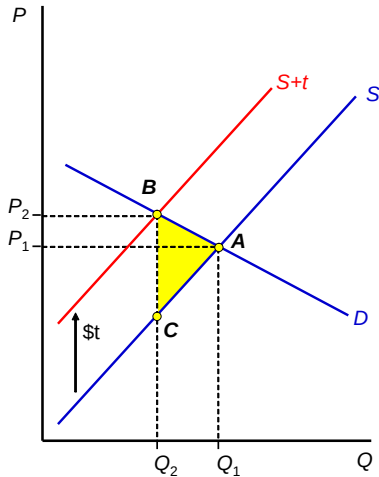


Comparative Statics

(a) Inelastic Demand



(b) Elastic Demand



- With many goods, the most efficient way to raise tax revenue is:
 - ① Tax inelastic goods more (e.g. medical drugs, food)
 - ② Spread taxes across all goods to keep tax rates relatively low on all goods (broad tax base)
- These are two countervailing forces; balancing them requires quantitative measurement of excess burden

Measuring Excess Burden: Marshallian Surplus

- How to measure excess burden? Three empirically implementable methods:
 - 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

Method 1: Supply and Demand Elasticities

$$EB = -\frac{1}{2}dQd\tau$$

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

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

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

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

Method 2: Distortions in Equilibrium Quantity

- 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 reduced-form regression:

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

- Identify β using exogenous variation in τ . Then:

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

Marginal Excess Burden of Tax Increase

- Excess burden of a tax τ is

$$EB(\tau) = -(1/2) \frac{dQ}{d\tau} \tau^2$$

- Consider EB from raising tax by $\Delta\tau$ given pre-existing tax τ :

$$\begin{aligned} EB(\Delta\tau) &= -(1/2) \frac{dQ}{d\tau} [(\tau + \Delta\tau)^2 - \tau^2] \\ &= -(1/2) \frac{dQ}{d\tau} \cdot [2\tau \cdot \Delta\tau + (\Delta\tau)^2] \\ &= -\tau \frac{dQ}{d\tau} \Delta\tau - (1/2) \frac{dQ}{d\tau} (\Delta\tau)^2 \end{aligned}$$

- First term is first-order in $\Delta\tau$; second term is second-order $((\Delta\tau)^2)$
- This is why taxing markets with pre-existing taxes generates larger marginal EB
 - EB of $\Delta\tau = 1\%$ is 10 times larger if $\tau = 10\%$ than if $\tau = 0$.

First vs. Second-Order Approximations

- Computing marginal excess burden by differentiating formula for excess burden gives:

$$\frac{dEB}{d\tau} \cdot \Delta\tau = -\tau \frac{dQ}{d\tau} \cdot \Delta\tau$$

- First derivative of $EB(\tau)$ only includes first-order term in Taylor expansion:

$$EB(\tau + \Delta\tau) = EB(\tau) + \frac{dEB}{d\tau} \Delta\tau + \frac{1}{2} \frac{d^2 EB}{d\tau^2} (\Delta\tau)^2$$

- First-order approximation is accurate when τ large relative to $\Delta\tau$
 - Ex: $\tau = 20\%$, $\Delta\tau = 5\%$ implies first term accounts for 90% of EB
 - But introduction of new tax ($\tau = 0$) generates EB only through second-order term

Method 3: Leakage in government revenue

- To first order, marginal excess burden of raising τ is:

$$\frac{\partial EB}{\partial \tau} = -\tau \frac{dQ}{d\tau}$$

- Observe that tax revenue $R(\tau) = Q\tau$

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

- Actual revenue gain: $\frac{\partial R}{\partial \tau} = Q + \tau \frac{dQ}{d\tau}$

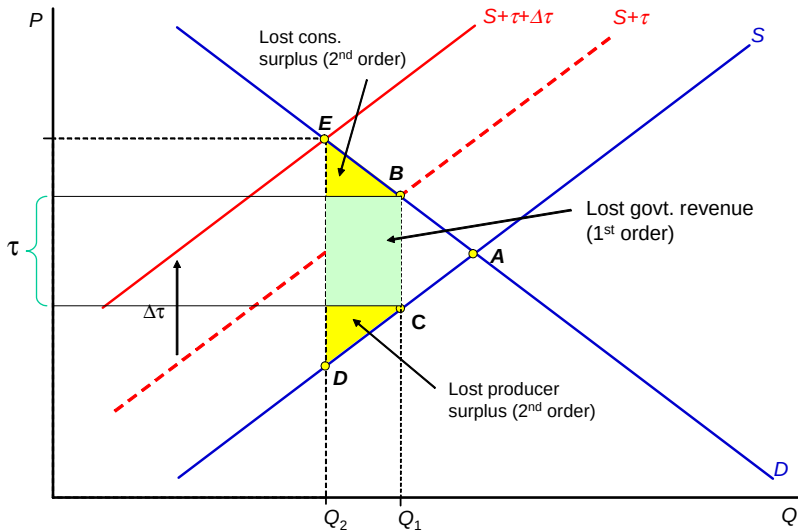
- MEB is the difference between mechanical and actual revenue gain:

$$\frac{\partial R}{\partial \tau}|_Q - \frac{dR}{d\tau} = Q - [Q + \tau \frac{dQ}{d\tau}] = -\tau \frac{dQ}{d\tau} = \frac{\partial EB}{\partial \tau}$$

First vs. Second-Order Approximations

- Why does leakage in govt. revenue only capture first-order term?
 - Govt revenue loss: rectangle in Harberger trapezoid, proportional to $\Delta\tau$
 - Consumer and producer surplus loss: triangles in trapezoid (proportional to $\Delta\tau^2$)
- Method 3 is accurate for measuring marginal excess burden given pre-existing taxes but not introduction of new taxes

Excess Burden of a Tax Increase: Harberger Trapezoid



General Model with Income Effects

- Drop quasilinearity assumption and consider an individual with utility

$$u(c_1, \dots, c_N) = u(c)$$

- Individual's problem:

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

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

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

Demand Functions and Indirect Utility

- Let λ denote multiplier on budget constraint
- First order condition in c_i :

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

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

Measuring Deadweight Loss with Income Effects

- Question: how much utility is lost because of tax beyond revenue transferred to government?
- Marshallian surplus does not answer this question with income effects
 - Problem: not derived from utility function or a welfare measure
 - Creates various problems such as “path dependence” with taxes on multiple goods

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

- 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
- Utility at initial price q^0 :

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

- 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
- Literature typically 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 1985, Section 3.2)

Efficiency Cost Formulas with Income Effects

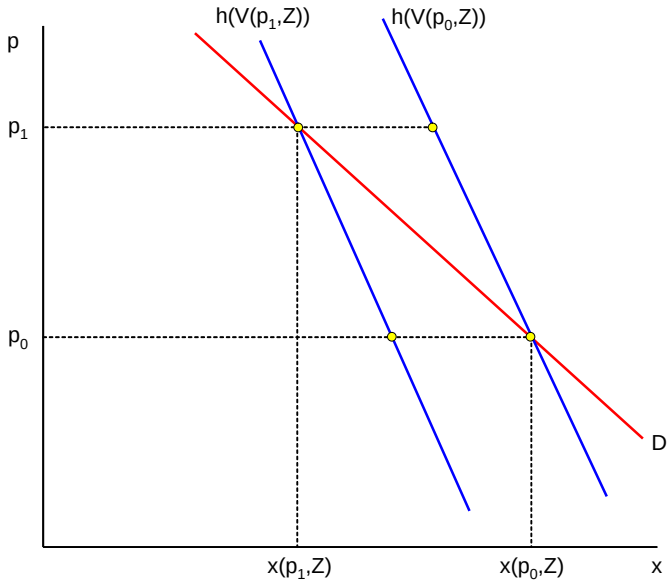
- Derive empirically implementable formulas using Hicksian demand (EV and CV)
- Assume p is fixed $\rightarrow$ flat supply, constant returns to scale
- The envelope thm implies that $e_{q_i}(q, u) = h_i$, and so:

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

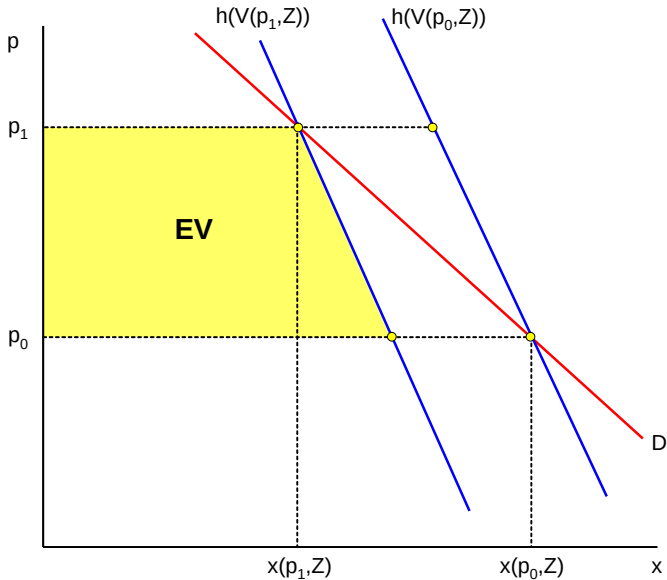
- If only one price is changing, this is the area under the Hicksian demand curve for that good
- Note that optimization implies that

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

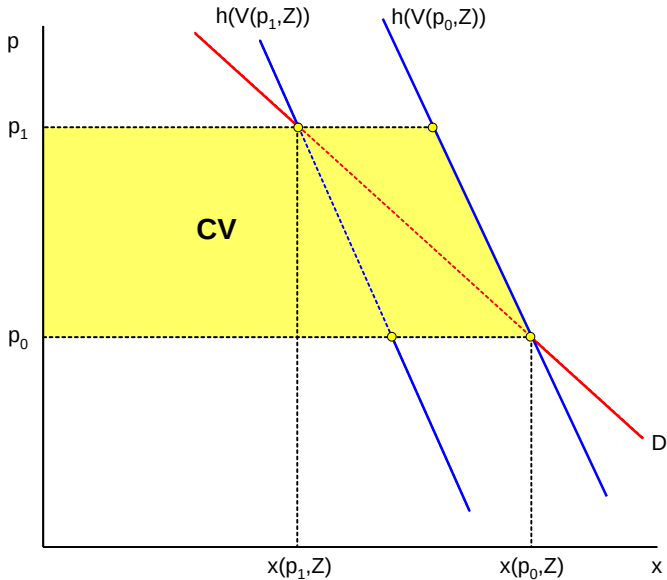
Compensating vs. Equivalent Variation



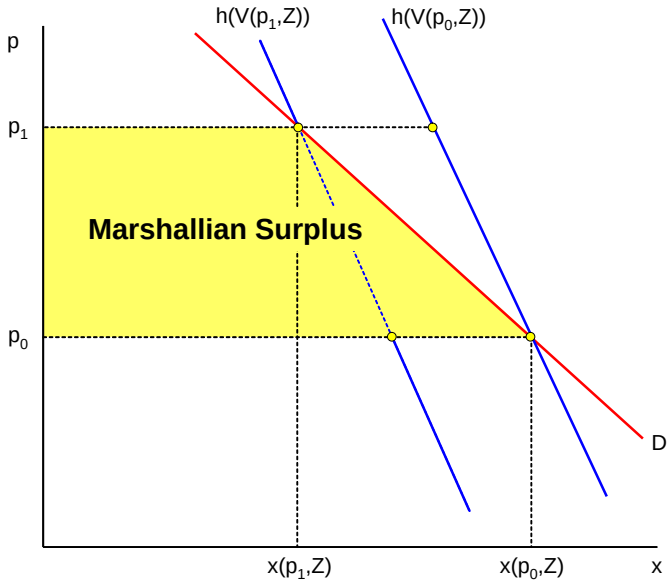
Compensating vs. Equivalent Variation



Compensating vs. Equivalent Variation



Marshallian Surplus



- With one price change:

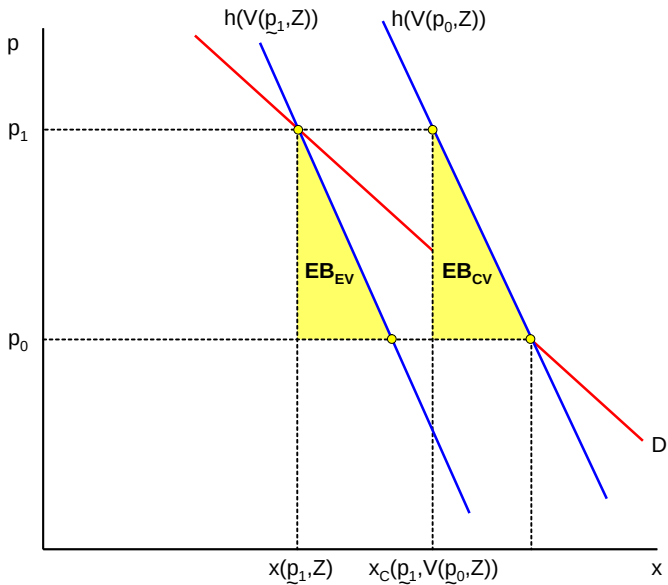
$$EV < \text{Marshallian Surplus} < CV$$

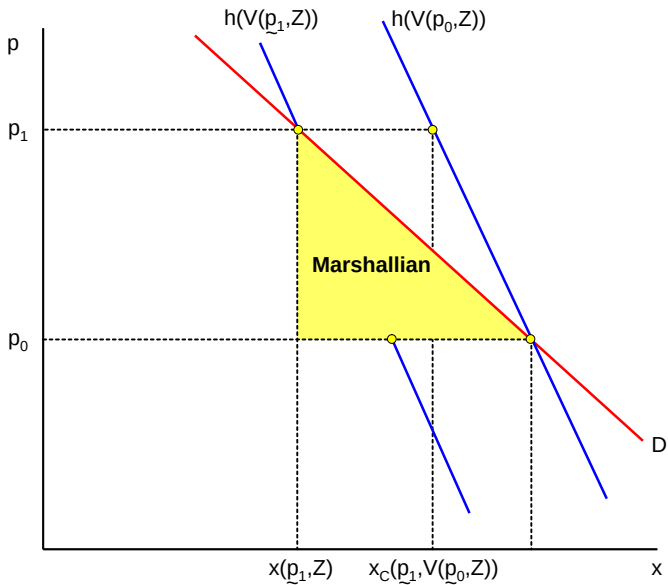
- But this is not true in general with multiple price changes because Marshallian Surplus is ill-defined

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

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

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





- 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
- $CV = EV = \text{Marshallian DWL}$ only with quasilinear utility (Chipman and Moore 1980)

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

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

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

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

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

Harberger Formula

- Without pre-existing tax, obtain “standard” Harberger formula:

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

- General lesson: use compensated (substitution) elasticities to compute EB , not uncompensated elasticities
- To implement empirically, estimate Marshallian price elasticity and income elasticity. Then apply Slutsky eqn:

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

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
- 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)

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

- Basic formula hard to implement because it requires estimates of all cross-price elasticities
- Goulder and Williams make formula empirically implementable by making 3 assumptions:
 - 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 $\frac{\partial I}{\partial \tau_k}$ 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? Consumption tax and labor income tax have equivalent effects
 - Price increase in all consumption goods has the same effect on labor supply as an increase in tax on labor:

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

Goulder and Williams Formula: Intuition

- Why do we only need to estimate η_L ?
- Step 1: consumption tax and labor income tax have equivalent effects
 - Price increase in all consumption goods has the same effect on labor supply as an increase in tax on labor:

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

Goulder and Williams Formula: Intuition

- Step 2: 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 of l w.r.t. τ_k is equivalent to $\eta_L s_k$
- Labor supply elasticity η_L sufficient to calculate cross-elasticity for good that has “average” level of substitutability

- 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 approximate
- Alternative approach: structural estimation of demand model
- Start by estimating Marshallian demand functions:

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

- Then integrate to recover underlying indirect utility function $v(q, Z)$
- Invert to obtain expenditure function $e(q, u)$ and 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
- Modern literature focuses on deriving “sufficient statistic” formulas that can be implemented using quasi-experimental techniques
- Now develop general 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)
 - But permit multiple goods (GE)

Sufficient Statistics vs Structural Methods

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

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

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

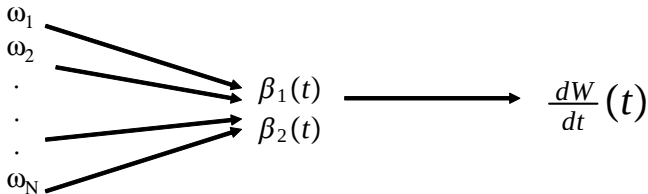
$$W(t) = \left\{ \max_x u(x_1, \dots, x_{N-1}) + Z - (p_1 + t)x_1 - \sum_{i=2}^{N-1} p_i x_i \right\} + tx_1$$

- 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)

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

- Envelope conditions for $(x_1, \dots, x_N)$ 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 $u_k(x_1, \dots, x_{N-1}) + x_N$
- Social welfare function under utilitarian criterion:

$$W(t) = \left\{ \max_x \sum_{k=1}^K [u_k(x_1^k, \dots, x_{N-1}^k) + Z - (p_1 + t)x_1^k - \sum_{i=2}^{N-1} p_i x_i^k] \right\} + \sum_{k=1}^K t x_1^k$$

- 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

- Harberger sufficient statistic also works with discrete choice
- Agents have value V_k for good 1; can either buy or not buy
- Let $F(V)$ denote distribution of valuations
- With 2 goods, utility of agent k is

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

- Social welfare:

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

- 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:

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

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

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

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

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

- 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: $l_1, \dots, l_N$
- 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:

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

- 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 :

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

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

Chetty (2009) Formula

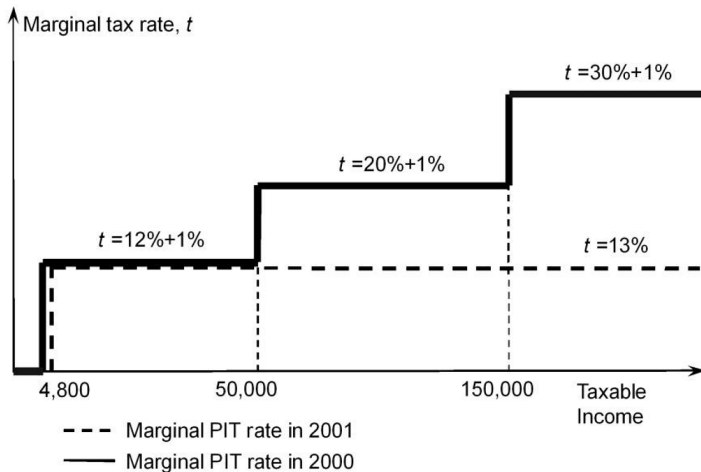
- With both transfer cost $z(e)$ and resource cost $g(e)$ of evasion:

$$\begin{aligned}\frac{dW}{dt} &= t \frac{dLI}{dt} - g'(e) \frac{de}{dt} \\ &= t \left\{ \mu \frac{dTl}{dt} + (1 - \mu) \frac{dLI}{dt} \right\} \\ &= -\frac{t}{1 - t} \left\{ \mu Tl \varepsilon_{Tl} + (1 - \mu) w l \varepsilon_{LI} \right\}\end{aligned}$$

- EB depends on weighted average of taxable income (ε_{Tl}) 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, $g'(e)$

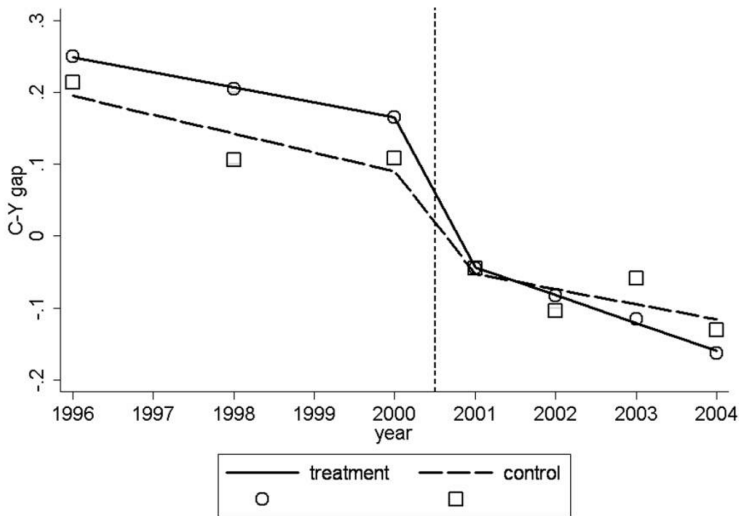
- Estimate ε_{LI} and ε_{TI} to implement formula that permits transfer costs
- Insight: consumption data can be used to infer ε_{LI}
- Estimate effect of 2001 flat tax reform in Russia on gap between taxable income and consumption, which they interpret as evasion

Marginal personal income tax rate before and after the reform



Source: Gorodnichenko, Martinez-Vazquez, and Peter 2009

Consumption-income gap dynamics



Source: Gorodnichenko, Martinez-Vazquez, and Peter 2009

- Taxable income elasticity $\frac{dTl}{dt}$ is large, whereas labor income elasticity $\frac{dLl}{dt}$ is not
→ 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?

- 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

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

- 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

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

Rosen (1982): $\varepsilon_{H,r} = -1$

Income elasticity: 0.75

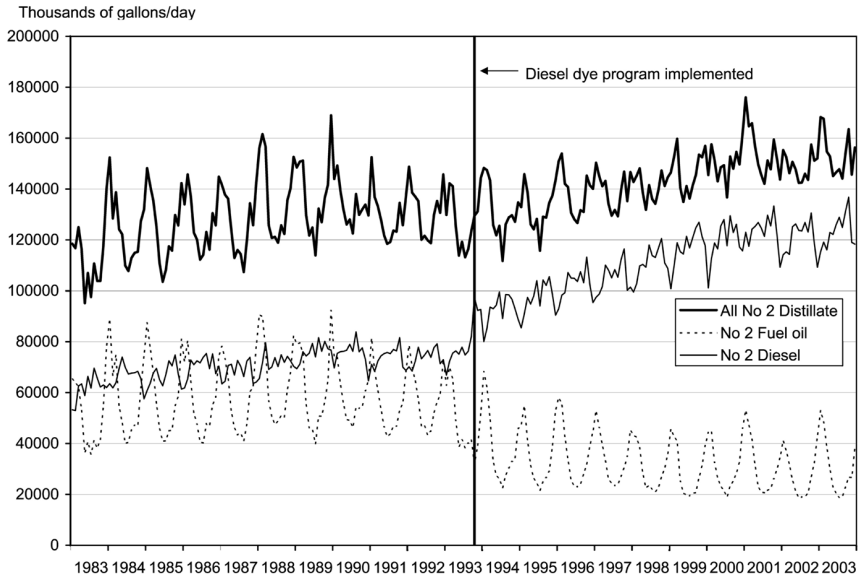
Housing share: 0.25

$$\Rightarrow \text{Compensated elasticity: } -1 + \frac{3}{4} \times \frac{1}{4} \simeq -0.8$$

- 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 $\rightarrow$ tax is effectively salient
- Calculates amount of overconsumption of housing and efficiency cost of housing subsidy

- 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

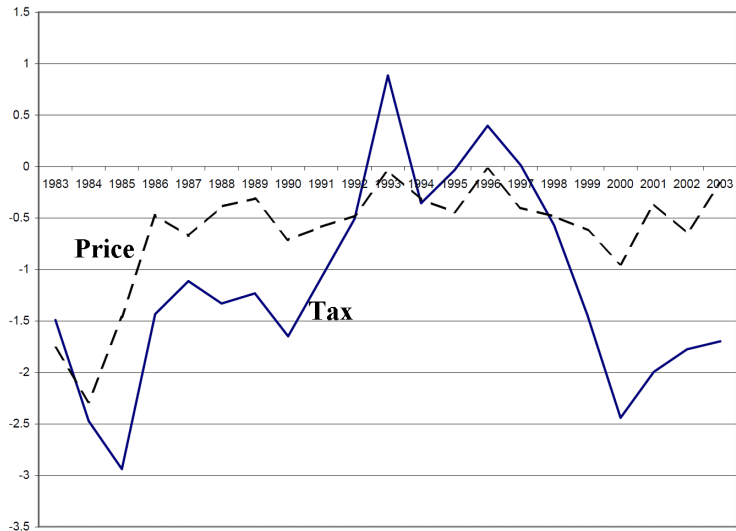
- Study deadweight cost from taxing diesel fuels, focusing on evasion
- Diesel fuel used for business purposes (e.g. trucking) is taxed, but residential purposes (e.g. heating homes) is not
- Substantial opportunity to evade tax
- 1993: government added red dye to residential diesel fuel
 - Easy to monitor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



Source: Marion and Muehlleger 2008

- Use reform to assess deadweight costs of evasion and taxation
 - Harder to evade $\rightarrow$ 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 Muehligler 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
- How to calculate efficiency costs when 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

- 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 model
 - 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

- Derive 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

- 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

- 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: $CV = 0$.

- 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

- 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
- Identifying “refinements” typically requires some insight into positive theory of behavior

- 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

- 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 the chosen consumption bundle.
Agent's indirect utility given tax of 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:

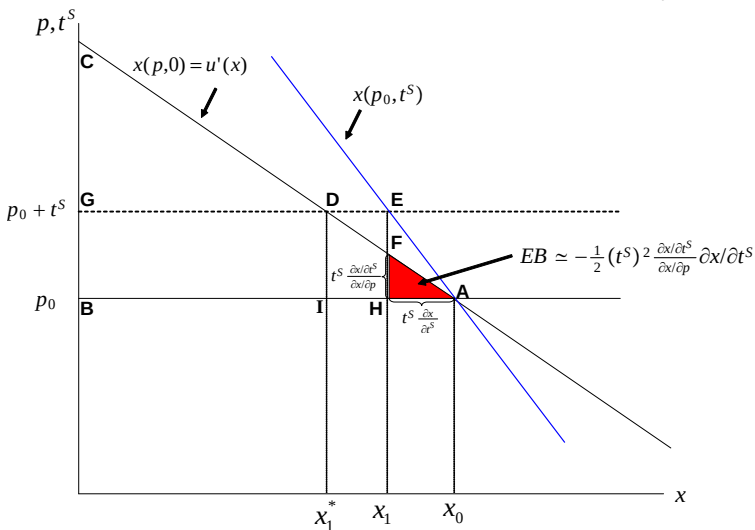
$$x(p, 0, Z) = x^*(p, 0, Z) = \arg \max_x u(x) + v(Z - px)$$

- A1 specifies ancillary condition: tax rate and salience does not enter utility directly
- A2 is a refinement: behavior when tax is salient reveals true preferences

Efficiency Cost with Salience Effects

- Two demand curves: price-demand $x(p, 0, Z)$ and tax-demand $x(p_0, t^S, Z)$
- 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

- Without income effects ($\frac{\partial x}{\partial Z} = 0$), excess burden of introducing a small tax t^S is

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

- 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:

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

- Compensated price demand: $dx^c / dp = dx / dp + x dx / dZ$
- Compensated tax demand: $dx^c / dt^S = dx / dt^S + x dx / dZ$
- 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

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

- 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
 - Value of tax simplification
 - Tax smoothing
- 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 2012

- ① Commodity Taxation: Ramsey Rule
- ② Capital Income Taxation and Retirement Savings
- ③ Income Taxation: Mirrlees Model
- ④ Optimal Transfer Programs

Optimal Commodity Taxation: Introduction

- 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?

Overview of Optimal Taxation

- From an efficiency perspective, would finance government purely through lump-sum taxation
- With redistributive concerns, would ideally levy individual-specific lump sum taxes
 - Tax higher-ability individuals a larger lump sum
- Problem: cannot observe individuals' types
 - Therefore must tax economic outcomes such as income or consumption, which leads to distortions

Ramsey vs. Mirrleesian Approaches

- Two approaches to optimal taxation
 - ① Ramsey: restrict attention linear ($t \cdot x$) tax systems
 - ② Mirrleesian: Non-linear ($t(x)$) tax systems, with no restrictions on $t(x)$
- Ramsey approach: rule out possibility of lump sum taxes by assumption and consider linear taxes
- Mirrleesian approach: permit lump sum taxes, but model their costs in a model with heterogeneity in agents' skills

Four Central Results in Optimal Tax Theory

- ① Ramsey (1927): inverse elasticity rule
- ② Chamley (1985), Judd (1986): no capital taxation in infinite horizon Ramsey models
- ③ Diamond and Mirrlees (1971): production efficiency
- ④ Atkinson and Stiglitz (1976): no consumption taxation with optimal *non-linear* income taxation

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

Ramsey Model: Key Assumptions

- ① Lump sum taxation prohibited
- ② Cannot tax all commodities (e.g., 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)
 - As in efficiency analysis, assume that individual does not internalize effect of τ_i on govt. budget
 - Captures idea that any one individual accounts for a small frac. of economy
- Individual maximizes utility

$$u(x_1, \dots, x_N, l)$$

subject to budget constraint

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

- Z = non wage income, w = wage rate

Ramsey Model: Consumer Behavior

- Lagrangian for individual's maximization problem:

$$\mathcal{L} = u(x_1, \dots, x_N, l) + \alpha(wl + Z - (q_1x_1 + \dots + q_Nx_N))$$

- 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 $q = (w, q_1, \dots, q_N)$

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:

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

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

- 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:

$$\begin{aligned} (\lambda - \alpha)x_i + \lambda \sum_j \tau_j [\partial h_j / \partial q_i - x_i \partial x_j / \partial Z] &= 0 \\ \Rightarrow \frac{1}{x_i} \sum_j \tau_j \frac{\partial h_j}{\partial q_j} &= -\frac{\theta}{\lambda} \end{aligned}$$

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 lump sum tax:
 - 1 Direct value for the government of λ
 - 2 Loss in welfare for individual of α
 - 3 Behavioral effect $\rightarrow$ loss in tax revenue of $\partial(\sum_j \tau_j x_j) / \partial Z$

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

Inverse Elasticity Rule

- Introducing elasticities, we can write Ramsey formula as:

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

- Consider special case where $\varepsilon_{ij} = 0$ if $i \neq j$

- Slutsky matrix is diagonal

- Obtain classic inverse elasticity rule:

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

Ramsey Formula: Limitations

- Ramsey solution: tax inelastic goods to minimize efficiency costs
- But does not take into account redistributive motives
- Necessities likely to be less elastic than luxuries
- Therefore, optimal Ramsey tax system is likely regressive
- Diamond (1975) extends Ramsey model to take redistributive motives into account
 - Basic intuition: replace multiplier λ with average marginal utility for consumers of that good

Application of Ramsey Approach to Taxation of Savings

- Standard lifecycle model of consumption

$$\max \sum_t u_t(c_t) \text{ s.t. } q_t c_t \leq W$$

$$\text{where } q_t = (1 + \tau_t)p_t$$

$$\text{and } \tau_0 \equiv 0$$

- Consumption in each period isomorphic to consumption of different goods
 - Can apply standard Ramsey formulas to calculate τ_t^*
- Capital income tax is a constant tax θ on interest rate:

$$q_t = \frac{1}{(1 + (1 - \theta)r)^t}$$

Optimal Capital Income Tax Rate

- For any $\theta > 0$, implied tax τ_t approaches ∞ as $t \rightarrow \infty$:

$$\begin{aligned}\frac{q_t}{p_t} &= 1 + \tau_t = \left(\frac{1+r}{1+(1-\theta)r} \right)^t \\ \Rightarrow \lim_{t \rightarrow \infty} \tau_t &= \infty\end{aligned}$$

- Ramsey formula implies that optimal τ_t cannot be ∞ for any good
- Therefore optimal capital income tax rate converges to 0 in long run (Judd 1985, Chamley 1986)
 - Best policy is for govt to tax capital until it accumulates sufficient assets to fund public goods and never tax capital again

Zero Capital Taxation in Ramsey Models

- Fairly robust result in pure Ramsey framework (Bernheim 2002)
- But not robust to:
 - Allowing for progressive income taxation (Golosov, Kocherlakota, Tsyvinski 03)
 - Allowing for credit market imperfections (Aiyagari 95, Farhi and Werning 11)
 - Finitely-lived agents with finite bequest elasts. (Piketty and Saez 12)
- More general issue: are agents this forward-looking when making savings choices?
 - Return to this in the context of corrective taxes later in the course

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)
[Spring Semester]
- ④ Optimal Transfer Programs: Saez (2002)

Key Concepts for Taxes/Transfers

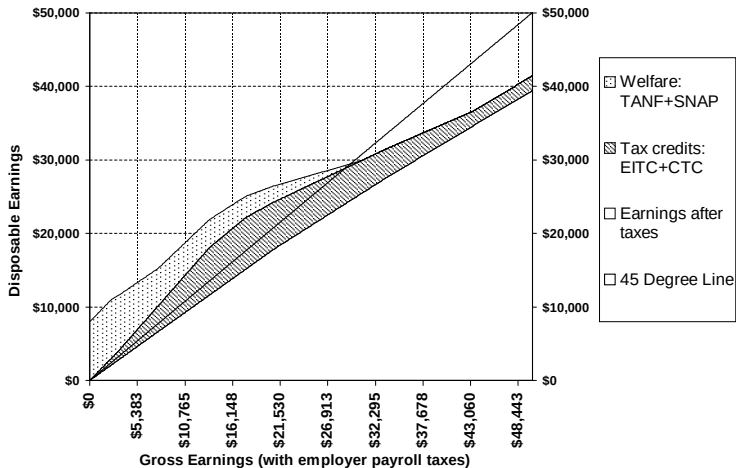
- Let $T(z)$ denote tax liability as a function of earnings z
- ① Transfer benefit with zero earnings $-T(0)$ [also called demogrant or lumpsum grant]
- ② Marginal tax rate $T'(z)$: individual keeps $1 - T'(z)$ for an additional \$1 of earnings (relevant for intensive margin labor supply responses)
- 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 $T(z^*) = 0$

US Tax/Transfer System, single parent with 2 children, 2009



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:

$$T(z) \quad : \quad 0 = \partial L / \partial T(z) = [-u'(z - T(z)) + \lambda]h(z)$$

$$\Rightarrow u'(z - T(z)) = \lambda$$

$$\Rightarrow z - T(z) = c \text{ constant for all } z$$

$$\Rightarrow c = \bar{z} - E$$

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, $T(\cdot)$ income tax

- Individuals differ in ability w distributed with density $f(w)$
- Govt social welfare maximization: Govt maximizes

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

where $G(\cdot)$ is increasing and concave

Social Welfare Function

- Mirrleesian approach: maximize weighted sum of utilities of ex-post consumption
- With equal weights G and diminishing marginal utility, we would equate everyone's income barring information constraints
- But is this what people really want?
 - Ex: many would argue that it is ok for an entrepreneur to keep the money he earned if he worked hard to get it
- Is maximizing total ex-post utility the right objective function?

Social Welfare Function

- Questions on appropriate social welfare function date to Rawls, Nozick, Sen, and others
 - Notions of “equality of opportunity” and “just deserts” rather than pure consequentialist perspective (Mankiw 2010, Weinzierl 2012)
- But no widely applied tractable framework besides Mirrleesian approach
 - One recent approach: Saez and Stantcheva’s (2012) “endogenous” welfare weights $G(c, T(w))$
 - Fertile area to make a fundamental contribution

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, had 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 :

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

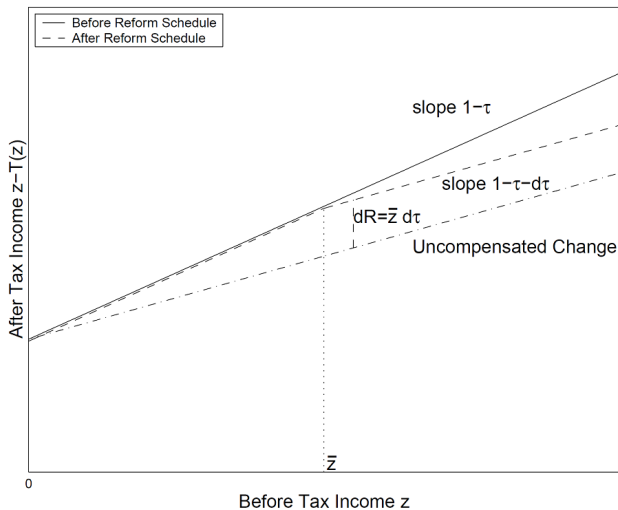
where $\varepsilon = [(1 - \tau)/Z]dz/d(1 - \tau)$ is the uncompensated taxable income elasticity w.r.t. $1 - \tau$

- Strictly inefficient to have $\tau > \tau^*$
- Technical note: why write ε as elasticity of z w.r.t. $1 - \tau$ instead of τ ?

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 (1998) 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 $1 - \tau$

FIGURE 1 – High Income Tax Rate Perturbation



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

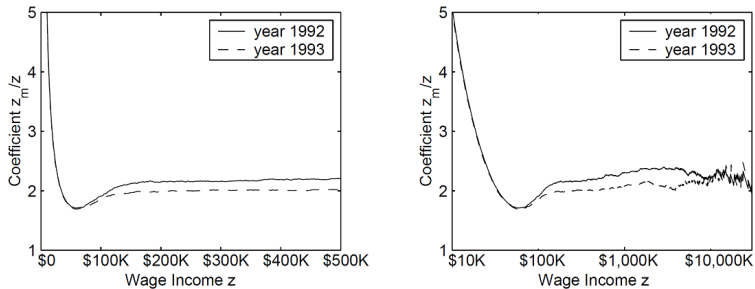
$$dM + dW + dB = Nd\tau \left\{ (1 - \bar{g})[z^m - \bar{z}] - \bar{\varepsilon} \frac{\tau}{1 - \tau} z^m \right\}$$

- 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 $\frac{z_m}{\bar{z}} = 2$
- Empirically, thickness parameter $z^m/\bar{z}$ unrelated to top tax rate τ (Saez 1999)
- How is this consistent with behavioral responses to taxation ($dz^m/d(1-\tau) > 0$)?
 - Increase in τ reduces *both* $\int_{z>\bar{z}} zh(z)dz$ and $1 - H(\bar{z})$
 - Leaves $z_m = \int_{z>\bar{z}} zh(z)dz / (1 - H(\bar{z}))$ constant
 - High taxes reduce number of people in tail, but could leave thickness of tail unchanged

Optimal Top Income Tax Rate

- Diamond (1998) shows that with Pareto skill distribution, income distribution is Pareto with parameter a invariant to τ
- With Pareto distribution ($f(z) = a \cdot k^a / z^{1+a}$), $\frac{a}{a-1} = \frac{z_m}{\bar{z}} \Rightarrow a = 2$

$$\Rightarrow \tau_{\text{TOP}}^* = \frac{1 - \bar{g}}{1 - \bar{g} + 2\bar{\varepsilon}}$$

- 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 $\bar{z} > z^S \Rightarrow z^m / \bar{z} \rightarrow 1$ when $\bar{z} \rightarrow z^T \Rightarrow$

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

- 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

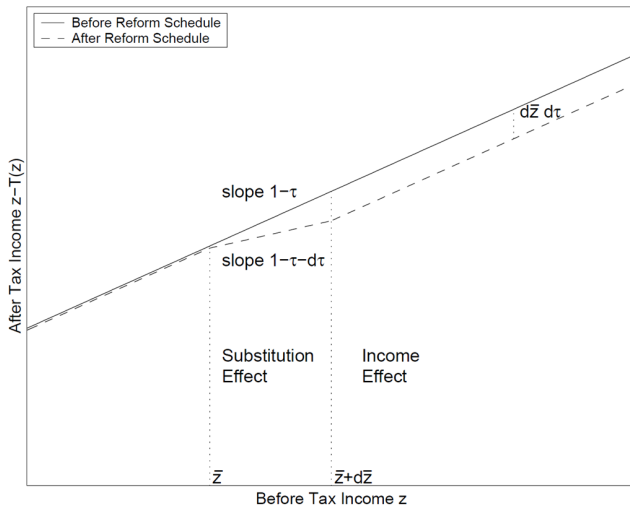
Connection to Revenue Maximizing Tax Rate

- 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 $\tau_{\text{TOP}} = \tau_{\text{MAX}} = 1/(1 + a \cdot \bar{\epsilon})$
- Example: $a = 2$ and $\bar{\epsilon} = 0.5 \Rightarrow \tau = 50\%$
- Laffer linear rate is a special case where $\bar{z} = 0$
$$\Rightarrow z^m / \bar{z} = \infty = a/(a - 1) \Rightarrow a = 1 \Rightarrow \tau_{\text{MAX}} = 1/(1 + \bar{\epsilon})$$

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 $T(\cdot)$]
- 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 $z \rightarrow \infty$
 - $G(z) \rightarrow \bar{g}$, $(1 - H(z))/(zh(z)) \rightarrow 1/a$
 - $\varepsilon_{(z)} \rightarrow \bar{\varepsilon} \Rightarrow T'(z) = (1 - \bar{g})/(1 - \bar{g} + a \cdot \bar{\varepsilon}) = \tau_{\text{TOP}}$

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 $z > 0$
 - Without income effects, $G(0) = 1$
 - Value of lump sum grant to all equals value of public good
 - Concave SWF $\rightarrow G'(z) < 0$
- $dB > 0$ 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

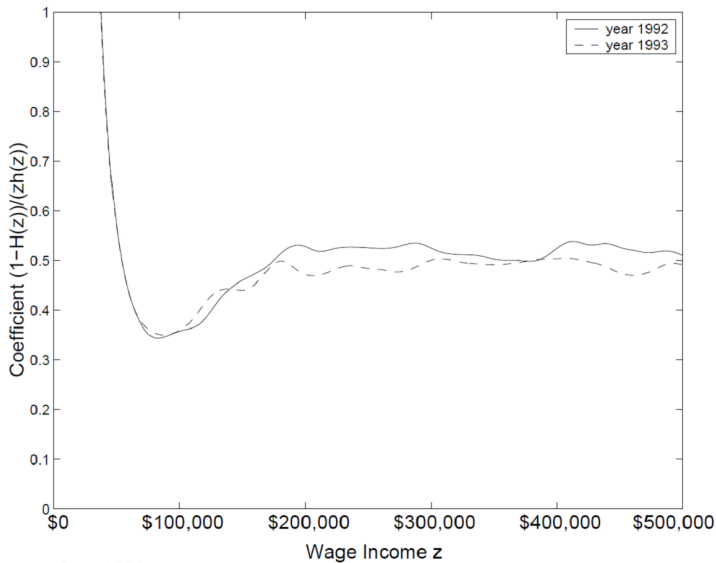
Numerical Simulations of Optimal Tax Schedule

- Formula above is a condition for optimality but not an explicit formula for optimal tax schedule
 - Distribution of incomes $H(z)$ endogenous to $T(\cdot)$
- Therefore need to use structural approach (specification of primitives) to calculate optimal $T(\cdot)$
- 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 $T(\cdot)$ yields **empirical** $H(z)$

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



Source: Saez 2001

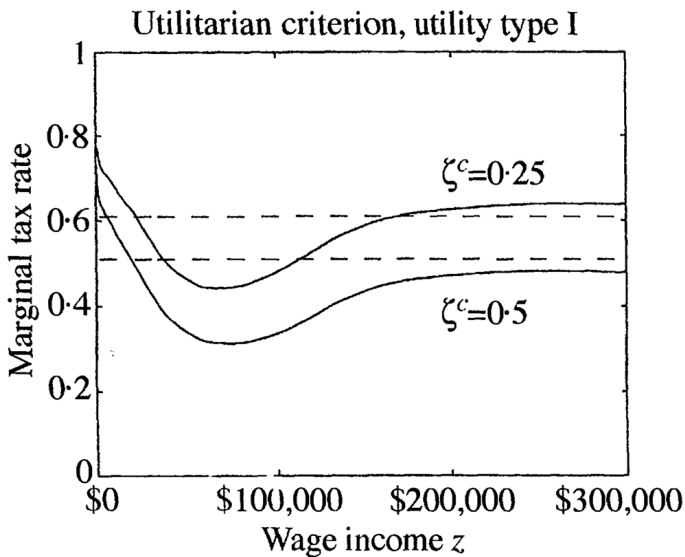
Numerical Simulations

- 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

- Atkinson and Stiglitz (1976) analyze optimal combination of income and commodity taxes
- K consumption goods $c = (c_1, \dots, c_K)$ with pre-tax price $p = (p_1, \dots, p_K)$
- Individual h has utility $u(c_1, \dots, c_K) - \phi^h(z)$
- Assumes (1) separability between c and z and (2) homogeneity of consumption sub-utility u
- Main result: commodity taxation is superfluous

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

Atkinson-Stiglitz: Implications for Capital Taxation

- Two period model: wage rate w in period 1, retired in period 2
- Let δ = discount rate, $\psi(\cdot)$ 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

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

- Capital income tax θ is equivalent to tax on c_2
- Atkinson-Stiglitz implies that $\theta^* = 0$ in the presence of an optimal income tax

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 reveal information about type and relax IC constraint

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
- In dynamic Mirrlees models, optimal capital tax is not zero (Golosov, Kocherlekota, and Tsyvinski 2003)
 - Optimum features 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

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: Mirrlees 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 lump sum grant quickly
- NIT optimal because it targets transfers to the most needy

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 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: $w_0 = 0 < w_1 < \dots < w_I$
- 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 of population earns w_i , with $\sum_i h_i = 1$
- 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 $\rightarrow \sum_i g_i h_i = 1 = \text{value of public good}$
- 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 (not true with Rawlsian SWF)

Saez 2002: Intuition for EITC

- Two types: doctors (wage w_h) and plumbers (w_l)
- Both can choose whether to work, but doctors cannot become plumbers
- Transfer to 0 income individuals $\rightarrow$ help plumbers but distort doctors' incentives to work
- Transfer to those with income of $w_l \rightarrow$ still help plumbers, but do not distort doctors' incentives
- Therefore better to have a larger transfer to w_l than 0, i.e. have a subsidy for work = EITC
- In pure ext margin model, transfer T_1 only distorts behavior of type 1
 - Higher types don't move down
 - But transfer T_0 distorts behavior of all types on extensive margin

Saez 2002: Optimal Tax Formula

- Small tax cut $dT_i < 0 \Rightarrow dc_i = -dT_i > 0$. Three effects:
 - ① Fiscal Effect: loss of tax revenue $dM = -h_i dc_i$
 - ② Welfare Effect: each worker in job i gains dT_i so welfare gain $dW = h_i g_i dc_i$
 - No first order welfare loss for switchers
 - ③ Behavioral Effect: $dh_i = e_i h_i dc_i / (c_i - c_0)$
→ 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

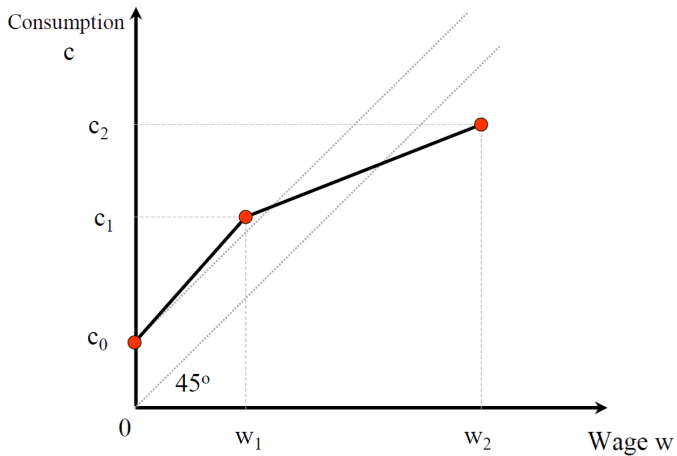


Figure 3a: Optimal Tax/Transfer Derivation (assuming $g_1 > 1$)

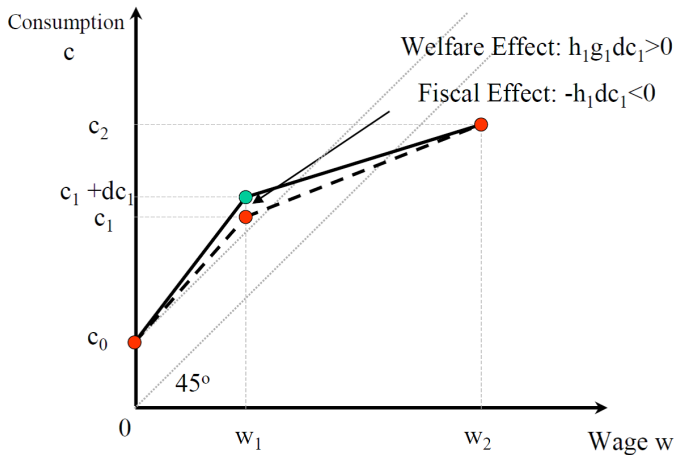


Figure 3a: Optimal Tax/Transfer Derivation (assuming $g_1 > 1$)

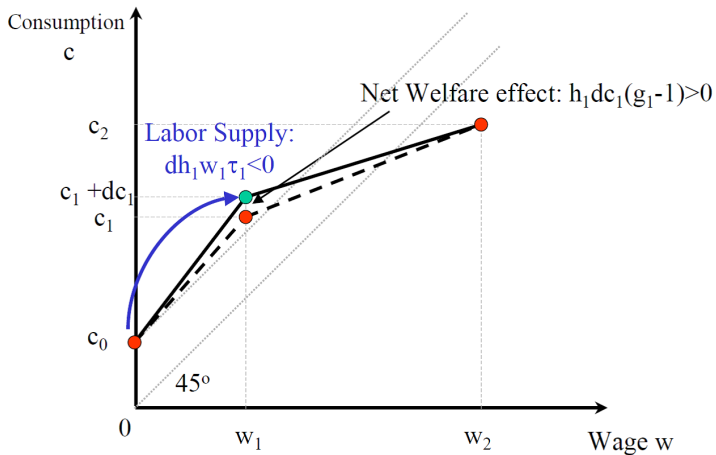
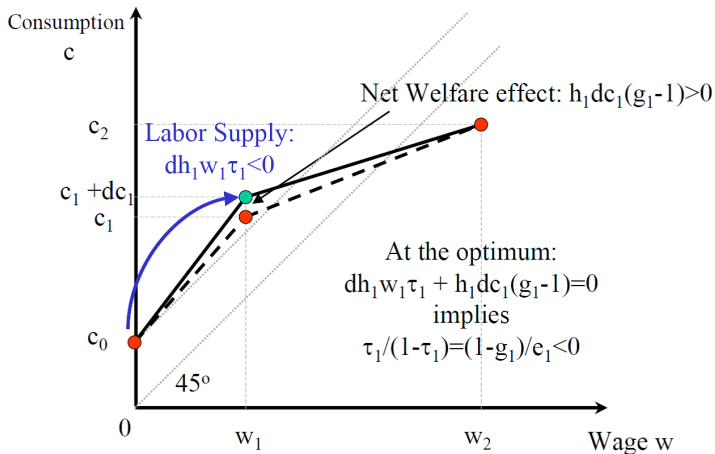
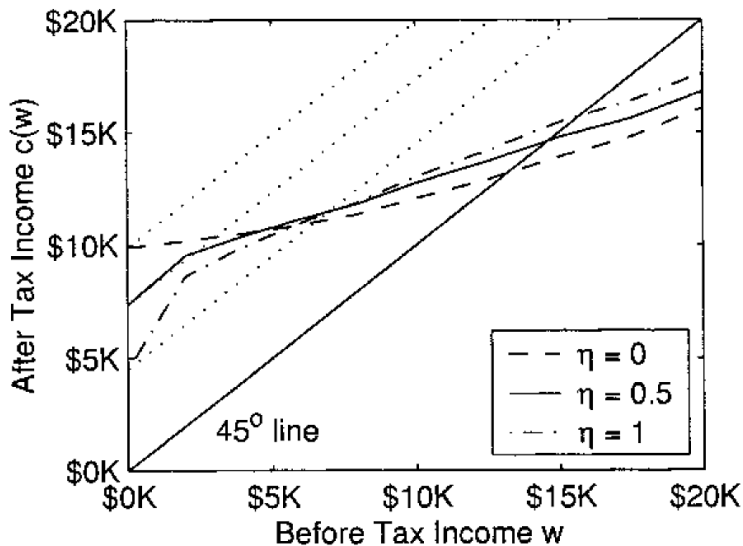


Figure 3a: Optimal Tax/Transfer Derivation (assuming $g_1 > 1$)



- 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

$$v = 1, \epsilon_H = 0.25, \epsilon_L = 0.25$$



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

- 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 ($> 6\text{ft}$) 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

- In first-best full information model, no reason for in-kind transfers
 - In-kind transfer is tradeable at market price $\rightarrow$ in-kind equivalent to cash
 - In-kind transfer non-tradeable $\rightarrow$ 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 :

$$U_p = 2S - .5W$$

$$U_r = S - 1W$$

- 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

$$U_p = 100$$

$$U_r = 50$$

$$SWF = 150$$

- 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 assumption in 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 $T(\cdot)$ 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, severance payments

Public Economics Lectures

Part 5: Income Taxation and Labor Supply

Raj Chetty and Gregory A. Bruich

Harvard University
Fall 2012

- 1 Labor Supply Elasticity Estimation: Overview
- 2 Non-linear budget set methods
- 3 Summary of elasticity estimates in static models
- 4 Intertemporal Labor Supply Models
- 5 Elasticity of Taxable Income
- 6 Micro vs Macro Elasticities
- 7 Implications for Preference Parameters

- 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 Gieritz (*JEL* 2011)

Theoretical Issues in Estimation

- Labor supply elasticity is a parameter of fundamental importance for income tax policy
 - Optimal tax rate depends inversely on $\varepsilon^c = \frac{\partial \log l}{\partial \log w} \bigg|_{U=\bar{U}}$, the compensated wage elasticity of labor supply
- First discuss econometric issues that arise in estimating these (and other) elasticities
- Use a simple labor supply model to organize the empirical issues

Baseline Labor-Leisure Choice Model: Key Assumptions

- 1 One period
- 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

$$\log l = \alpha + \varepsilon \log(1 - \tau)w$$

- Here y does not matter because u is quasilinear
- Log-linearization of first order condition for general utility $u(c, l)$ would yield a labor supply fn of the form:

$$\log l = \alpha + \varepsilon \log(1 - \tau)w - \eta y$$

- 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 Progressive taxes [non-linear budget set methods]
- ⑤ Frictions [macro comparisons, bounds]

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

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

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

Measurement Error and Division Bias

- Mis-measurement of hours causes a spurious link between hours and wages
- Estimate a regression of the following form:

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

- Then

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

- Problem: $\mathbb{E}\hat{\beta}_2 \neq \beta_2$ 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 $I_i > 0$
- This can bias OLS estimates: low wage earners must have very high unobserved propensity to work to find it worthwhile
- In cross-sections, requires a parametric selection correction (e.g. Heckman 1979)
- Non-parametric approach: use panel data to estimate within-person intensive-margin changes

Extensive vs. Intensive Margin

- Related issue: want to characterize 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)
- 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 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

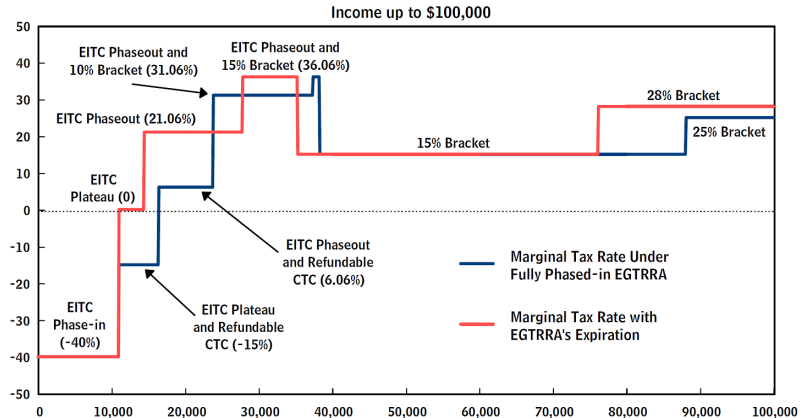
Progressive Taxes and Labor Supply

- 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.

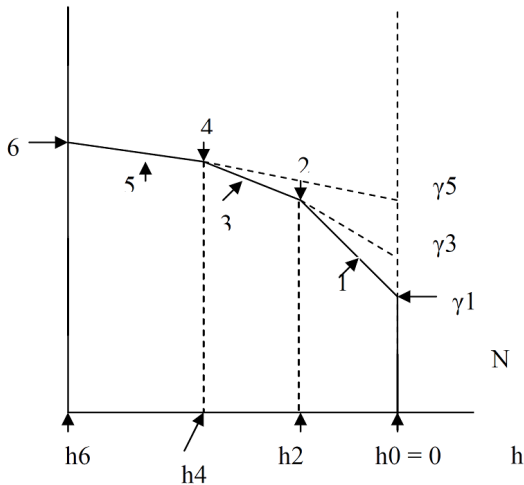
Effective Marginal Federal Income Tax Rates for a Married Couple with Two Children Before and After the Expiration of EGTRRA

(Percent)



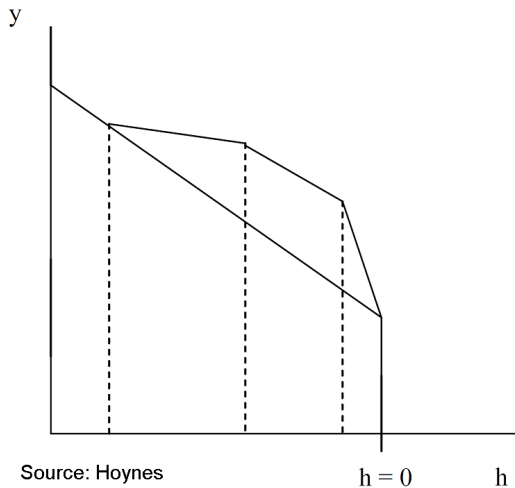
Note: EGTRRA is the 2001 Bush Tax Reform

Example 1: Progressive Income Tax

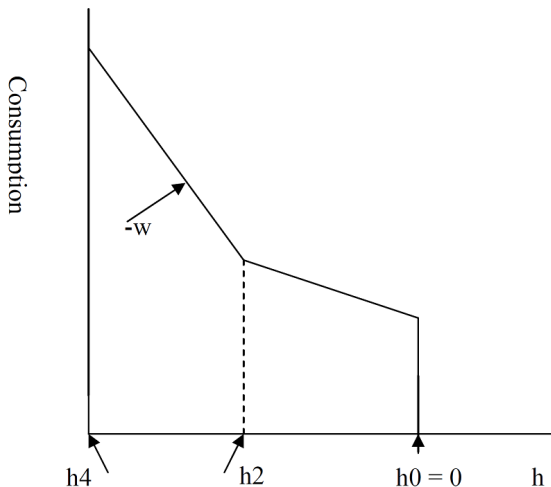


Source: Hoynes

Example 2: EITC

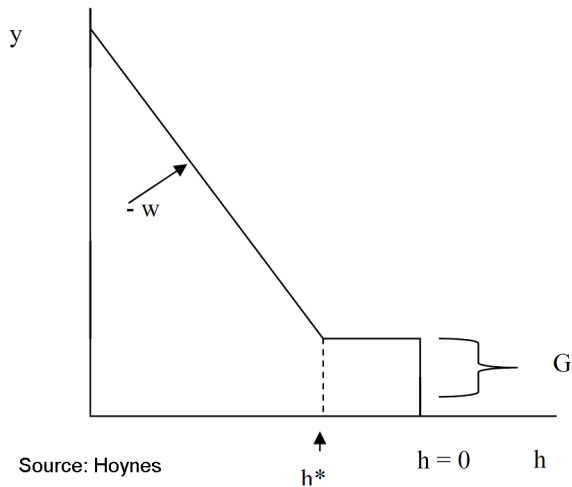


Example 3: Social Security Payroll Tax Cap



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

Optimization Frictions

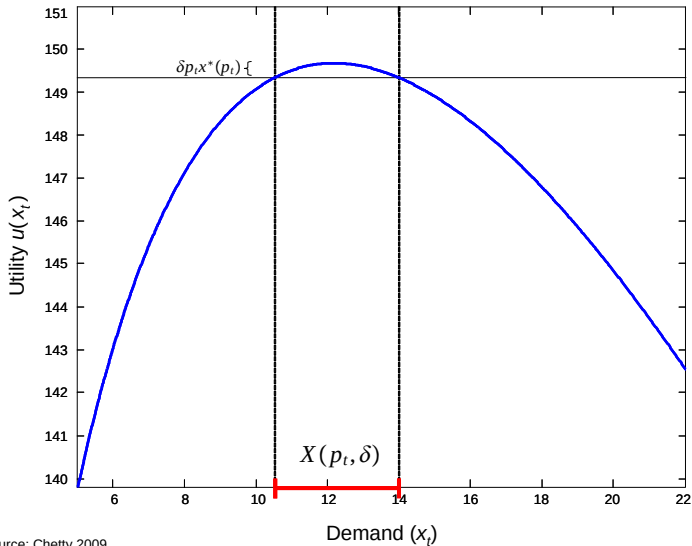
- Standard methods assume that agents can costlessly adjust hours of work
- In practice, most hours changes occur with job switches (Altonji and Paxson 1992)
- And many individuals may be inattentive to change in tax rates
- Implies that long-run impacts of policies may not be identified from short run variation

- Chetty (2012) formalizes how frictions affect identification of elasticities
- Agents can choose any x_t that generates a utility loss less than exogenous threshold δ :

$$U(x_i^*) - U(x_i) < \delta p x_i^*$$

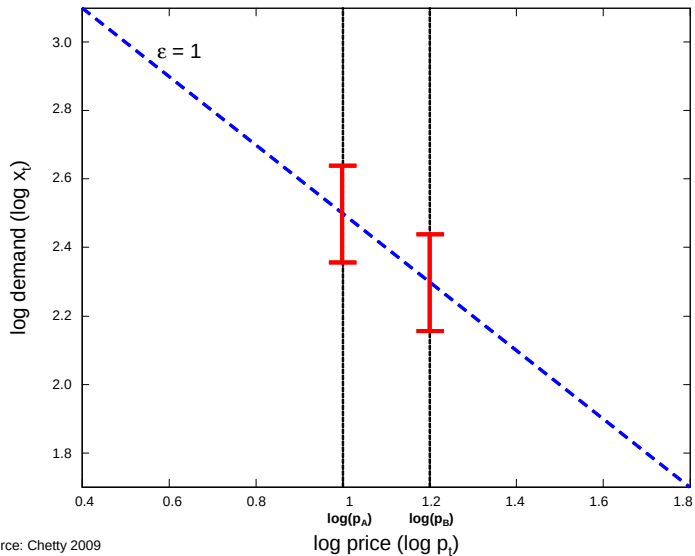
- A given price p produces a choice **set** $X(p, \delta)$ instead of a single point $x^*(p)$

Construction of Choice Set



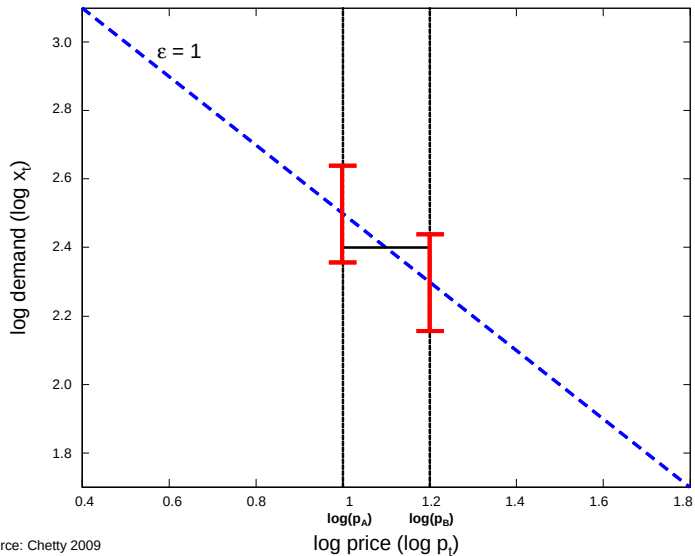
Source: Chetty 2009

Identification with Optimization Frictions



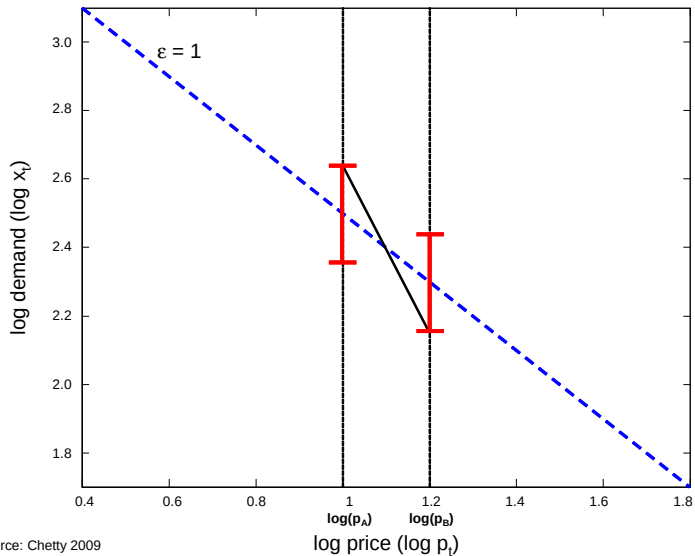
Source: Chetty 2009

Identification with Optimization Frictions



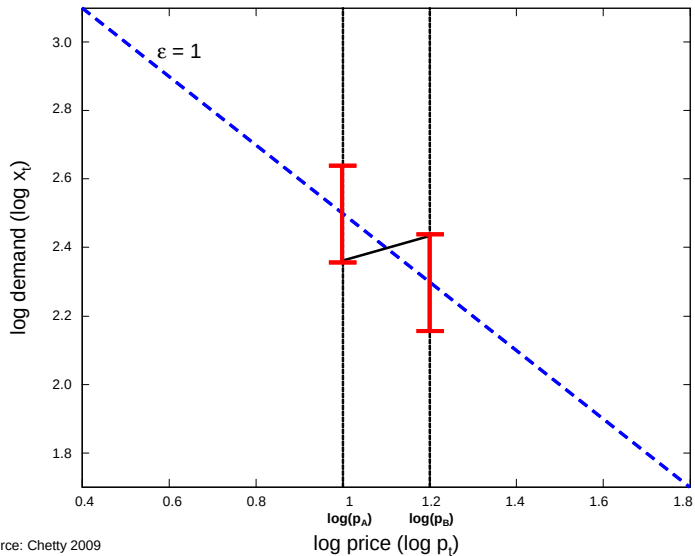
Source: Chetty 2009

Identification with Optimization Frictions



Source: Chetty 2009

Identification with Optimization Frictions



Source: Chetty 2009

- Identification problem: Multiple observed elasticities $\hat{\varepsilon}$ can be generated by a model with a given structural elasticity when $\delta > 0$
- Conversely, multiple structural elasticities consistent with observed $\hat{\varepsilon}$
- Note that this is not a finite-sample problem; does not disappear as sample size approaches ∞
- One focus of current research: how to deal with such frictions and recover ε ?

Estimating Elasticities Using Variation in Tax Rates

- Now discuss literature estimating intensive-margin elasticities using variation in tax rates for identification
- Begin by analyzing how to estimate elasticities with progressive taxes in models without frictions
 - First discuss traditional NLBS estimation using maximum likelihood methods
 - Then discuss recent literature on bunching estimators
- Then discuss most recent work on frictions and bunching estimators

Non-Linear Budget Set Methods

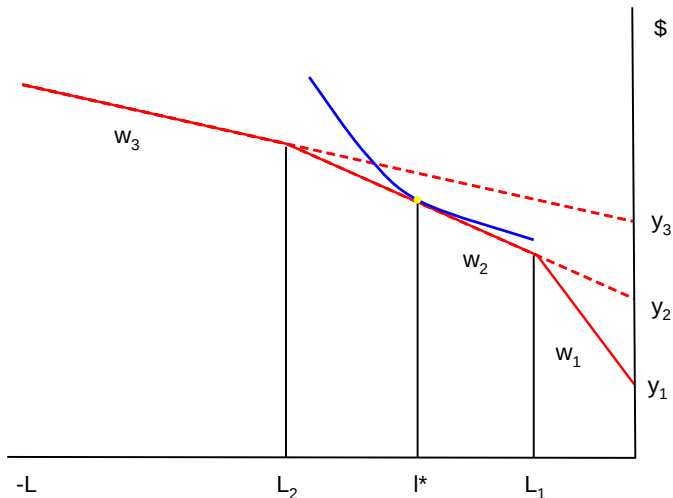
- Traditional approach to estimating elasticities with non-linear budget sets pioneered by Hausman (1981)

- Assume an uncompensated labor supply equation:

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

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

Non-Linear Budget Set Estimation: Virtual Incomes



Source: Hausman 1985

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: $0 < l_i < l_K$

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

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

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

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

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

Maximum Likelihood Estimation

- Log likelihood function is $\ell = \sum_i \log L_i$

- Final step is solving

$$\max \ell(\alpha, \beta, \gamma, \sigma)$$

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

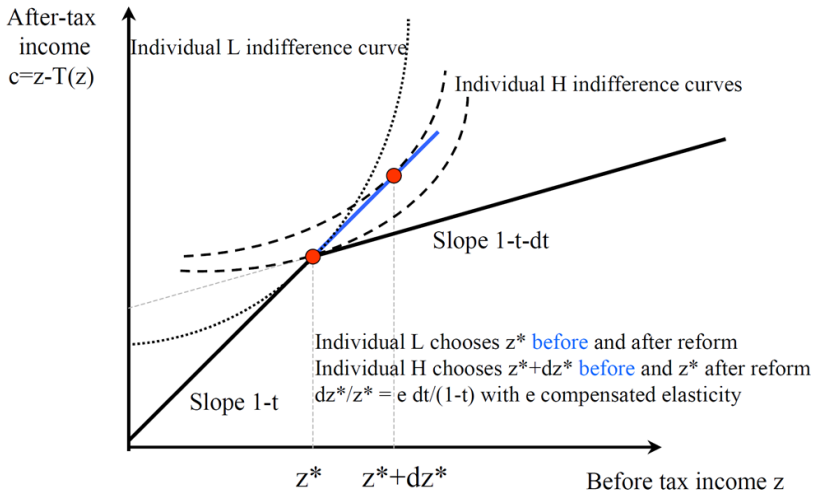
Saez 2010: Bunching at Kinks

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

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

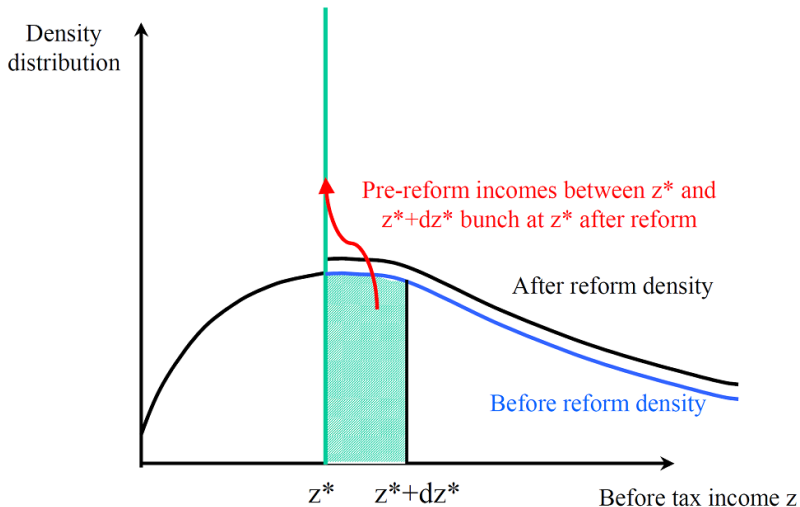
- Currently a popular approach (esp. when adapted to account for frictions) because it yields highly credible estimates

A. Indifference curves and bunching



Source: Saez (2009)

B. Density Distributions and Bunching

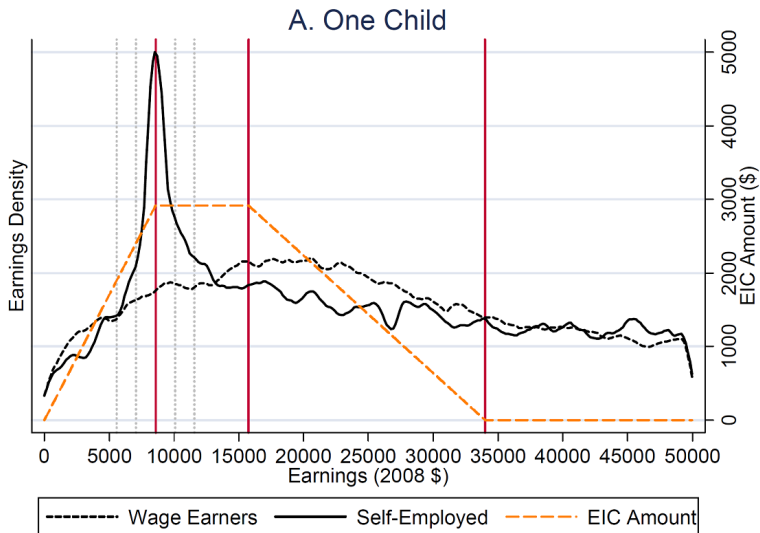


Source: Saez (2009)

Saez 2010: Bunching at Kinks

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

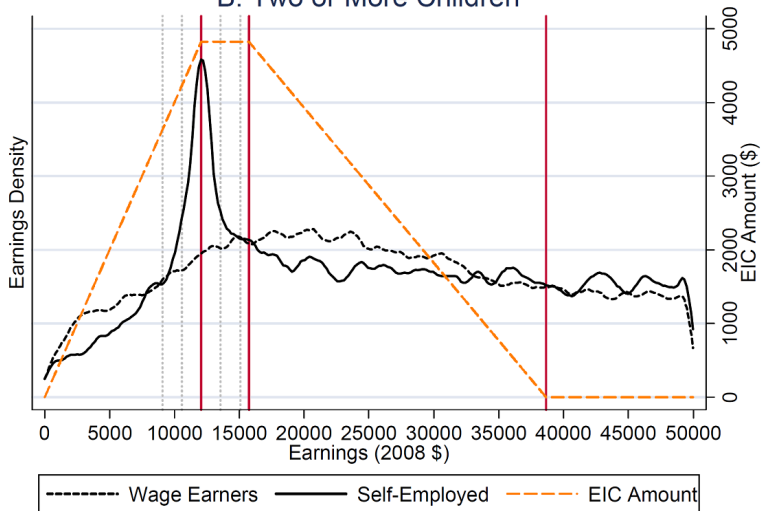
Earnings Density and the EITC: Wage Earners vs. Self-Employed



Source: Saez (2009)

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

B. Two or More Children

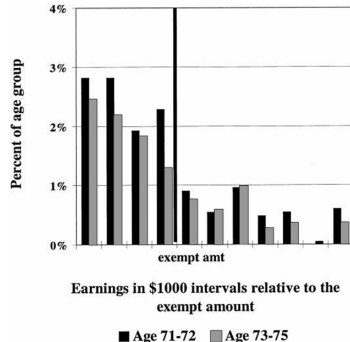
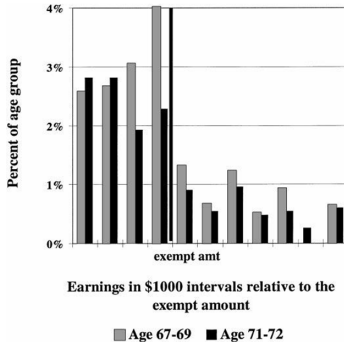


Source: Saez (2009)

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



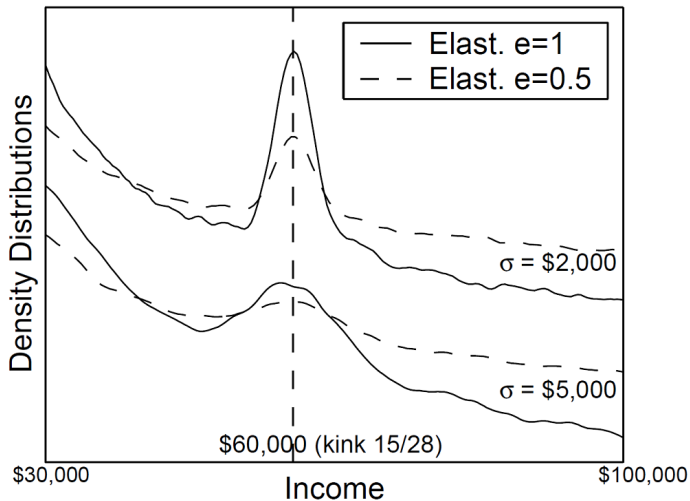
Source: Friedberg (REStat 2000)

- Estimates elasticities using Hausman method, finds relatively large compensated and uncompensated elasticities
 - Haider and Loughran (2008) replicate these results in admin. data over more years
 - Find that degree of bunching and implied elasticities are 3 times larger in data with less measurement error
 - Ironically, lost social security benefits are considered delayed retirement with an actuarial adjustment of future benefits
- So the one kink where we do find real bunching is actually not real!

Why not more bunching at kinks?

- ① Small structural elasticity
- ② Noise in income generation process

A. Simulation kink 15 to 28%



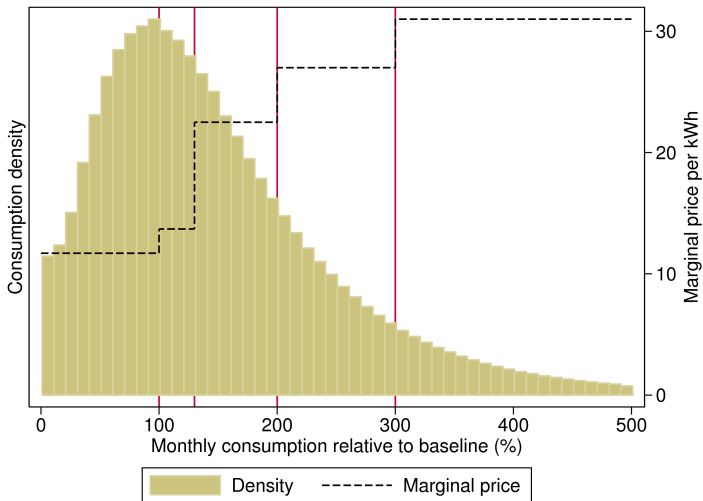
Source: Saez 2002

Why not more bunching at kinks?

- ① Small structural elasticity
- ② Noise in income generation process
- ③ Price misperceptions and salience effects
 - Liebman and Zeckhauser (2009): “Schmeduling”
 - Ito (2012): empirical evidence that average price matters more than marginal price

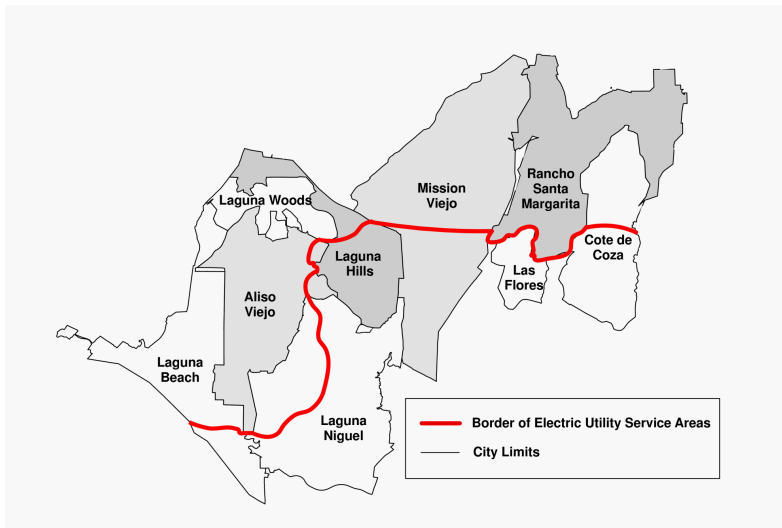
- Ito (2012) presents evidence that individuals pay more attention to average prices than marginal prices
 - Paper can be interpreted as estimating price-perception function $\tilde{p}(p + \tau)$
- Studies electricity consumption in Orange County, CA with two sources of price variation
 - Kinks in rate structure generate variation in marginal price
 - Spatial discontinuity coupled with sharp price increase in 2000 by one utility changes average price

Consumption Density and Price Schedule in 2007: Bunching Around Kink Points



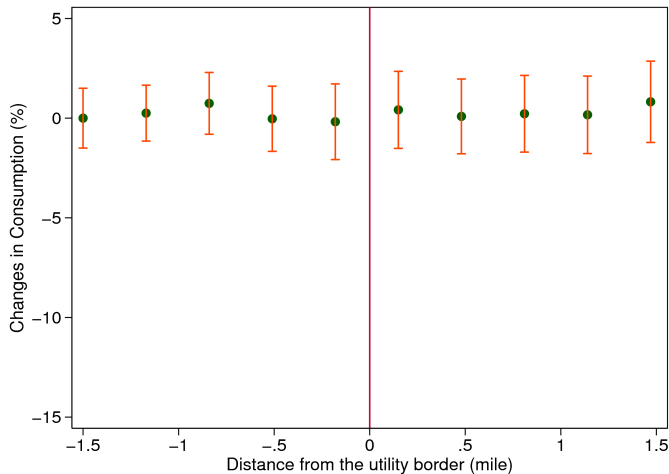
Source: Ito 2012

A Spatial Discontinuity in Electric Utility Service Areas in Orange County, California



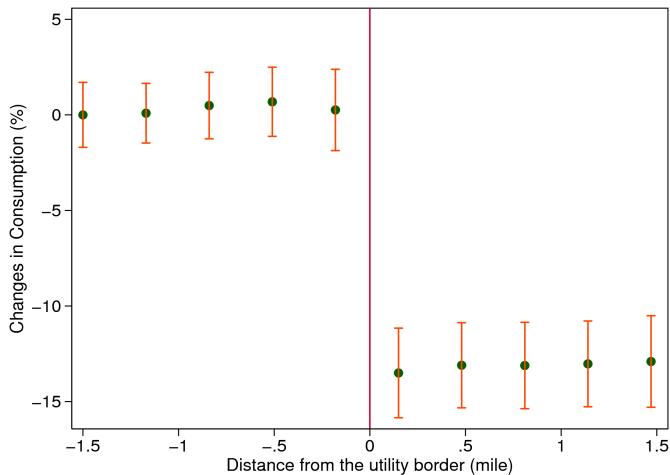
Source: Ito 2012

Changes in Consumption from July 1999 to July 2000, by Distance from the Utility Border



Source: Ito 2012

Changes in Consumption from August 1999 to August 2000, by Distance from the Utility Border



Source: Ito 2012

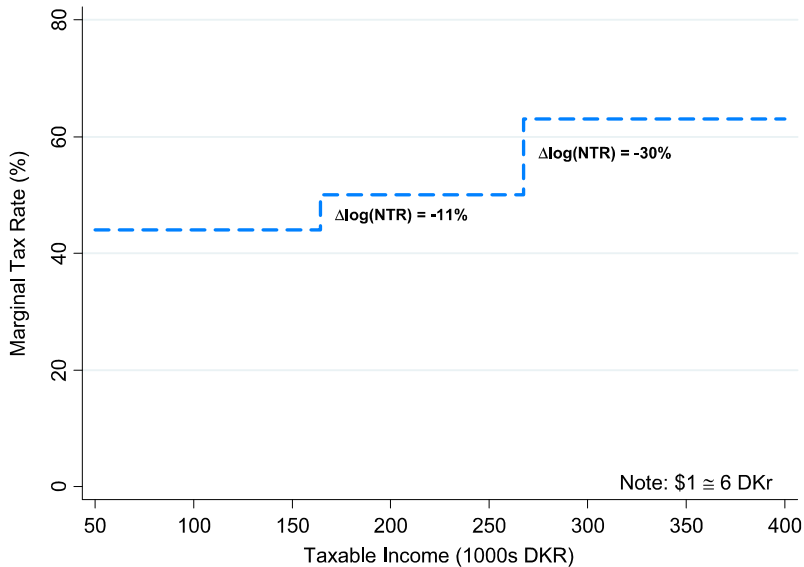
Why not more bunching at kinks?

- ① Small structural elasticity
- ② Noise in income generation process
- ③ Price misperceptions and salience effects
- ④ Optimization frictions and rigidities in job offers
 - Chetty, Friedman, Olsen, Pistaferri (2011)
 - Chetty, Friedman, Saez (2012)
 - Kleven and Waseem (2012)
 - Gelber, Jones, and Sacks (2012)
 - Bastani and Selin (2012)

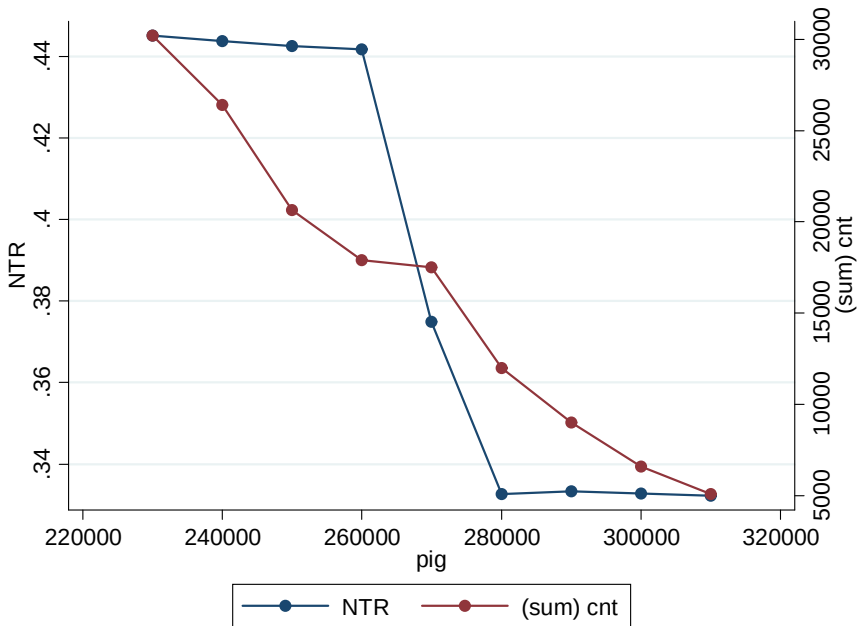
- Firms post jobs with different hours offers
- Workers draw from this distribution and must pay search cost to reoptimize
- Firm cater to aggregate worker preferences: posted distribution fits aggregate tastes
- Therefore not all workers locate at optimal hours
- Bunching at kink and observed responses to tax reforms attenuated

- Model generates two key 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

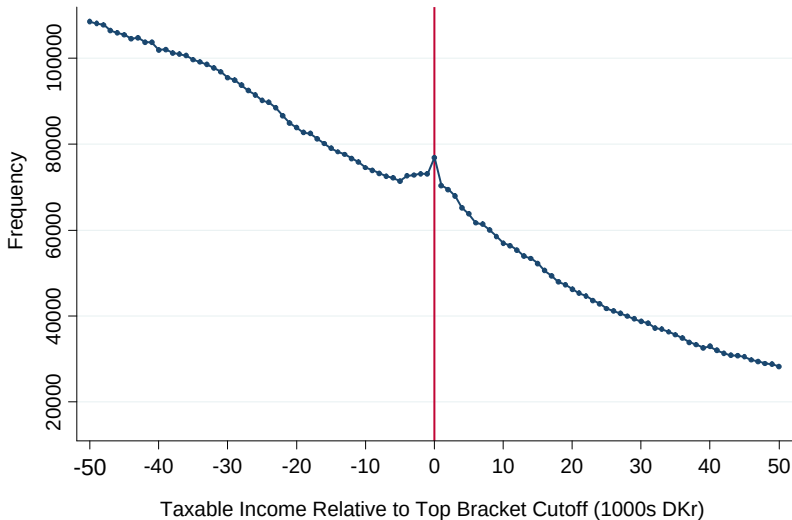
Marginal Tax Rates in Denmark in 2000



Source: Chetty et al. 2011

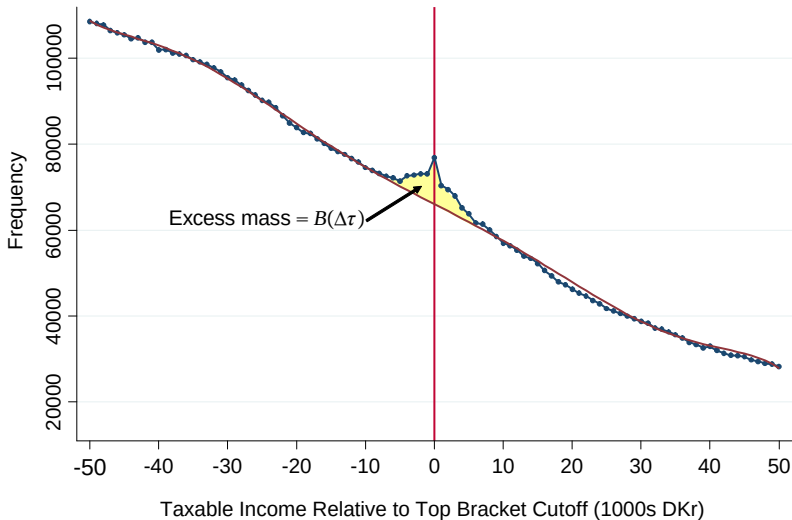


Income Distribution for Wage Earners Around Top Tax Cutoff



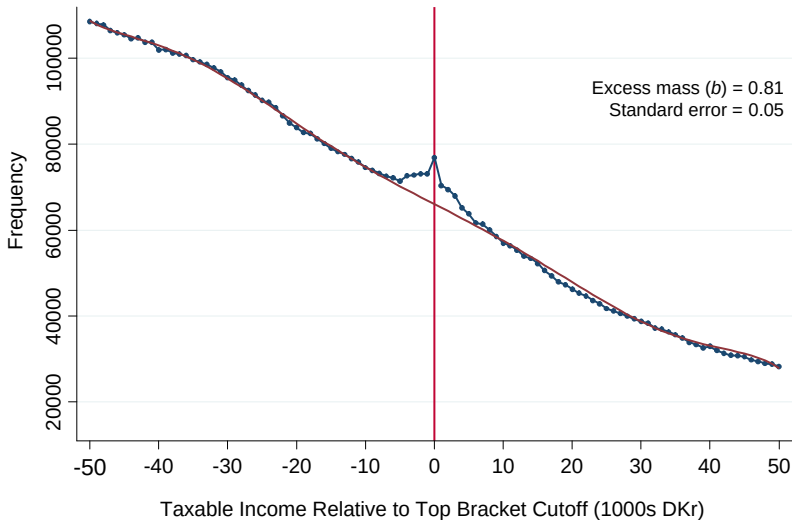
Source: Chetty et al. 2009

Income Distribution for Wage Earners Around Top Tax Cutoff



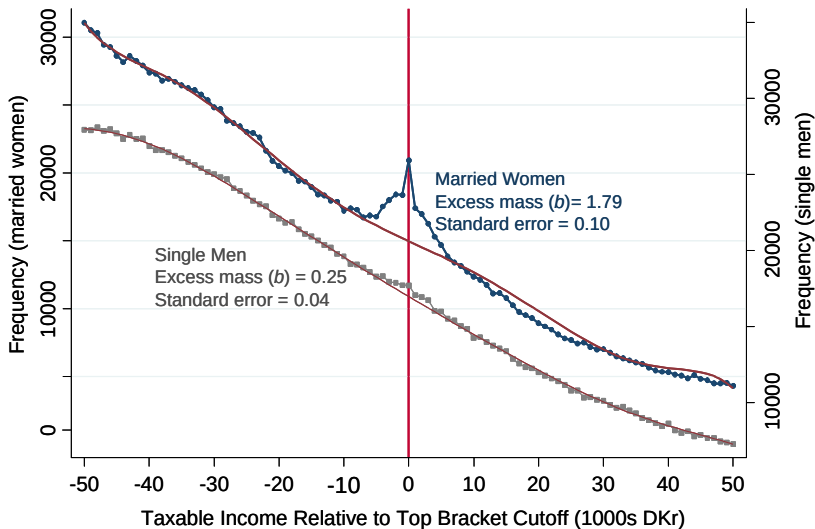
Source: Chetty et al. 2009

Income Distribution for Wage Earners Around Top Tax Cutoff



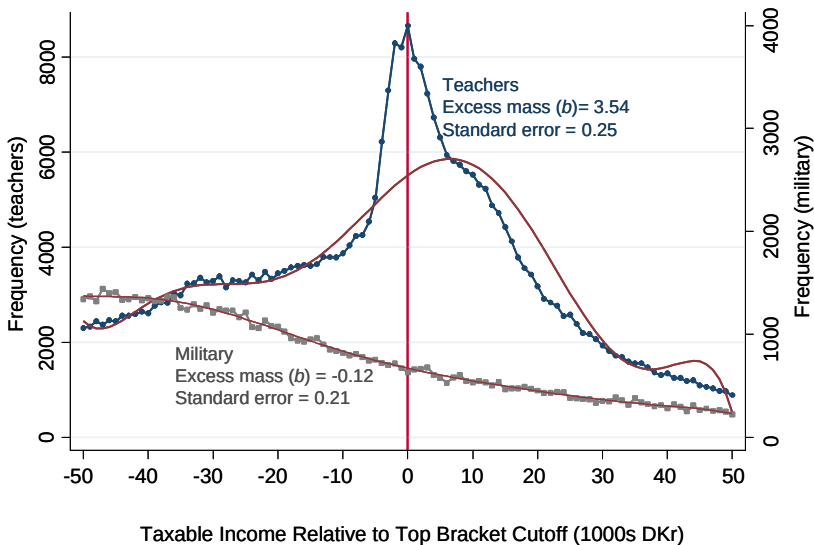
Source: Chetty et al. 2009

Married Women vs. Single Men



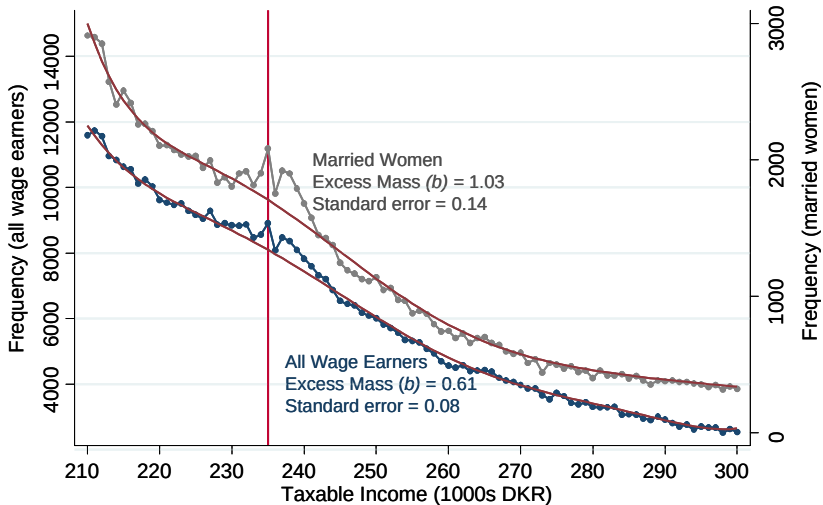
Source: Chetty et al. 2009

Teachers vs. Military



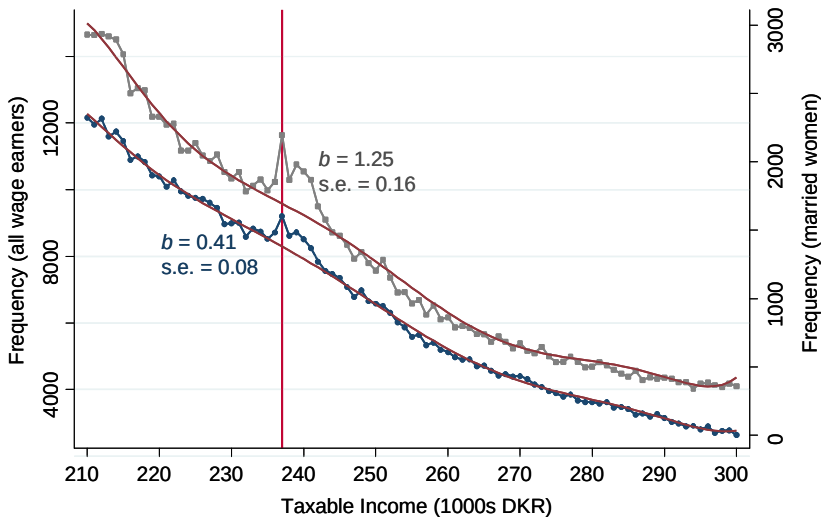
Source: Chetty et al. 2009

Taxable Income Distributions in 1994



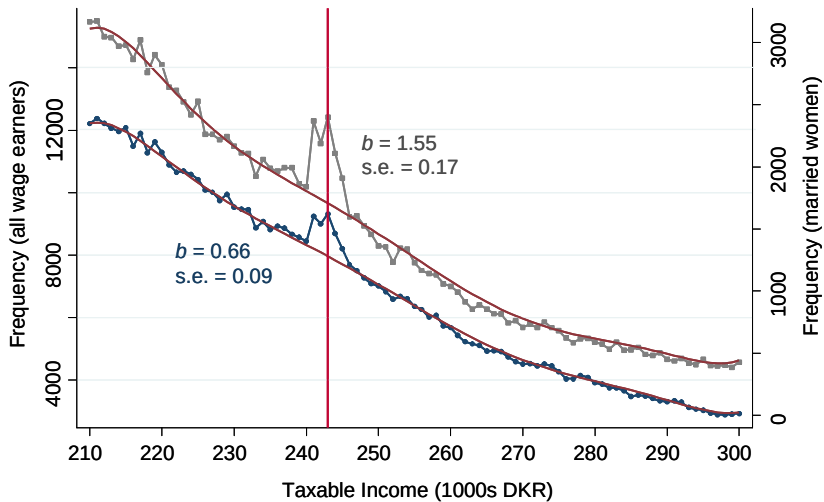
Source: Chetty et al. 2009

1995



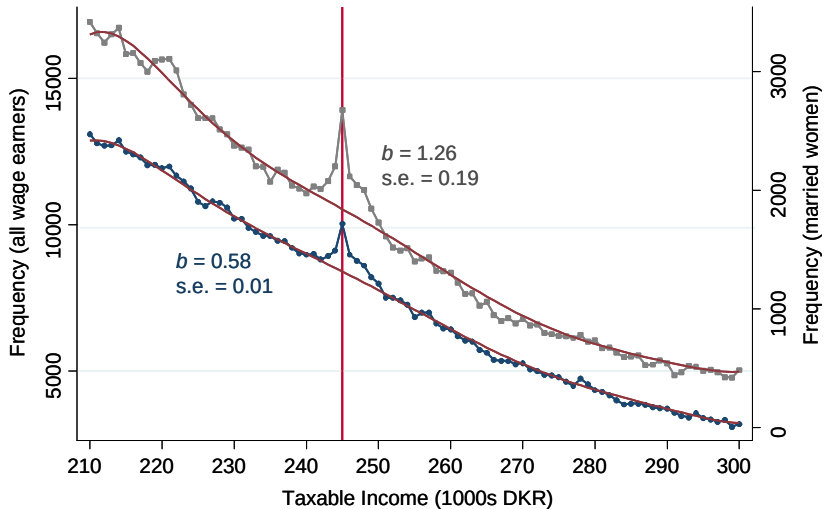
Source: Chetty et al. 2009

1996



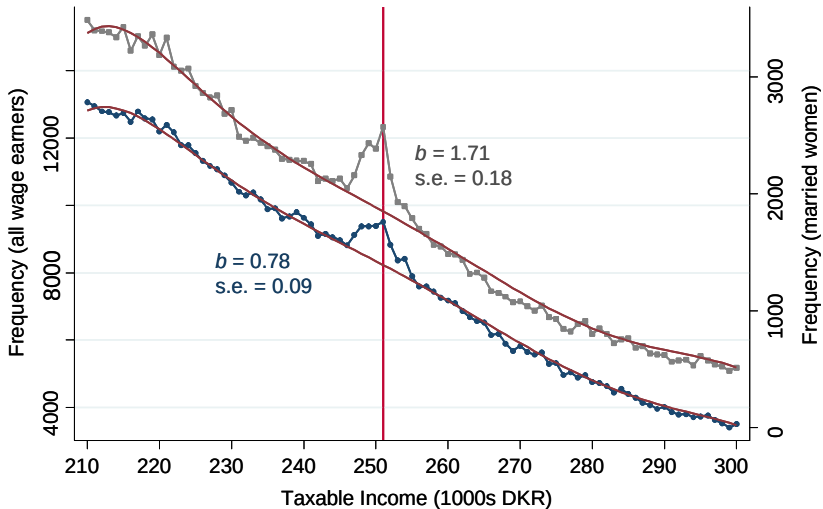
Source: Chetty et al. 2009

1997



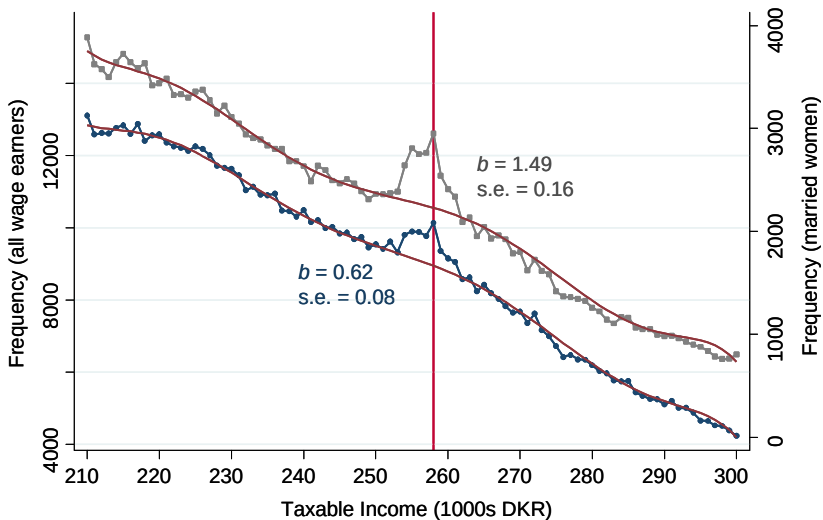
Source: Chetty et al. 2009

1998



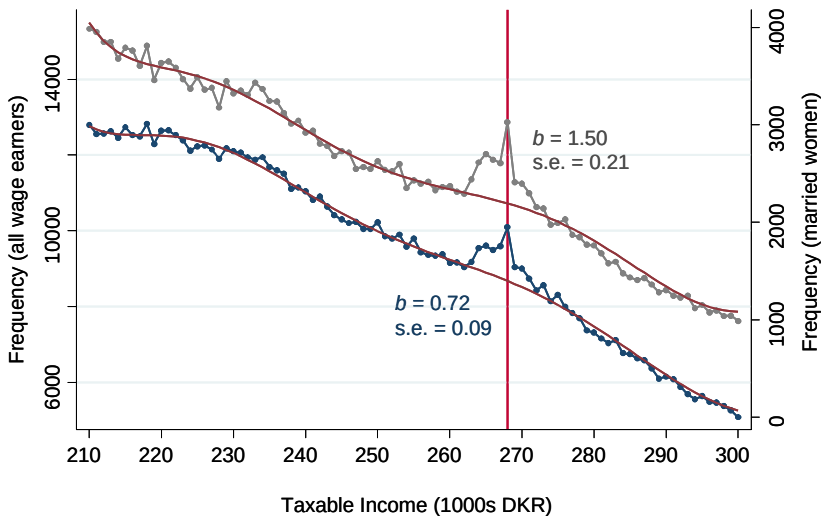
Source: Chetty et al. 2009

1999



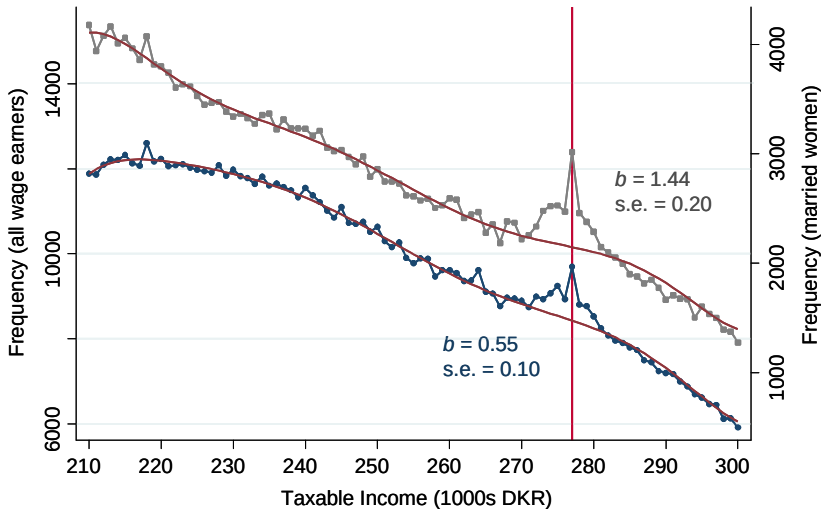
Source: Chetty et al. 2009

2000



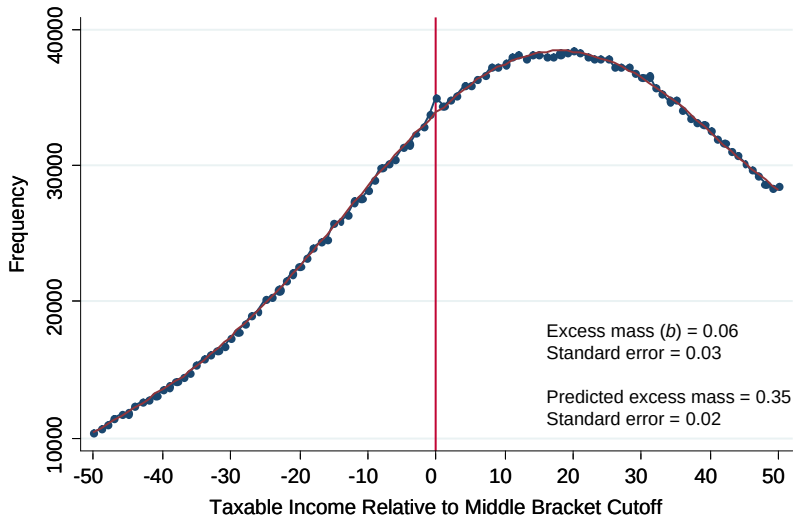
Source: Chetty et al. 2009

2001



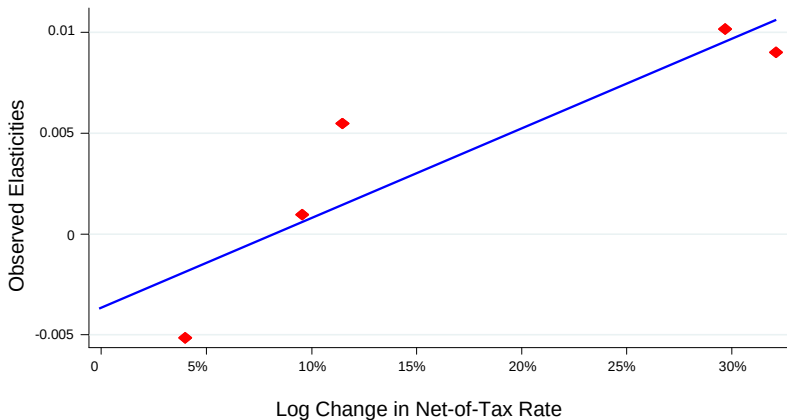
Source: Chetty et al. 2009

Married Women: Taxable Income Distribution at Middle Tax Cutoff



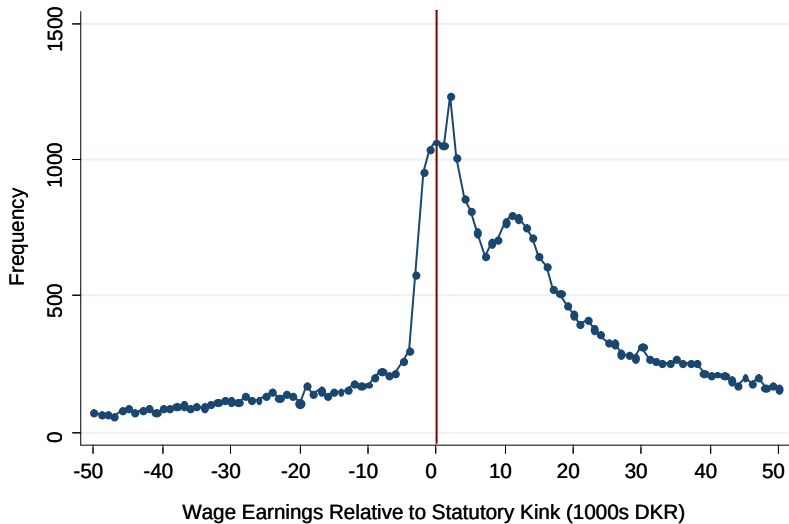
Source: Chetty et al. 2009

Observed Elasticity vs. Size of Tax Change All Wage Earners



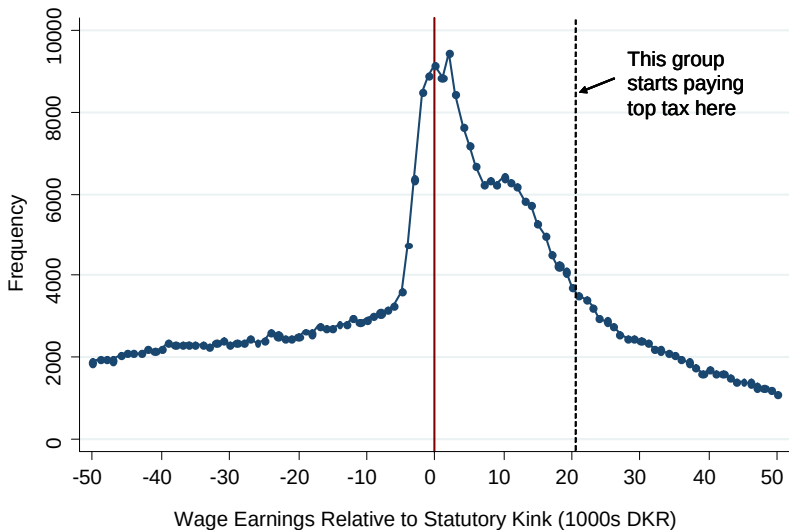
Source: Chetty et al. 2009

All Teachers

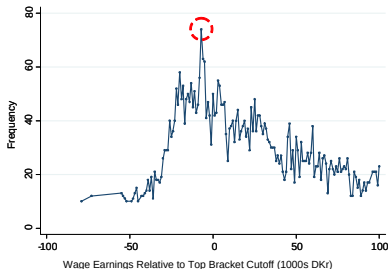


Source: Chetty et al. 2009

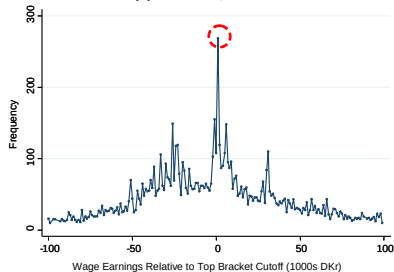
Wage Earnings Distribution: Teachers with Deductions > DKr 20,000



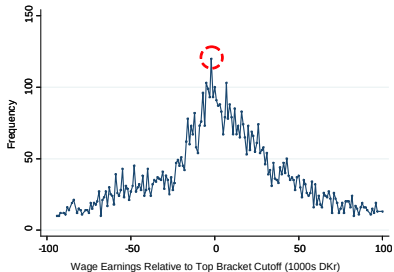
(a) Electricians, 2000



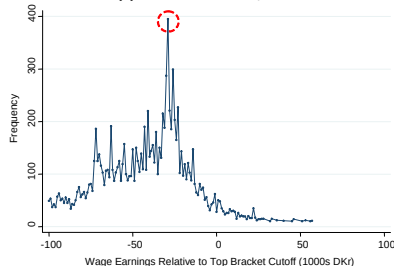
(b) Salesmen, 1996



(c) Nurses and Midwives, 2001

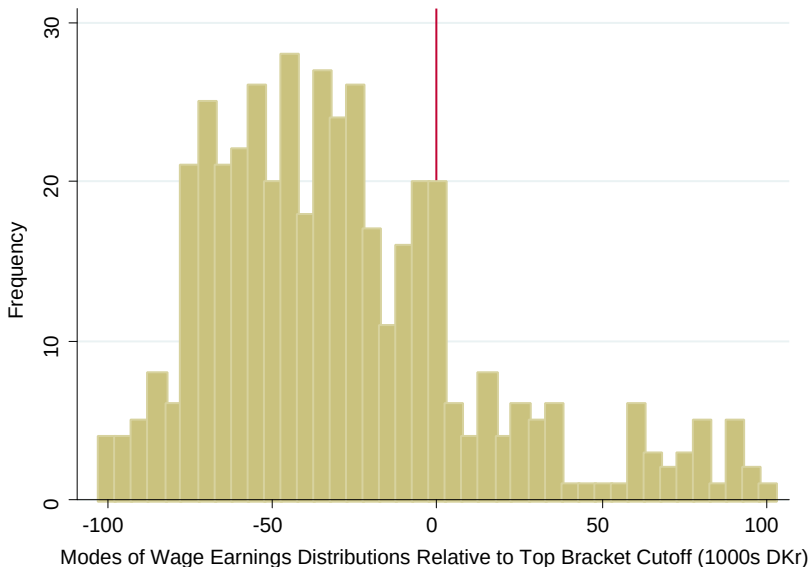


(d) Tellers and Clerks, 1998



Source: Chetty et al. 2009

Modes of Occupation-Level Wage Earnings Distributions

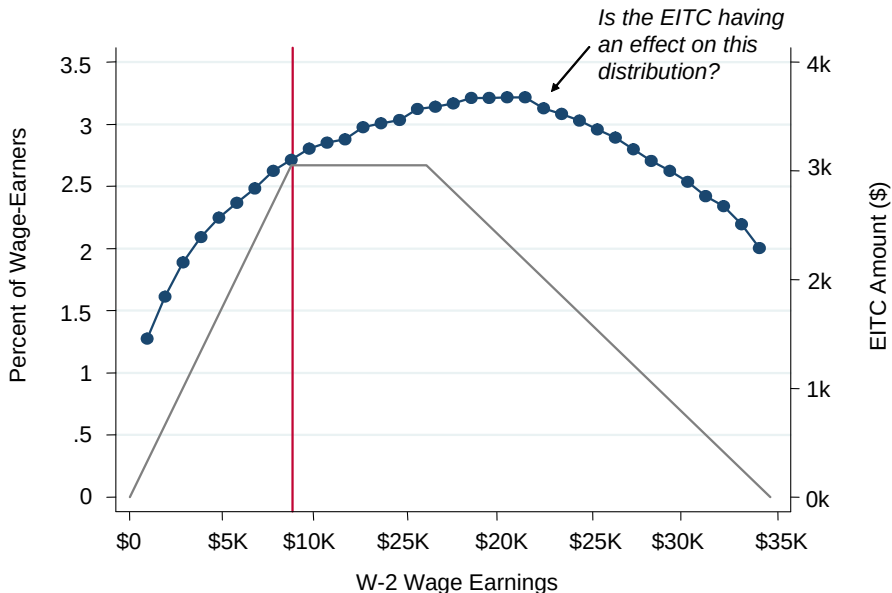


Source: Chetty et al. 2009

- Frictions and coordinated firm responses play a crucial role behavioral responses to taxation
 - NLBS models may fit data better if these factors are incorporated
 - Standard methods of estimating elasticities that ignore these factors likely to underestimate elasticities
- Limitation: does not yield an estimate of structural elasticity ε or actual impact of tax system on earnings distribn.

- Identifies impacts of EITC on earnings distribution given existence of frictions
 - Use areas with no knowledge about the EITC schedule as a counterfactual for earnings distribution in absence of EITC
 - Results suggest that earlier Danish study may have significantly understated impact of tax system on earnings distribution

Income Distribution For Single Wage Earners with One Child



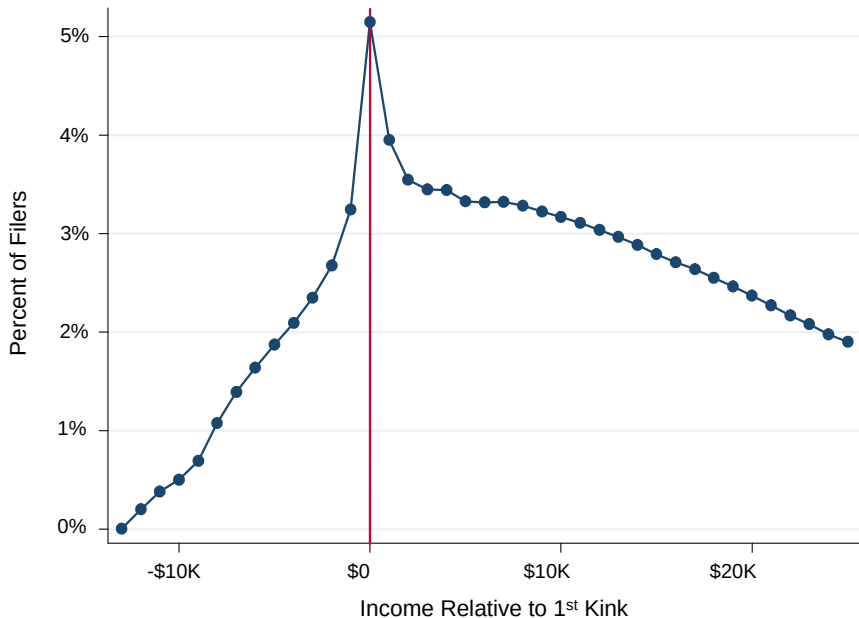
Empirical Analysis: Two Parts

- First, develop a proxy for local knowledge about EITC schedule based on income manipulation by self-employed individuals
 - Self-employment income is self-reported → easy to manipulate
 - Audit data reveal very high misreporting rates of SE income
- Second, compare W-2 wage earnings distributions across areas to uncover impacts of EITC of “real” earnings behavior
 - Wage earnings are directly reported to IRS by employers → virtually no scope for misreporting

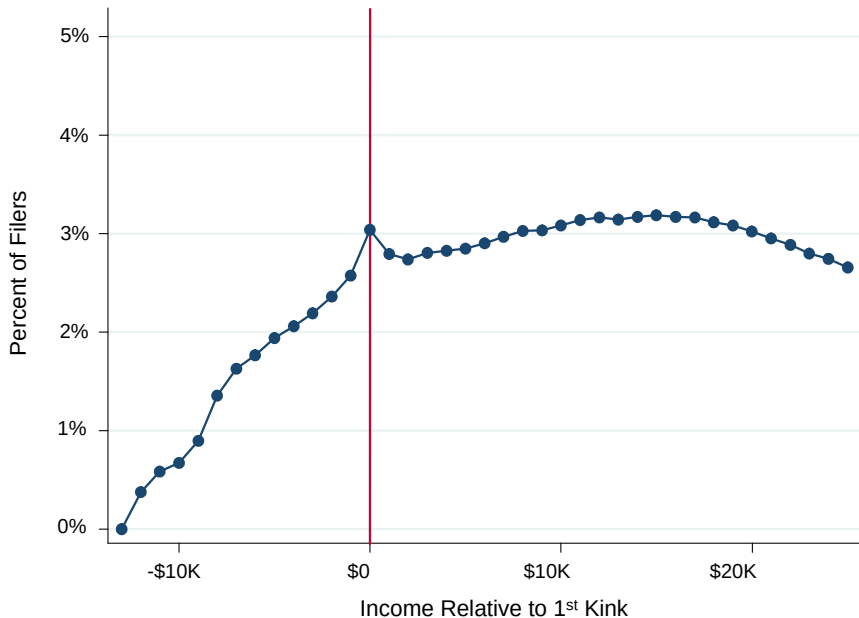
Outline of Empirical Analysis

- Step 1: Document variation across neighborhoods in sharp bunching among self-employed

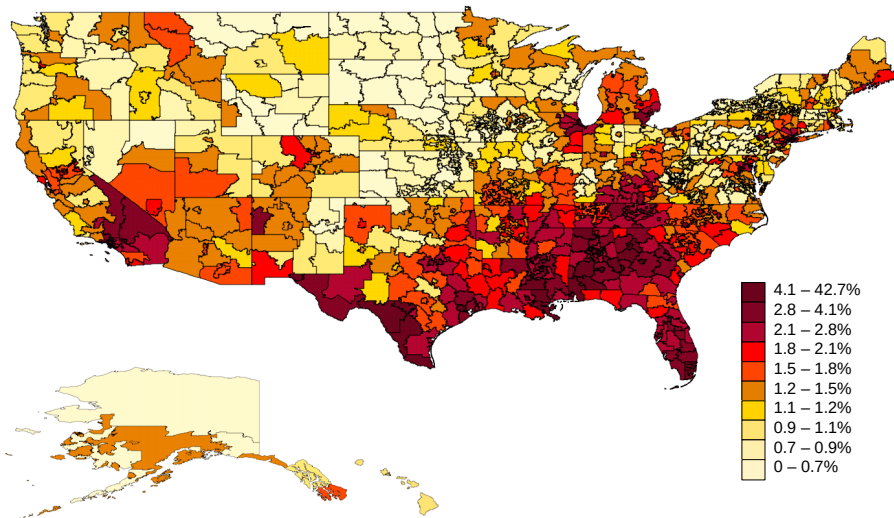
Earnings Distribution in Texas



Earnings Distribution in Kansas



Fraction of Tax Filers Who Report SE Income that Maximizes EITC Refund in 2008



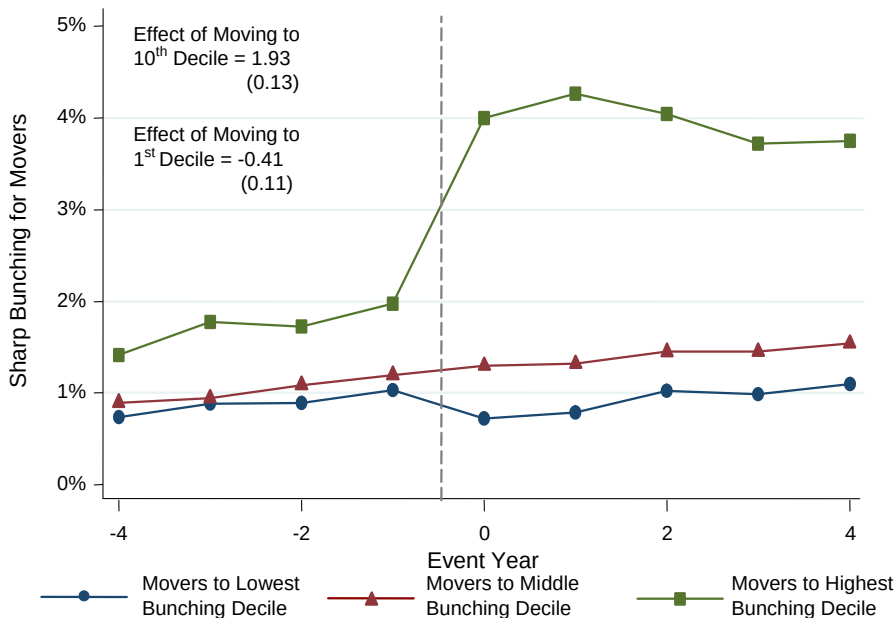
Outline of Empirical Analysis

- Step 1: Document variation across neighborhoods in sharp bunching among self-employed
- Step 2: Establish that variation in sharp bunching across neighborhoods is driven by differences in knowledge about EITC schedule

Movers: Neighborhood Changes

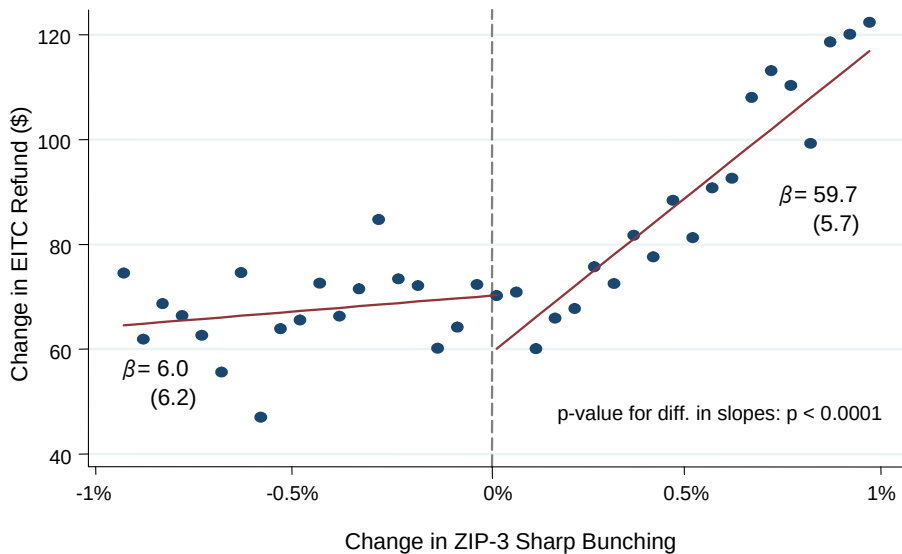
- Consider individuals who move across neighborhoods to isolate causal impacts of neighborhoods on elasticities
 - 54 million observations in panel data on cross-zip movers
 - Analyze how changes in neighborhood sharp bunching affect movers' behavior

Event Study of Sharp Bunching Around Moves

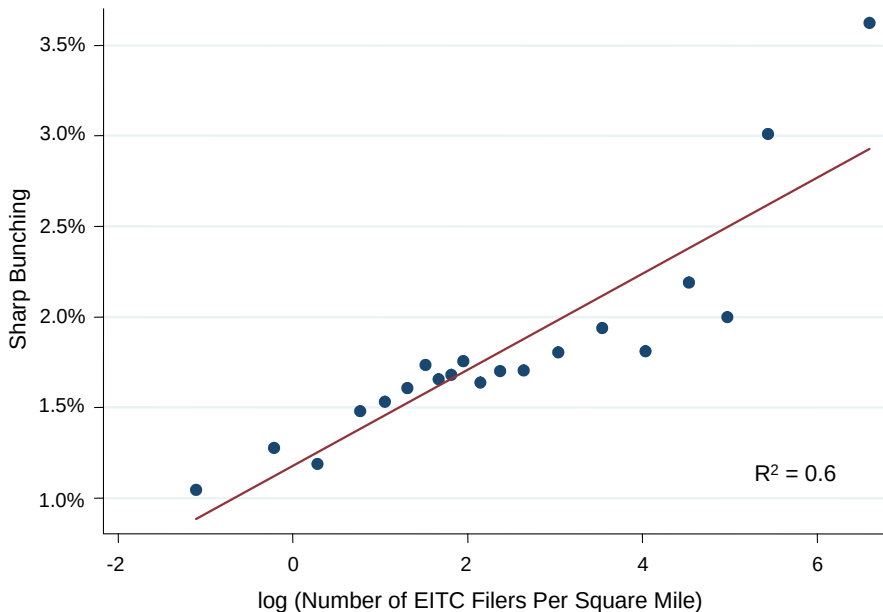


- Knowledge model predicts asymmetric impact of moving:
 - Moving to a higher-bunching neighborhood should raise EITC refund
 - Moving to a lower-bunching should not affect EITC refund

Change in EITC Refunds vs. Change in Sharp Bunching for Movers



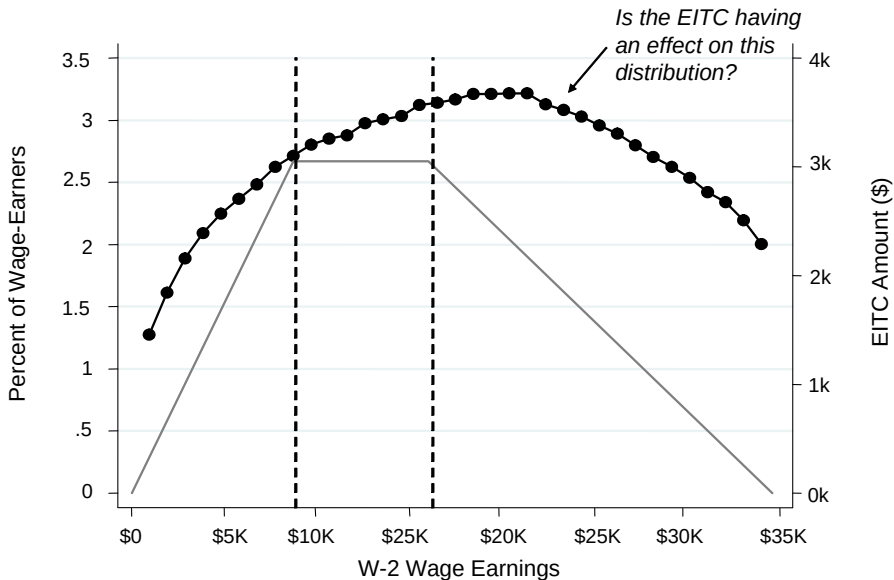
Agglomeration: Sharp Bunching vs. EITC Filer Density by ZIP Code



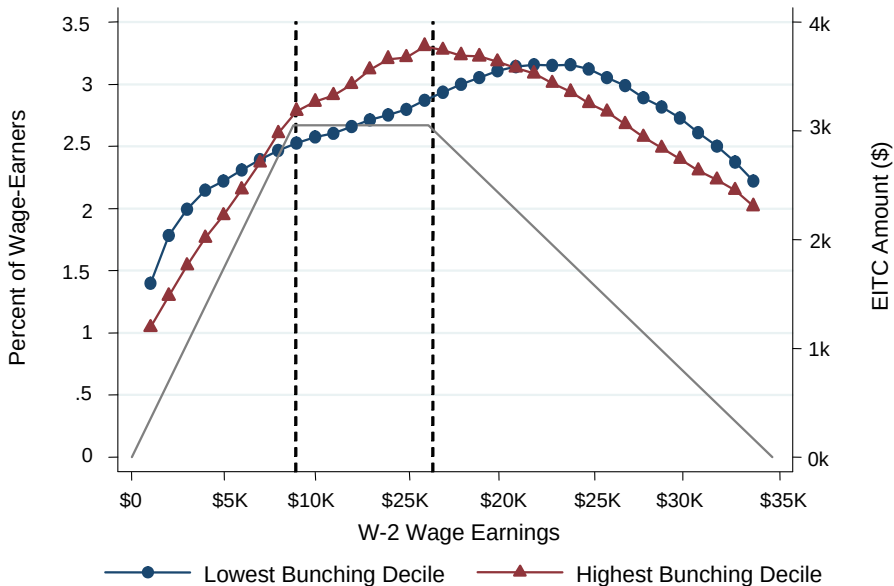
Outline of Empirical Analysis

- Step 1: Document variation across neighborhoods in sharp bunching among self-employed
- Step 2: Establish that variation in sharp bunching across neighborhoods is driven by differences in knowledge about EITC schedule
- Step 3: Compare wage earnings distributions across low- and high-knowledge neighborhoods to uncover impacts of EITC on earnings

Income Distribution For Single Wage Earners with One Child



Income Distribution For Single Wage Earners with One Child High vs. Low Bunching Areas

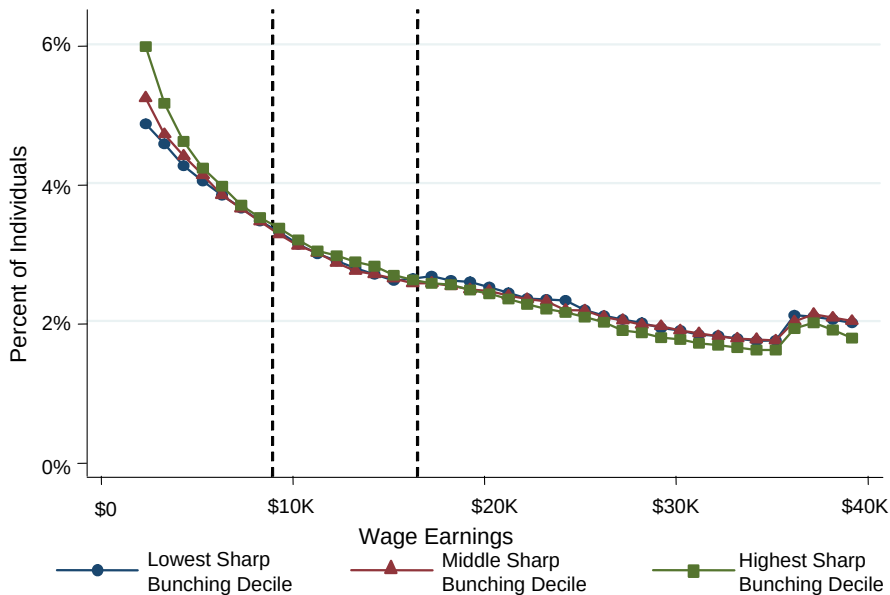


Outline of Empirical Analysis

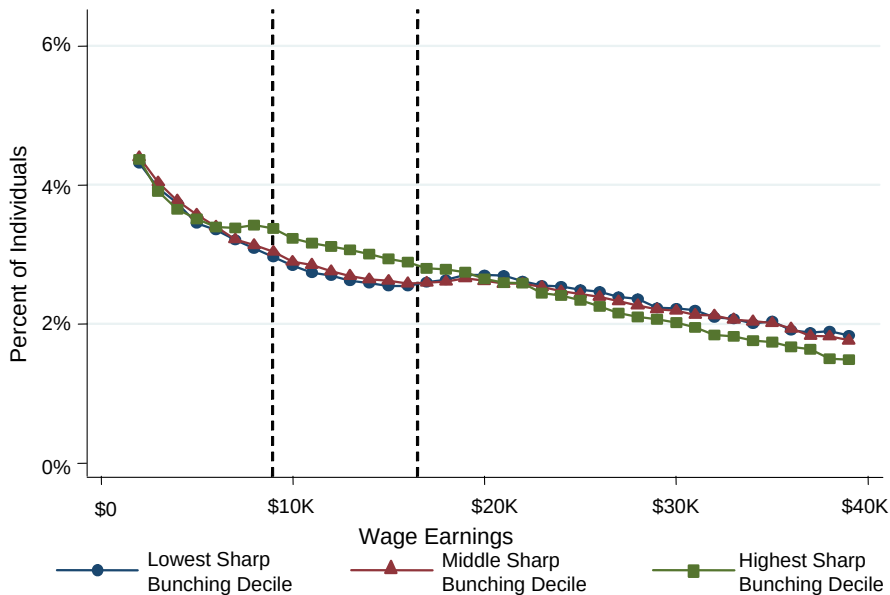
- Step 1: Document variation across neighborhoods in sharp bunching among self-employed
- Step 2: Establish that variation in sharp bunching across neighborhoods is driven by differences in knowledge about EITC schedule
- Step 3: Compare wage earnings distributions across low- and high-knowledge neighborhoods to uncover impacts of EITC on earnings
- Step 4: Compare impacts of changes in EITC subsidies on earnings across low vs. high knowledge nbhds. to account for omitted variables

- Individuals without children are essentially ineligible for the EITC
- Birth of a child therefore generates sharp variation in marginal incentives
- Birth affects labor supply directly, but cross-neighborhood comparisons provide good counterfactuals

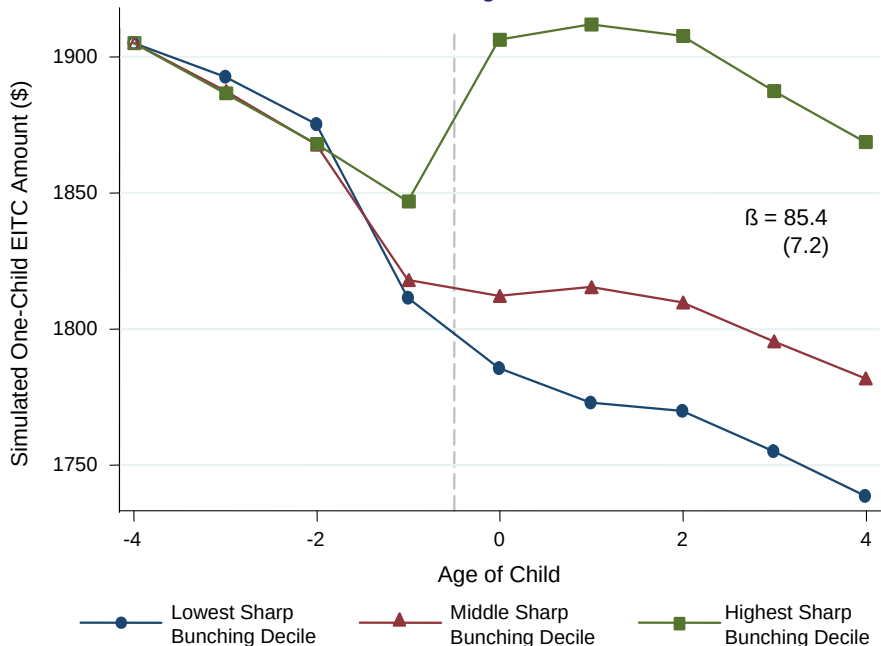
Earnings Distribution in the Year Before First Child Birth for Wage Earners



Earnings Distribution in the Year of First Child Birth for Wage Earners



Simulated EITC Credit Amount for Wage Earners Around First Child Birth



Changes in Simulated EITC around Births for Wage Earners



Changes in Simulated EITC around Births for Wage Earners



Composition of Wage Earnings Responses

- Where is the increase in EITC refunds coming from?
- By studying distributional shifts, we estimate that:
 - 73% comes from increase in earnings in phase-in region
 - 27% from reduction in earnings in phase-out region
- Use neighborhoods with no self-emp. sharp bunching as counterfactual to characterize causal impacts of EITC on earnings distribution

Impact of EITC on Income Distribution

	Percent of EITC-Eligible Households Below Threshold			
	50% of Poverty Line	100% of Poverty Line	150% of Poverty Line	200% of Poverty Line
No EITC Counterfactual	13.7%	31.9%	54.3%	77.3%
EITC, No Behavioral Response	9.4%	22.0%	42.1%	71.1%
EITC, with Avg. Behavioral Response	8.2%	21.0%	42.0%	71.3%
EITC with Top Decile Behavioral Response	6.2%	19.6%	42.0%	71.7%

Elasticity Estimates Based on Change in EITC Refunds Around Birth of First Child

	Mean Elasticity	Phase-in Elasticity	Phase-out Elasticity	Extensive Elasticity
<i>A. Wage Earnings</i>				
Elasticity in U.S. 2000-2005	0.10 (0.008)	0.14 (0.011)	0.06 (0.006)	0.10 (0.009)
Elasticity in top decile ZIP-3's	0.46 (0.017)	0.58 (0.021)	0.30 (0.021)	0.59 (0.033)
<i>B. Total Earnings</i>				
Elasticity in U.S. 2000-2005	0.22 (0.013)	0.34 (0.020)	0.08 (0.004)	0.18 (0.012)
Elasticity in top decile ZIP-3's	0.95 (0.026)	1.32 (0.036)	0.34 (0.012)	1.05 (0.039)

Traditional Labor Supply Elasticity Estimates

- Return to simple model with linear tax
- 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)

- 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)
- But does not decompose estimates into income and substitution effects
- Hard to identify the elasticities relevant to predict labor supply effects of other programs

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

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

NOTE.—Terms are explained in text.

* Denotes mean is more than twice its standard error.

Source: Ashenfelter and Plant 1990

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

Problems with Experimental Design

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

1 Self reported earnings

- Treatments had financial incentives to under-report earnings.
- Reported earnings not well correlated with actual payments
→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
→No longer a random sample of treatment + controls

Tax Reform Variation

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

Table IIa
Marginal Tax Rate

Group	Before TRA86	After TRA86	Change	Relative Change
High	.521 (.002)	.382 (.001)	-.139 (.002)	
75 th Percentile	.365 (.001)	.324 (.001)	-.041 (.001)	-.098 (.002)
90 th 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%}	
75 th Percentile	0.687 (.010) [3799]	0.740 (.010) [3613]	0.053 (.010) {7.2%}	0.037 (.028) {12.3%}
90 th 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%}	
75 th Percentile	1504.1 (14.3) [2610]	1558.9 (13.9) [2676]	54.8 (20.0) {3.6%}	108.6 (65.1) {9.4%}
90 th Percentile	1434.1 (16.4) [2303]	1530.1 (15.9) [2348]	96.0 (22.8) {6.8%}	67.3 (64.8) {6.2%}

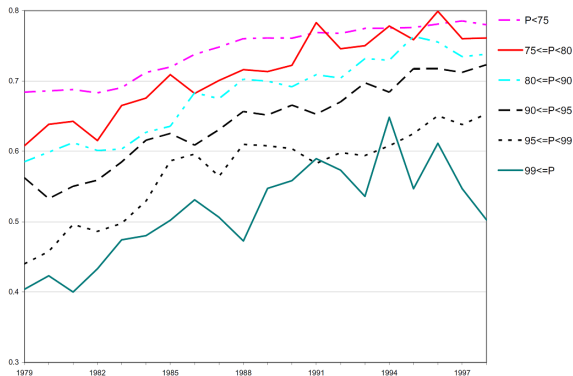
Each cell contains the mean for that group, along with standard errors in (), number of observations in [], and % increase in {}. Means are unweighted.

Source: Eissa 1995

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

- 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 $\leq 75^{\text{th}}$ percentile. Group 75 is $>75^{\text{th}}$ percentile and $\leq 80^{\text{th}}$ percentile. Group 80 is $>80^{\text{th}}$ and $\leq 90^{\text{th}}$. Group 90 is $>90^{\text{th}}$ and $\leq 95^{\text{th}}$. Group 95 is $>95^{\text{th}}$ and $\leq 99^{\text{th}}$. Group 99 is $>99^{\text{th}}$.

Source: Liebman and Saez (2006)

- 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

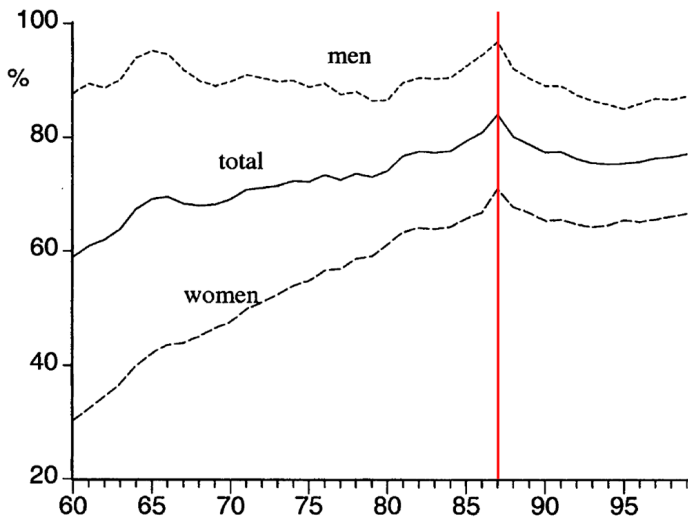


FIGURE 1. THE EMPLOYMENT RATE IN ICELAND,
1960–1996

Source: Bianchi, Gudmundsson, and Zoega 2001

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

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

- Estimates imply earnings elasticity w.r.t. marginal tax rate of roughly 0.37 (Chetty 2012)
 - 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?
- Recent literature has focused on welfare reform in mid-1990's
- Reform modified AFDC cash welfare program to provide more incentives to work by
 - ① Requiring recipients to go to job training
 - 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

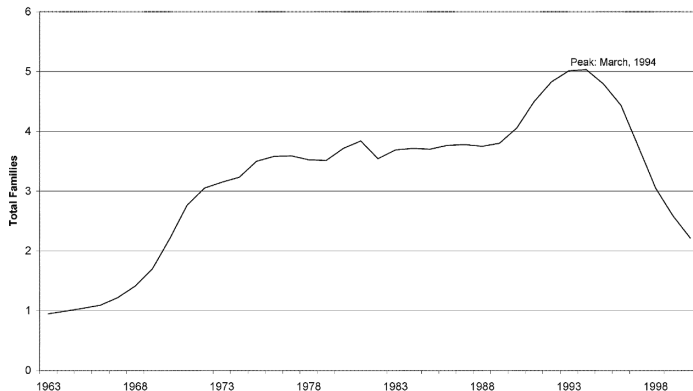


Fig. 1. Average monthly AFDC/TANF caseloads (1963–2000) (in millions).

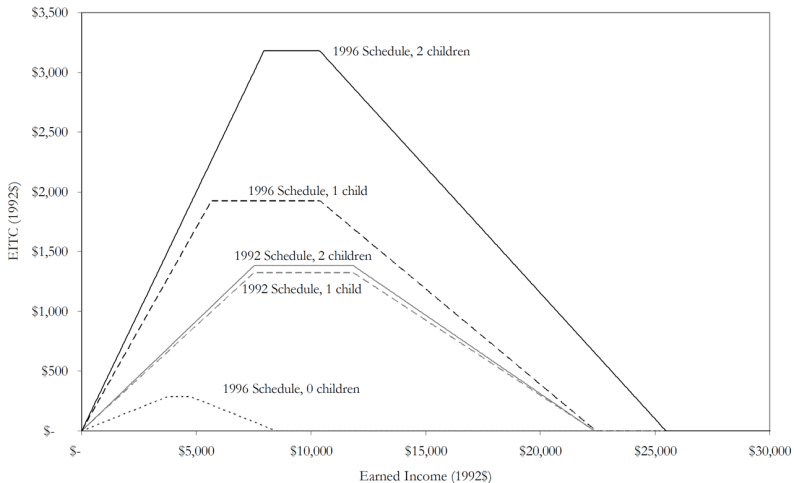
Source: Meyer and Sullivan 2004

Welfare Reform: Two Empirical Questions

- ① Incentives: did welfare reform actually increase labor supply
 - Test whether EITC expansions and changes in welfare policies 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

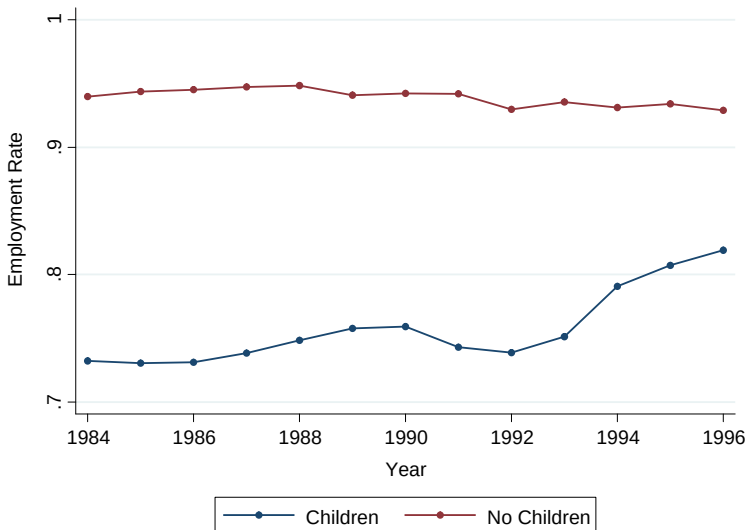
- Study the first question: impact of welfare reforms and EITC on labor supply
- Document dramatic (6 pp, 10%) increase in LFP for single women with children around EITC expansion
- No change for women without children

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



Source: Rothstein 2005

Employment Rates for Single Women with and without Children



Source: Meyer and Rosenbaum 2001

- Problem: EITC 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
- Fit a regression model with following policy variables at state level (interacted with # of kids, etc.):
 - EITC
 - AFDC benefits
 - Medicaid
 - Waivers
 - Training
 - Child Care

- 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%

- 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

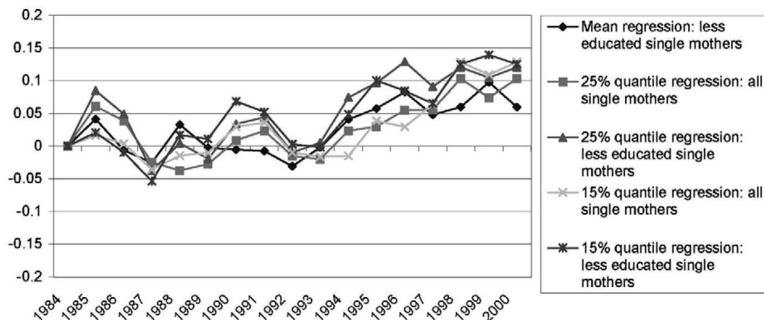


Fig. 2. Total consumption: single mothers, 1984–2000.

Source: Meyer and Sullivan 2004

Relative Consumption: single women with/without children

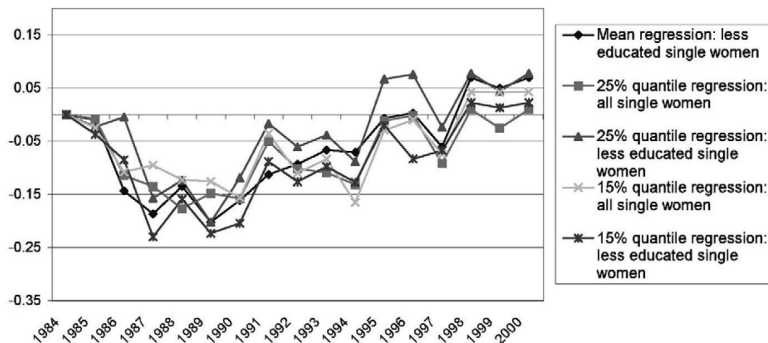


Fig. 3. Relative total consumption: single mothers vs. single women without children, 1984–2000.

Source: Meyer and Sullivan 2004

- 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?

- 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 in *short-run*
 - 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 tractable method (two stage budgeting) to identify preference parameters in a life-cycle model of labor supply
 - See Chetty (2006) for a simple exposition of two-stage budgeting

Life Cycle Model of Labor Supply

- With time-separable utility, agents maximize

$$U = \sum_{t=0}^T \beta^t u(c_t, l_t)$$

- First order conditions

$$\begin{aligned} l_t &: \beta^t u_{l_t} + \lambda w_t / (1+r)^t = 0 \\ c_t &: \beta^t u_{c_t} + \lambda / (1+r)^t = 0 \end{aligned}$$

- 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:

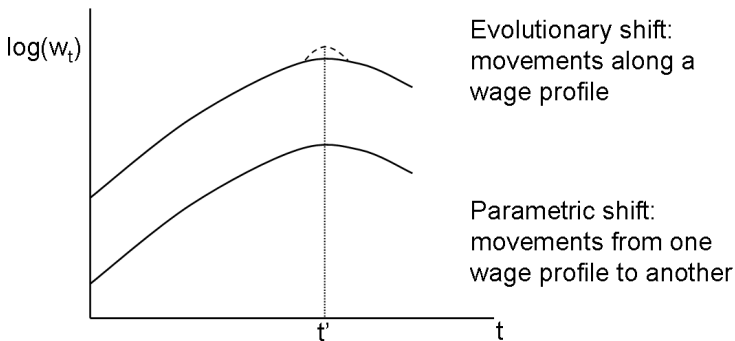
$$\delta = \left(\frac{w_t}{l_t} \right) \frac{\partial l}{\partial w_t} \Big|_{\lambda}$$

- 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:

$$\begin{aligned} l_t &= v'^{-1}(\lambda w_t / [\beta(1+r)]^t) \\ \Rightarrow \frac{\partial l}{\partial w_t} \Big|_{\lambda} &= \frac{\lambda}{\beta(1+r)^t v''(l_t)} > 0 \end{aligned}$$

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

Frisch elasticity $\geq$ Compensated static elasticity

Compensated static elasticity $\geq$ Uncompensated static elasticity

- 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

Frisch elasticity $\geq$ Compensated static elasticity

Compensated static elasticity $\geq$ Uncompensated static elasticity

- Without income effects, all three elasticities are equal
- Otherwise inequalities are strict
- Difference in elasticities related to anticipated vs. unanticipated changes
- Frisch elasticity is of central interest for calibration of macro business cycle models

Frisch Elasticities Implied by Hicksian Elasticity of 0.33

$$\varepsilon^F = \varepsilon + \rho \left(\frac{d[w l_{i,t}^*]}{dY_{i,t}} \right)^2 \frac{A_{i,t}}{w l_{i,t}^*}$$

Income Effect: $-d[w l^*]/dY$

		0.00	0.11	0.22	0.33	0.44	0.55	0.66	0.35
EIS (ρ)	0.00	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.50
	0.20	0.33	0.34	0.35	0.36	0.38	0.41	0.44	0.55
	0.40	0.33	0.34	0.36	0.39	0.43	0.49	0.55	0.60
	0.60	0.33	0.34	0.37	0.42	0.48	0.56	0.66	0.65
	0.80	0.33	0.35	0.38	0.44	0.53	0.64	0.77	0.70
	1.00	0.33	0.35	0.39	0.47	0.58	0.71	0.88	0.75
	1.20	0.33	0.35	0.41	0.50	0.63	0.79	0.99	0.79
	1.40	0.33	0.35	0.42	0.53	0.67	0.87	1.10	0.84

Source: Chetty 2011

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 (1991) 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

- Use quasi-experimental variation coupled with ITLS model to identify Frisch elasticity
- Group individuals by birth cohort (decade) interacted with education (e.g. high school or more)
- Tax reforms in the UK in 1980s affected groups very differently
- Key innovation over pure reduced-form studies: use consumption data as a control for permanent income
- Yields a structurally interpretable (λ constant) elasticity estimate

TABLE IV
ELASTICITIES: GROUPING INSTRUMENTS: COHORT AND EDUCATION

	Wage	Compensated Wage	Other Income	Group Means:		
				Hours	Wage	Income
No Children	0.140 (0.075)	0.140 (0.088)	0.000 (0.041)	32	2.97	88.63
Youngest Child 0–2	0.205 (0.128)	0.301 (0.144)	–0.185 (0.104)	20	3.36	129.69
Youngest Child 3–4	0.371 (0.150)	0.439 (0.159)	–0.173 (0.139)	18	3.10	143.64
Youngest Child 5–10	0.132 (0.117)	0.173 (0.127)	–0.102 (0.109)	21	2.86	151.13
Youngest Child 11 +	0.130 (0.107)	0.160 (0.117)	–0.063 (0.084)	25	2.83	147.31

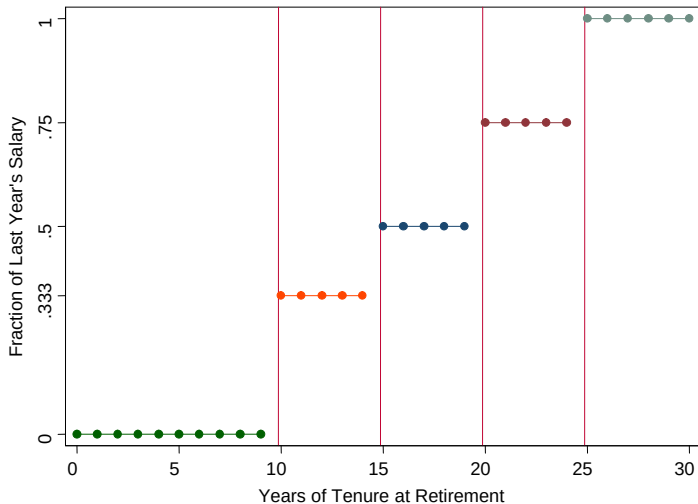
Note: Asymptotic standard errors in parentheses.

Source: Blundell, Duncan, and Meghir 1998

- 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
- Reforms in 80s went in opposite directions at different times
→ Secular trends cannot explain everything

- Use variation in retirement benefits as a function of job tenure in Austria to estimate intertemp subst. elasticity
 - Administrative panel for full population of Austria, 1980-2005
- Question: how much do people delay retirement in order to get higher (anticipated) benefits?
- Rough estimate of intertemp subst. elasticity on extensive margin of 0.2 at annual level

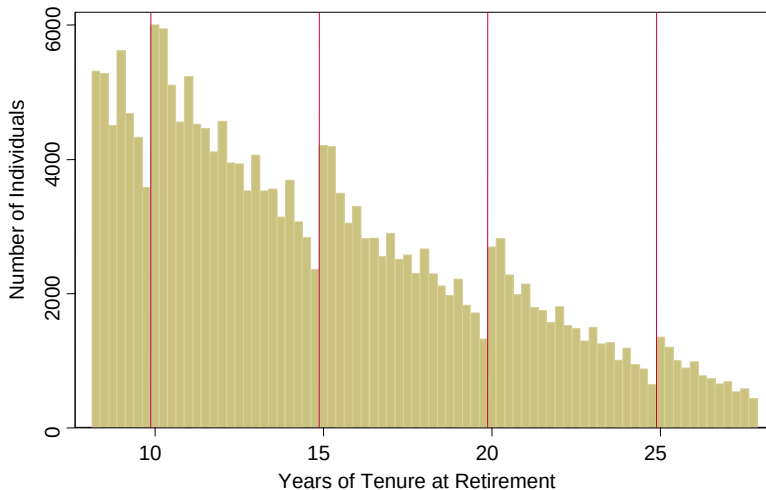
Lump-Sum Severance Payments at Retirement



Source: Manoli and Weber 2009

Distribution of Tenure at Retirement

Quarterly Frequency



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. 2010 (JEL)

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

- 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 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

- 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 $\hat{e} = \frac{e \cdot 20\% - 0}{20\% - 10\%} = 2e$

- 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

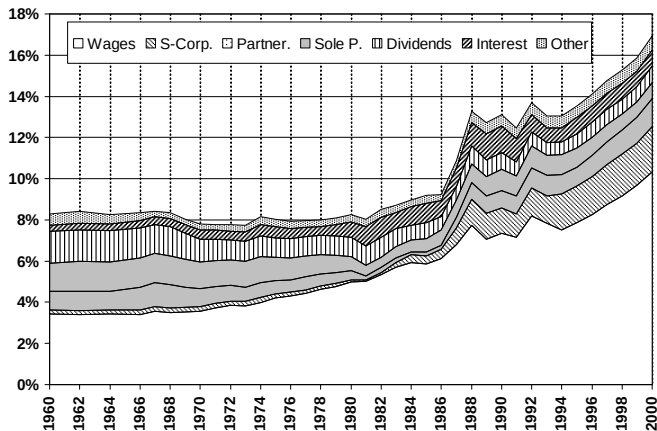


FIGURE 7

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 elasticity
- 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)	
$\ln(1 - \text{tax}_t)$	1.159 (.119)	.394 (.139)	.810 (.178)	2.218 (.281)	.290 (.311)	1.289 (.128)	.150 (.073)
$\ln(1 - \text{tax}_{t+1})$	-.763 (.106)	-.051 (.132)	-.433 (.158)	-1.663 (.240)	-.181 (.279)	-.853 (.115)	-.060 (.065)
$\ln(1 - \text{tax}_t) \times [I > 0]$	.282 (.140)		.851 (.639)	.140 (.189)	.943 (.344)	.175 (.153)	.187 (.094)
$\ln(\text{market value})$	.610 (.014)	.337 (.015)	.559 (.021)	.999 (.033)	.518 (.041)	.619 (.015)	.289 (.008)
Earnings/assets	.510 (.056)	.311 (.059)	.681 (.089)	.823 (.144)	.344 (.129)	.542 (.062)	.423 (.035)
Time	.077 (.008)	.068 (.009)	.073 (.012)	.061 (.019)	.060 (.020)	.079 (.008)	.082 (.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

- 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

- 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

$$MTR_{t+3}^P = f_{t+3}(y_t, X_t)$$

$$MTR_{t+3} = f_{t+3}(y_{t+3}, X_{t+3})$$

- Step 2 [first stage]: Regress $\log(1 - MTR_{t+3}) - \log(1 - MTR_t)$ on $\log(1 - MTR_{t+3}^P) - \log(1 - MTR_t)$
 - 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 (0.120)	-0.462 (0.194)	0.170 (0.106)	0.611 (0.144)	0.120 (0.106)	0.400 (0.144)
Dummy for marrieds	-0.008 (0.010)	-0.062 (0.018)	0.045 (0.014)	0.049 (0.023)	0.050 (0.012)	0.055 (0.021)
Dummy for singles	-0.037 (0.012)	-0.053 (0.019)	-0.034 (0.013)	-0.032 (0.022)	-0.036 (0.013)	-0.027 (0.021)
Log(income) control			-0.083 (0.015)	-0.167 (0.021)		

Source: Gruber and Saez 2002

- Find an elasticity of roughly 0.1-0.4 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

Observed Earnings Responses to Small Tax Reforms



Source: Chetty et al. 2009

- 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 $d[wl]/dy = -0.1$

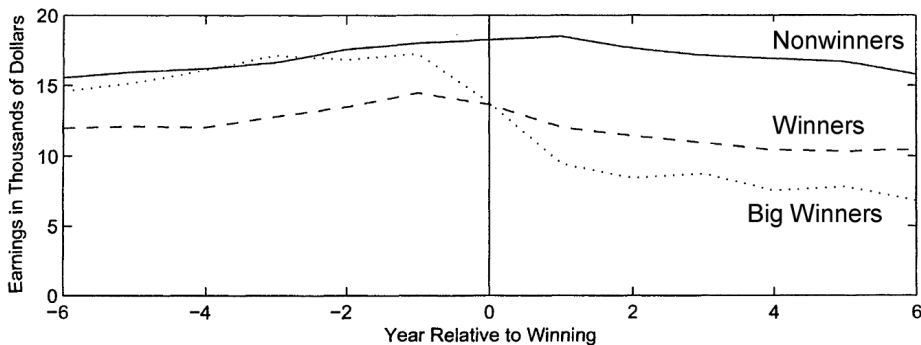


FIGURE 1. AVERAGE EARNINGS FOR NONWINNERS, WINNERS, AND BIG WINNERS

Source: Imbens, Rubin, and Sacerdote 2001

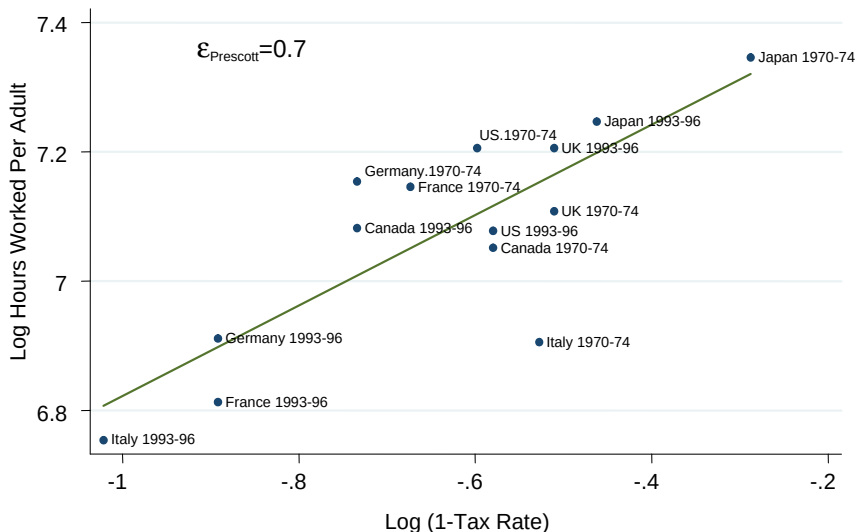
Taxable Income Literature: Summary

- Large responses for the rich, mostly intertemporal substitution and shifting
- Responses among lower incomes small in short run
- Pattern confirmed in many settings (e.g. Kopczuk 2009 - Polish flat tax reform)
- But many methodological problems remain to be resolved
 - Econometric issues: mean reversion, appropriate counterfactuals
 - Which elasticity is being identified?

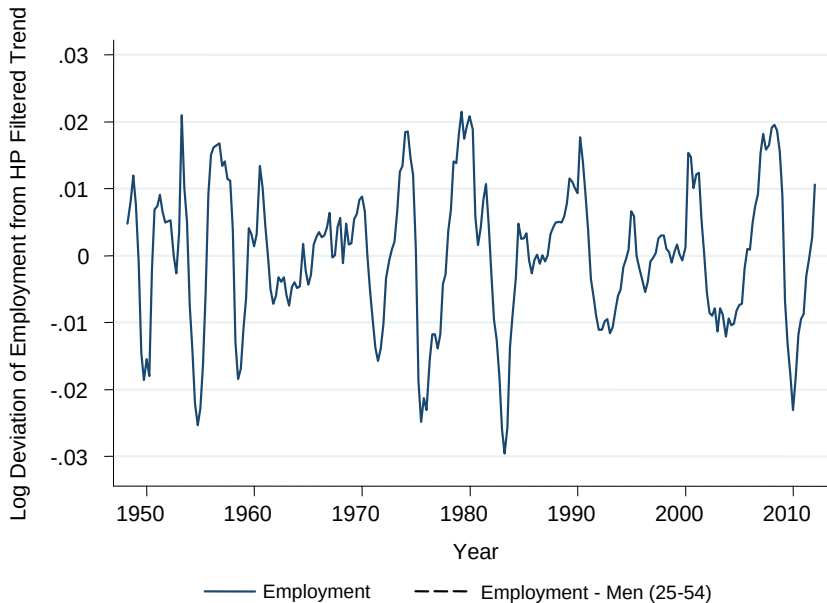
- Macroeconomists estimate/calibrate elasticities by examining long-term trends/cross-country comparisons
- Identification more tenuous but estimates perhaps more relevant to long-run policy questions of interest
- Use aggregate hours data and aggregate measures of taxes (average tax rates)
- But highly influential in calibration of macroeconomic models
 - Macro models require high elasticities to fit both business cycle and cross-country data

- Uses data on hours worked by country in 1970 and 1995 for 7 OECD countries
- Technique to identify elasticity: calibration
- Rough logic: posit a utility function $u(c, l; \varepsilon)$
- Hicksian elasticity of $\varepsilon = 0.8$ best matches cross-country evidence
- Note that this is analogous to a regression without controls for other variables abstracting from GE effects

Aggregate Hours vs. Net-of-Tax Rates Across Countries (Prescott Data)



Business Cycle Fluctuations in Employment Rates in the U.S.



Reconciling Micro and Macro Estimates

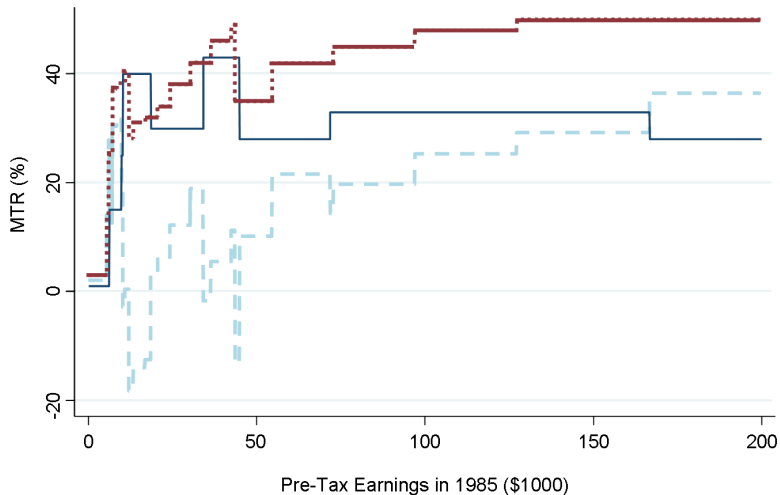
- Recent interest in reconciling micro and macro elasticity estimates
- Three potential explanations
 - 1 Statistical Bias: regulations, culture differs in countries with higher tax rates [Alesina, Glaeser, Sacerdote 2005]
 - 2 Extensive vs. Intensive margin: “Indivisible Labor” [Rogerson 1988; Rogerson 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 2012]

- Chetty (2012) 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)
- To illustrate potential importance of frictions, first calculate utility loss from **ignoring** tax changes under neoclassical model with $\varepsilon = 0.5$

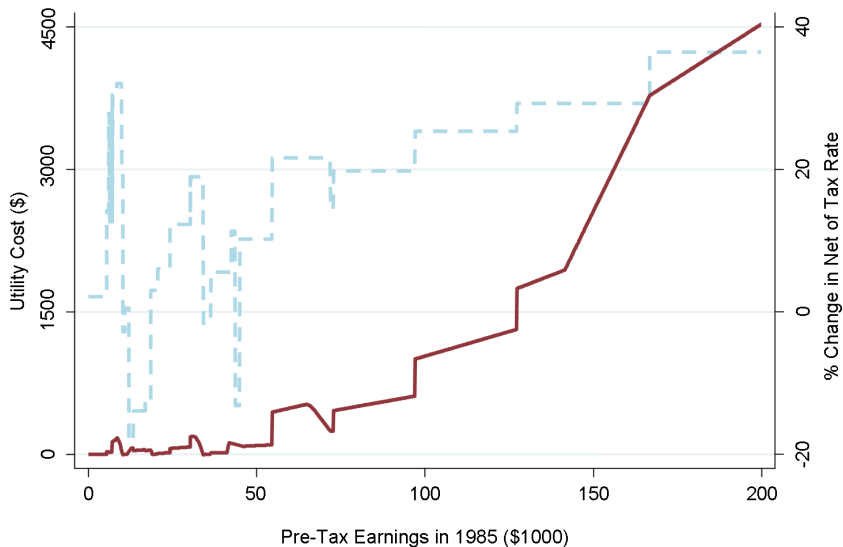
Tax Reform Act of 1986: Change in Marginal Tax Rates



Source: Chetty 2011

— MTR in 1985 — MTR in 1988 - - - $\Delta \log(1-MTR)$

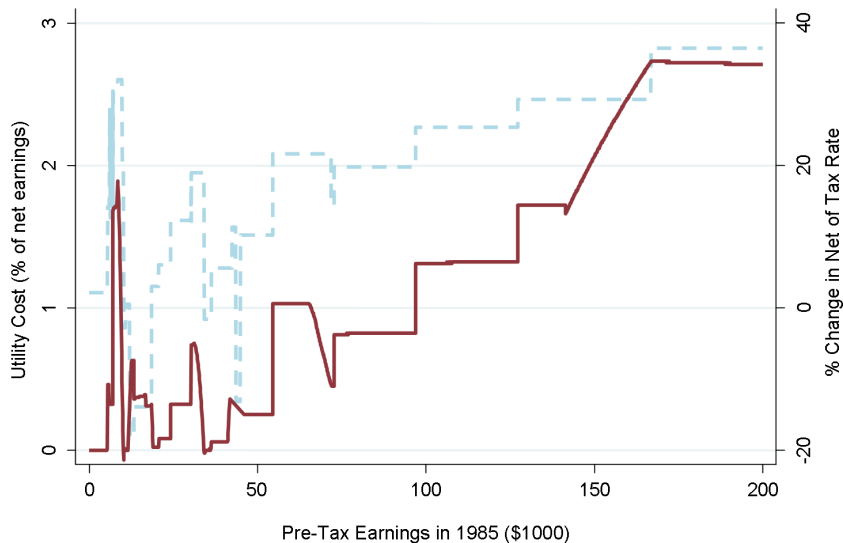
Utility Cost of Ignoring TRA86 (\$)



Source: Chetty 2011

— Utility Cost (\$) - - - $\Delta \log(1-MTR)$

Utility Cost of Ignoring TRA86 (% of net earnings)



Source: Chetty 2011

— Utility Cost (\$) - - - $\Delta \log(1-MTR)$

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 :

$$\begin{aligned}x_i^*(p) &= \left(\frac{a_i}{p}\right)^\varepsilon \\ \Rightarrow \log x_i^*(p) &= \alpha - \varepsilon \log p + v_i\end{aligned}$$

where $v_i = \alpha_i - \alpha$ denotes i 's deviation from mean demand

- Under orthogonality condition $\mathbb{E} v_i | p = 0$,

$$\varepsilon = \frac{\mathbb{E} \log x_1^* - \mathbb{E} \log x_0^*}{\log p_1 - \log p_0}$$

→ 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

$$\hat{\varepsilon} = \frac{\mathbb{E} \log x_1 - \mathbb{E} \log x_0}{\log p_1 - \log p_0}$$

- Observed elasticity confounds **structural** elasticity ε with adjustment cost distribution:

$$\hat{\varepsilon} = P(\Delta u_i > k_i) \varepsilon$$

- Behavioral example: price misperception $\tilde{p}(p)$

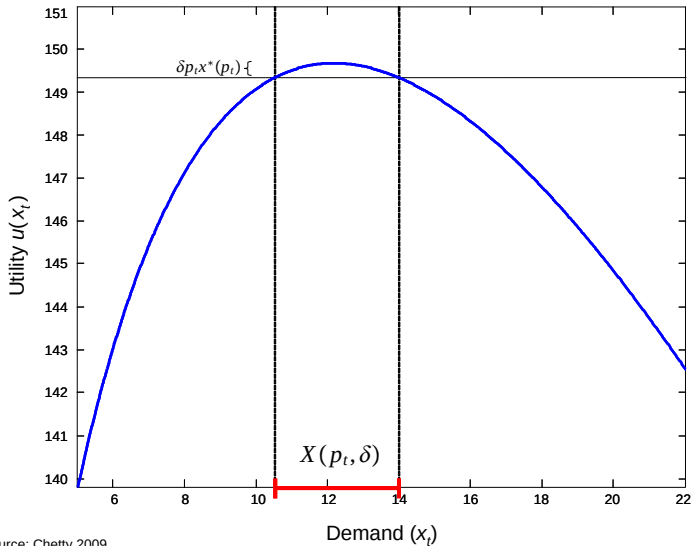
$$\hat{\varepsilon} = \varepsilon \frac{\mathbb{E} \log \tilde{p}(p_1) - \mathbb{E} \log \tilde{p}(p_0)}{\log p_1 - \log p_0}$$

- Restrict size of frictions by requiring that utility loss is less than exogenous threshold δ :

$$U(x_i^*) - U(x_i) < \delta p x_i^*$$

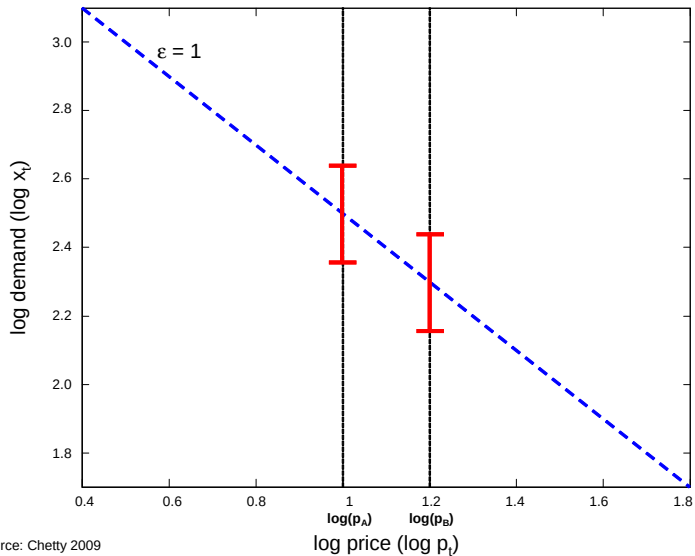
- This restriction generates a class of models around nominal model
 - Includes adjustment cost models, inattention, etc.
- A δ class of models maps price to a choice **set** $X(p, \delta)$ instead of a single point $x^*(p)$

Construction of Choice Set



Source: Chetty 2009

Identification with Optimization Frictions



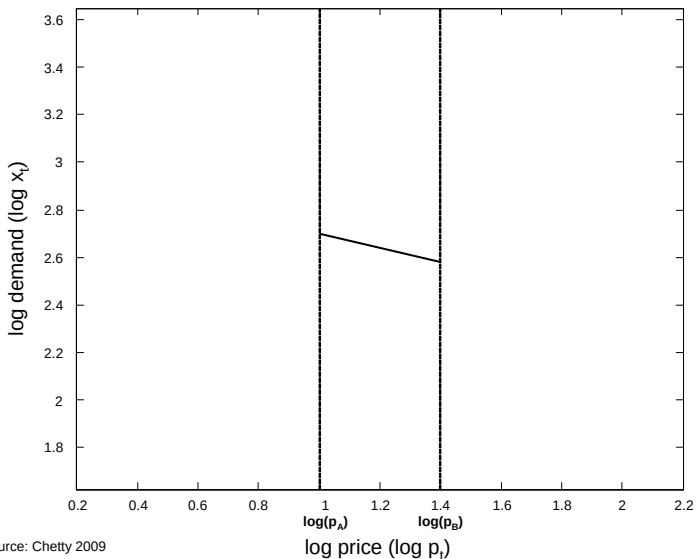
Source: Chetty 2009

Bounds and Partial Identification

- Chetty (2012) derives bounds on ε given estimates of $\hat{\varepsilon}$
- Approach is closely related to modern econometrics literature on partial identification
 - Also called “set identification” or “moment inequalities” in IO
- Pioneered by Manski (1993)
- See Tamer (2010 Annual Review) for a good summary

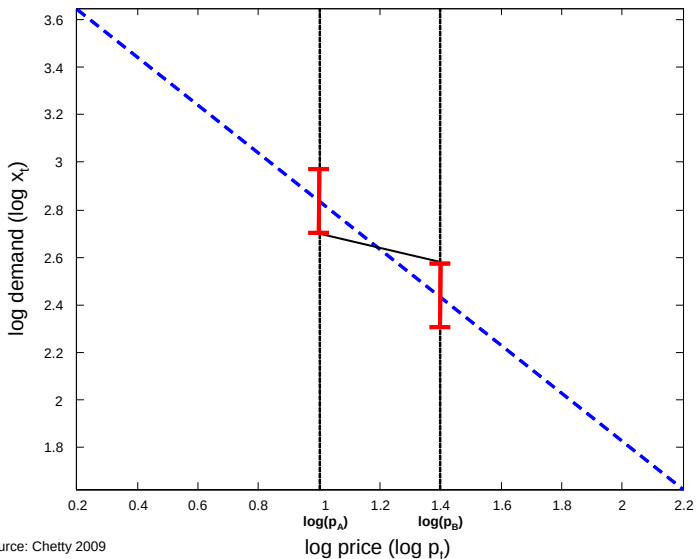
- Main concept: avoid making strong assumptions about data generating process
- Instead, derive non-parametric bounds on parameters under worst-case scenarios
- Classic example: missing data with a binary outcome
 - Traditionally, assume data is missing at random
 - But easy to derive bounds by assuming missing data is all either 0 or 1
- Chetty (2012) applies similar logic to handle model uncertainty rather than missing data

Calculation of Bounds on Structural Elasticity



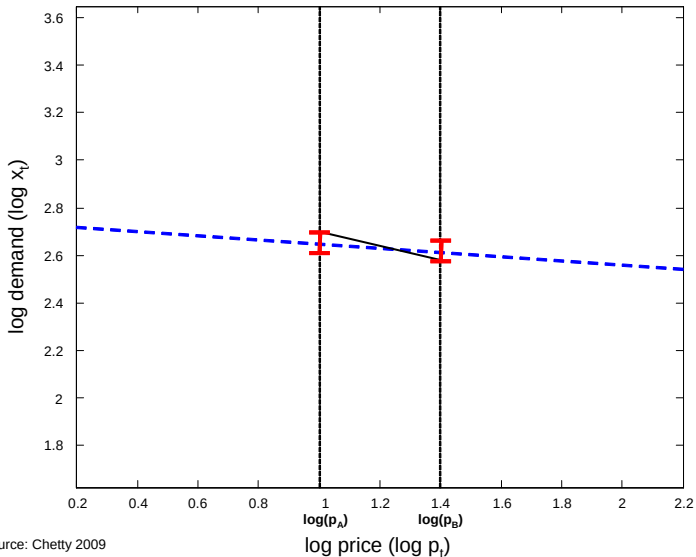
Source: Chetty 2009

a) Upper Bound on Structural Elasticity



Source: Chetty 2009

b) Lower Bound on Structural Elasticity



Source: Chetty 2009

Bounds on Elasticity with Optimization Frictions

- For small δ , the range of structural elasticities consistent with an observed elasticity $\hat{\varepsilon}$ in a δ class of models is approximately

$$\left[\hat{\varepsilon} + \frac{4\delta}{(\Delta \log p)^2} (1 - \rho), \hat{\varepsilon} + \frac{4\delta}{(\Delta \log p)^2} (1 + \rho) \right]$$

$$\text{where } \rho = \left(1 + \frac{1}{2} \frac{\hat{\varepsilon}}{\delta} (\Delta \log p)^2 \right)^{1/2}$$

- 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 $(\Delta \log p)^2$

Extensive Margin Responses

- Now consider bounds on extensive margin elasticities

- Assume that $x \in \{0, 1\}$ and flow utility is

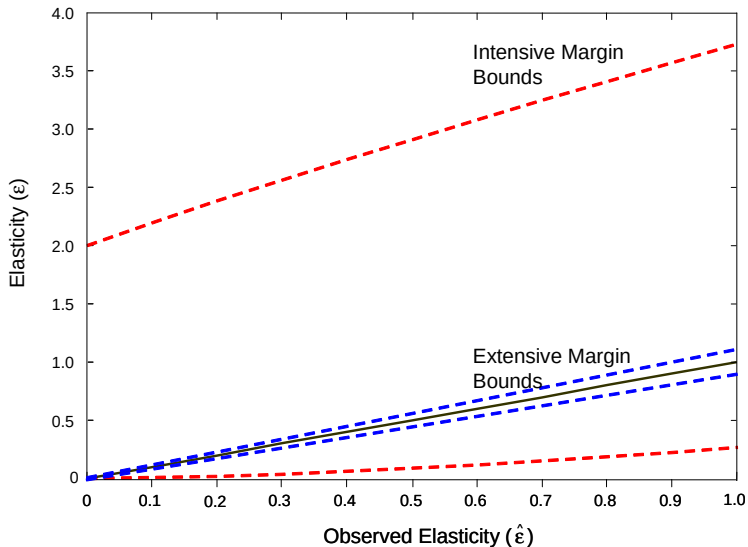
$$u_i(x, y) = y + b_i x$$

- Let $F(b_i)$ denote distribution of tastes for x
- Agents optimally buy x if taste $b_i > p \rightarrow \theta^* = 1 - F(p)$
- Let structural extensive elasticity be denoted by

$$\eta = \frac{\log \theta_A^*(p_A) - \log \theta_B^*(p_B)}{\log p_A - \log p_B}$$

- Let θ = observed participation rate and $\hat{\eta}$ = observed extensive elasticity

Bounds on Structural Elasticities: $\delta = 1\%$, $\Delta \log p = 20\%$



Source: Chetty 2011

Extensive vs. Intensive Margin Bounds

- Bounds on extensive margin elasticities shrink linearly with δ rather than in proportion to $\sqrt{\delta}$
- Intuition: agents are not near optima to begin with on extensive margin $\rightarrow$ first-order utility losses from failing to reoptimize
 - Marginal agent loses benefit of price cut if he doesn't enter market

- What can be learned about structural elasticity from existing estimates?
- Collect estimates from a broad range of studies that estimate intensive margin Hicksian elasticities
- Calculate bounds on the intensive margin structural elasticity with frictions of $\delta = 1\%$ of net earnings

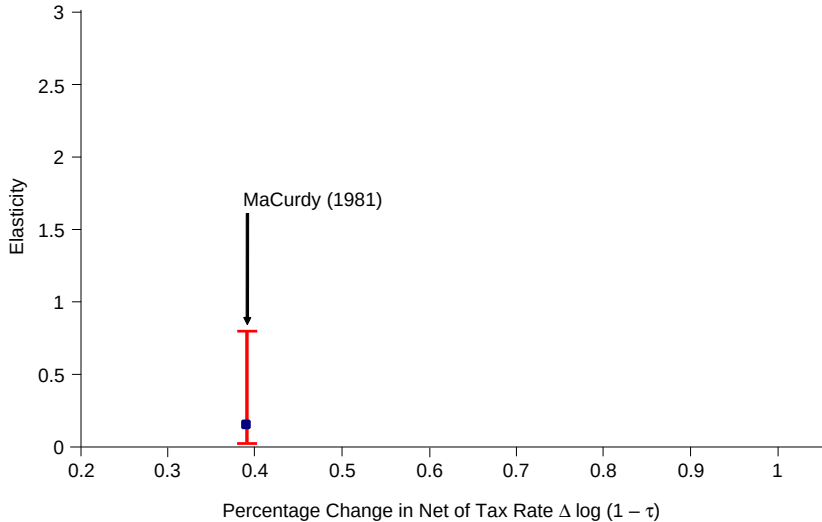
Bounds on Intensive-Margin Hicksian Elasticities with $\delta = 1\%$ Frictions

Study (1)	Identification (2)	$\hat{\varepsilon}$ (3)	$se(\hat{\varepsilon})$ (4)	$\Delta \log(1-\tau)$ (5)	ε_L (6)	ε_U (7)
<i>A. Hours Elasticities</i>						
1. MaCurdy (1981)	Lifecycle wage variation, 1967-1976	0.15	0.15	0.39	0.03	0.80
2. Eissa and Hoynes (1998)	U.S. EITC, 1984-1996, Men	0.20	0.07	0.07	0.00	15.29
3. Eissa and Hoynes (1998)	U.S. EITC, 1984-1996, Women	0.09	0.07	0.07	0.00	15.07
4. Blundell et al. (1998)	U.K. Tax Reforms, 1978-1992	0.14	0.09	0.23	0.01	1.78
5. Ziliak and Kniesner (1999)	Lifecycle wage, tax variation 1978-1987	0.15	0.07	0.39	0.03	0.80
	Mean observed elasticity	0.15				
<i>B. Taxable Income Elasticities</i>						
6. Bianchi et al. (2001)	Iceland 1987 Zero Tax Year	0.37	0.05	0.49	0.15	0.92
7. Gruber and Saez (2002)	U.S. Tax Reforms 1979-1991	0.14	0.14	0.14	0.00	4.42
8. Saez (2004)	U.S. Tax Reforms 1960-2000	0.09	0.04	0.15	0.00	3.51
9. Jacob and Ludwig (2008)	Chicago Housing Voucher Lottery	0.12	0.03	0.36	0.02	0.84
10. Gelber (2010)	Sweden, 1991 Tax Reform, Women	0.49	0.02	0.71	0.28	0.86
11. Gelber (2010)	Sweden, 1991 Tax Reform, Men	0.25	0.02	0.71	0.12	0.54
12. Saez (2010)	U.S., 1st EITC Kink, 1995-2004	0.00	0.02	0.34	0.00	0.70
13. Chetty et al. (2011a)	Denmark, Top Kinks, 1994-2001	0.02	0.00	0.30	0.00	0.93
14. Chetty et al. (2011a)	Denmark, Middle Kinks, 1994-2001	0.00	0.00	0.11	0.00	6.62
15. Chetty et al. (2011a)	Denmark Tax Reforms, 1994-2001	0.00	0.00	0.09	0.00	9.88
	Mean observed elasticity	0.15				

Bounds on Intensive-Margin Hicksian Elasticities with $\delta = 1\%$ Frictions

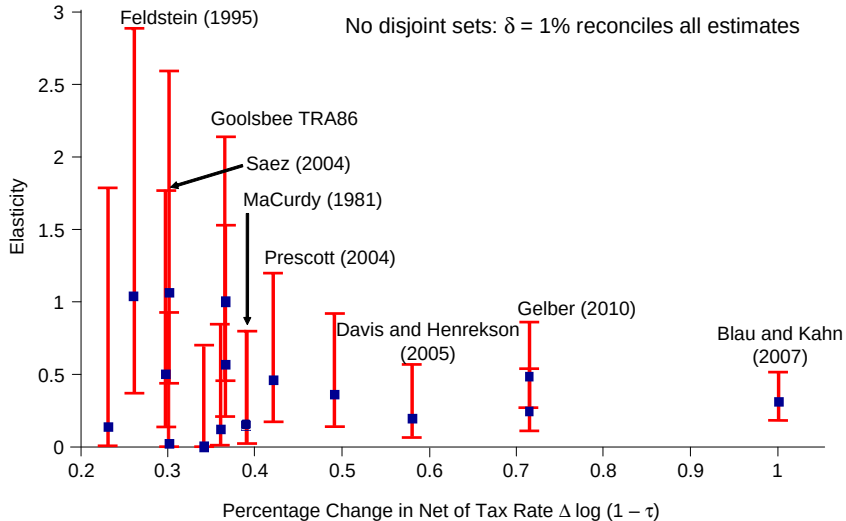
Study (1)	Identification (2)	$\hat{\varepsilon}$ (3)	$se(\hat{\varepsilon})$ (4)	$\Delta \log(1-\tau)$ (5)	ε_L (6)	ε_U (7)
<i>C. Top Income Elasticities</i>						
16. Feldstein (1995)	U.S. Tax Reform Act of 1986	1.04		0.26	0.37	2.89
17. Auten and Carroll (1999)	U.S. Tax Reform Act of 1986	0.57	0.12	0.37	0.21	1.53
18. Goolsbee (1999)	U.S. Tax Reform Act of 1986	1.00	0.15	0.37	0.47	2.14
19. Saez (2004)	U.S. Tax Reforms 1960-2000	0.50	0.18	0.30	0.14	1.77
20. Kopczuk (2010)	Poland, 2002 Tax Reform	1.07	0.22	0.30	0.44	2.58
	Mean observed elasticity	0.84				
<i>D. Macro/Cross-Sectional</i>						
21. Prescott (2004)	Cross-country Tax Variation, 1970-96	0.46	0.09	0.42	0.18	1.20
22. Davis and Henrekson (2005)	Cross-country Tax Variation, 1995	0.20	0.08	0.58	0.07	0.57
23. Blau and Kahn (2007)	U.S. wage variation, 1980-2000	0.31	0.004	1.00	0.19	0.51
	Mean observed elasticity	0.32				

Bounds on Intensive-Margin Hicksian Elasticities with $\delta=1\%$



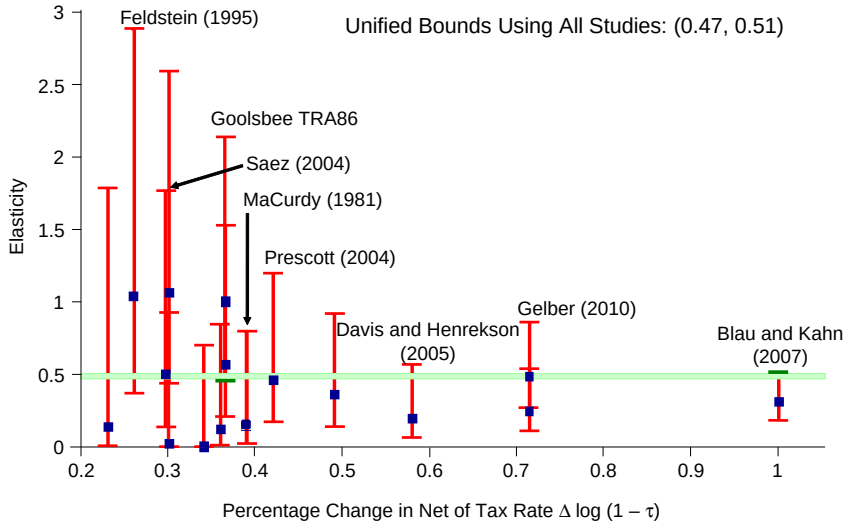
Source: Chetty 2011

Bounds on Intensive-Margin Hicksian Elasticities with $\delta=1\%$



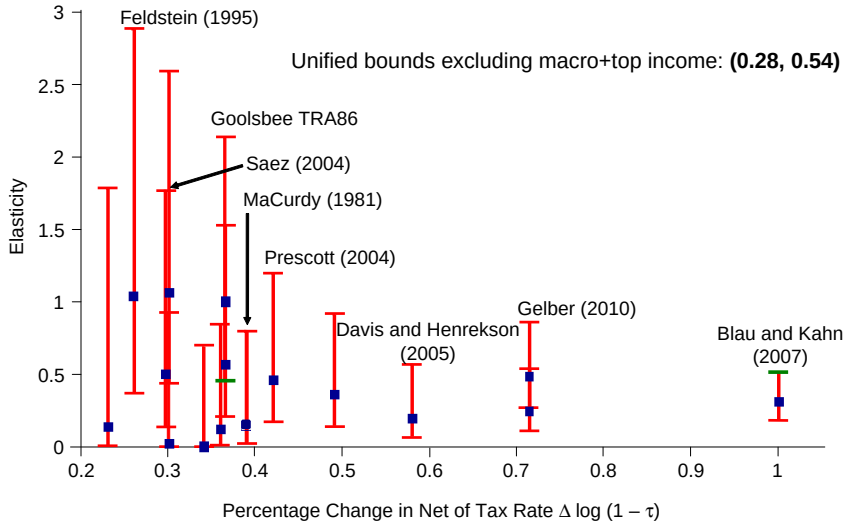
Source: Chetty 2011

Bounds on Intensive-Margin Hicksian Elasticities with $\delta=1\%$



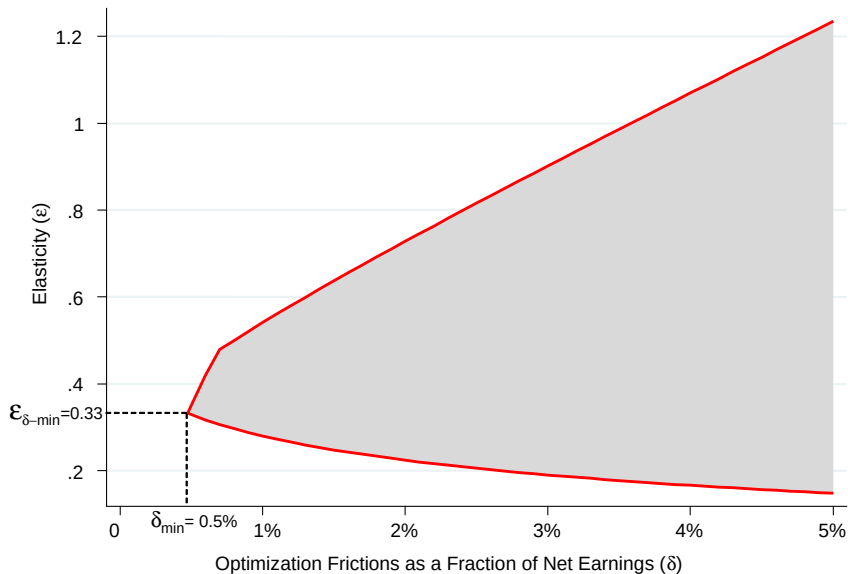
Source: Chetty 2011

Bounds on Intensive-Margin Hicksian Elasticities with $\delta=1\%$



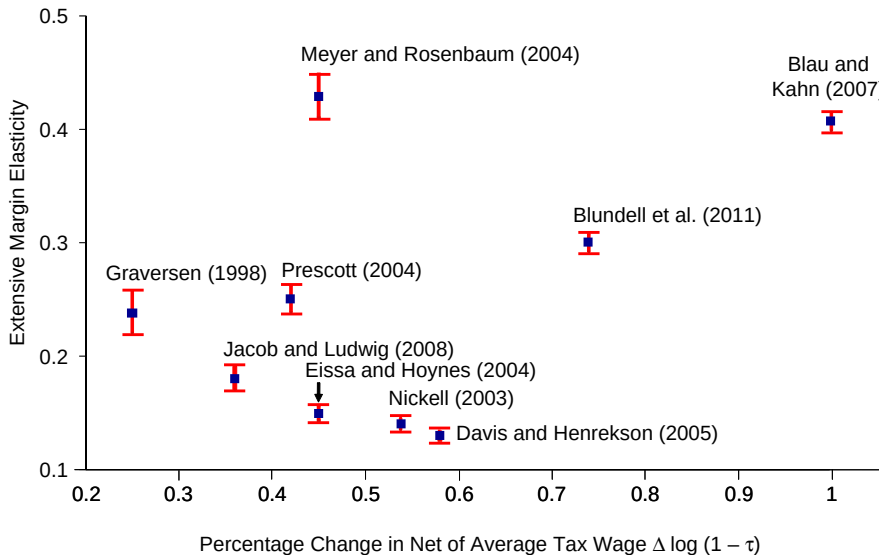
Source: Chetty 2011

Unified Bounds on Intensive Margin Elasticity vs. Degree of Frictions



Source: Chetty 2011

Bounds on Extensive-Margin Hicksian Elasticities with $\delta=1\%$ Frictions



Source: Chetty 2011

Micro vs. Macro Labor Supply Elasticities

- Chetty, Guren, Manoli, and Weber (2012): can frictions explain gap between micro and macro elasticities?
- Collect estimates of intensive and extensive margin elasticities adjusted for frictions and evaluate macro predictions

Aggregate Hours vs. Net-of-Tax Rates Across Countries (Prescott Data)

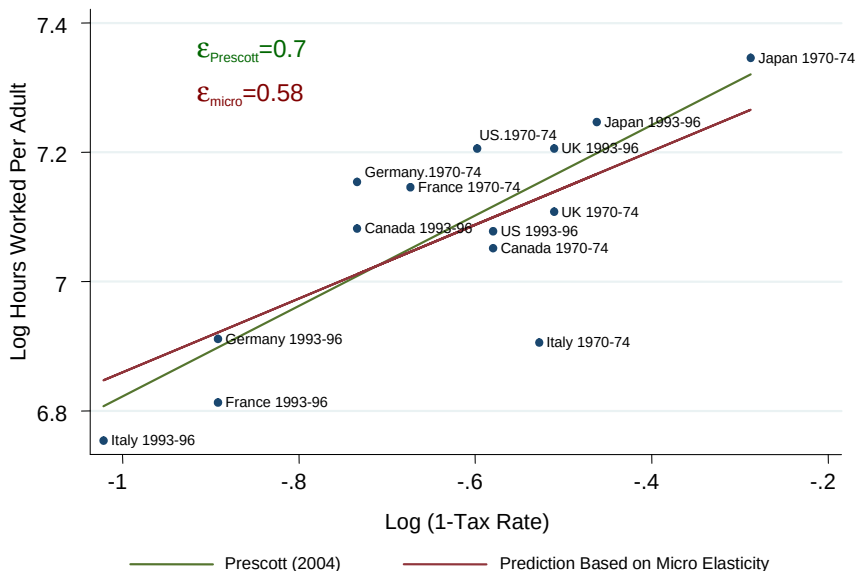


Table 2: Micro vs. Macro Labor Supply Elasticities

		Extensive Margin	Intensive Margin	Aggregate Hours
Steady State (Hicksian)	micro	0.25	0.33	0.58
	macro	0.17	0.33	0.50
Intertemporal Substitution (Frisch)	micro			
	macro			

→ Indivisible labor + frictions reconcile micro and macro steady-state elasticities

Business Cycle Fluctuations in Employment Rates in the U.S.

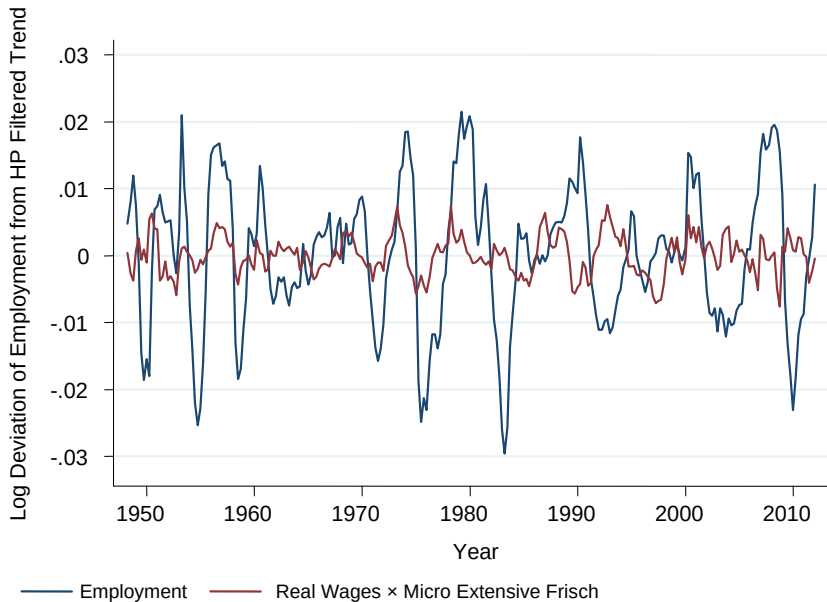


Table 2: Micro vs. Macro Labor Supply Elasticities

		Extensive Margin	Intensive Margin	Aggregate Hours
Steady State (Hicksian)	micro	0.25	0.33	0.58
	macro	0.17	0.33	0.50
Intertemporal Substitution (Frisch)	micro	0.32	0.54	0.86
	macro	[2.77]	[0.54]	3.31

→ Even with indivisible labor, Frisch elasticity of aggregate hours >1 is inconsistent with micro evidence

→ Challenge: matching employment flucs. with extensive Frisch of 0.3

- Search/labor wedge models provide one solution

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)

- 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 $\varepsilon_{l,w} \ll 0$
 - 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

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

$$wu_c(y + wl, l) = -u_l(y + wl, l)$$

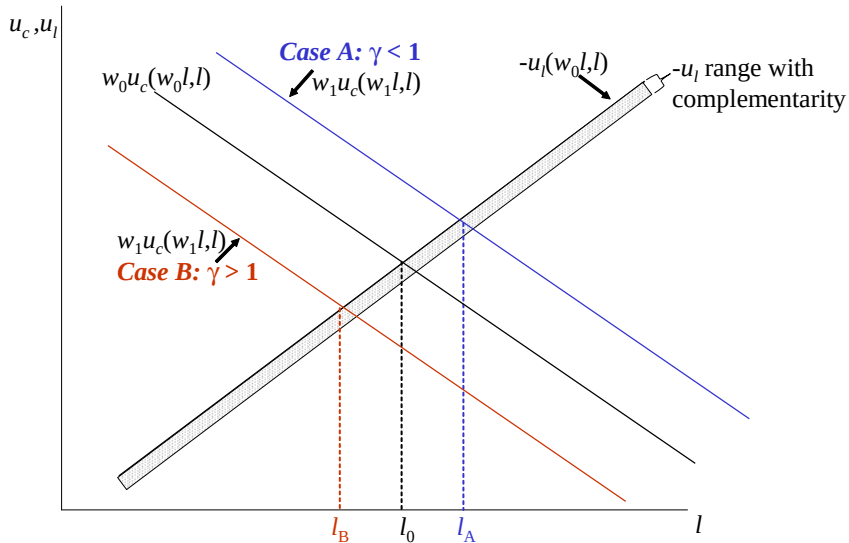
- Work until point where marginal utility of an additional dollar is offset by marginal disutility of work required to earn that dollar
- Comparative statics of this condition implies (if $u_{cl} = 0$):

$$\gamma = -\left(1 + \frac{wl}{y}\right) \frac{\varepsilon_{l,y}}{\varepsilon_{lc,w}}$$

- Risk aversion directly related to ratio of income effect to substitution effect

Labor Supply and Risk Aversion: 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, $\gamma < 1 \iff \varepsilon_{l,w} > 0$
- 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)
<i>A. Hours</i>						
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
<i>B. Participation</i>						
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
<i>C. Earned Income</i>						
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
<i>D. Macroeconomic/Trend Evidence</i>						
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

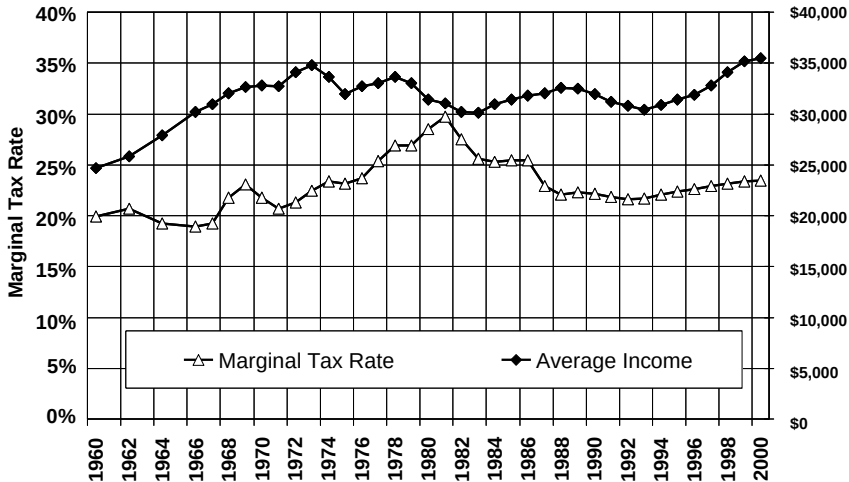
- Labor supply evidence justifies use of $u(c) = \log c$
- Formula $\gamma = -\left(1 + \frac{wl}{y}\right) \frac{\varepsilon_{l,y}}{\varepsilon_{lc,w}}$ useful in tax, insurance, and other applications

- We have covered evidence on two of the three elements critical for optimal income taxation
 - ① Labor supply elasticities
 - ② Measurement of preferences/social welfare weights
 - ③ Measurement of income distribution
- Third piece can be well measured using tax data, even for high incomes (Piketty and Saez 2004)

Saez 2004: Long-Run Evidence

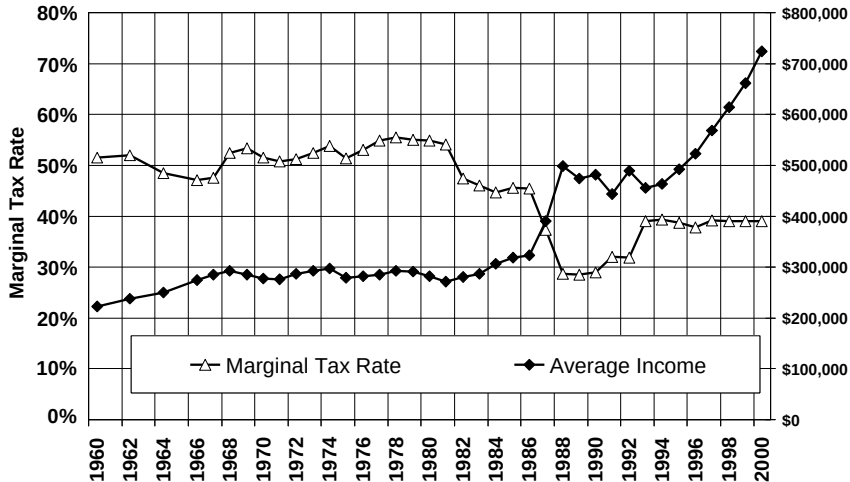
- Compares top 1% relative to the bottom 99%
- Bottom 99% real income increases up to early 1970s and stagnates since then
- Top 1% increases slowly up to the early 1980s and then increases dramatically up to year 2000.
 - Corresponds to the decrease in MTRs
- Pattern exemplifies general theme of this literature: large responses for top earners, no response for rest of the population

Bottom 99% Tax Units



Source: Saez 2004

Top 1% Tax Units



Source: Saez 2004

Public Economics Lectures

Part 6: Social Insurance

Raj Chetty and Gregory A. Bruich

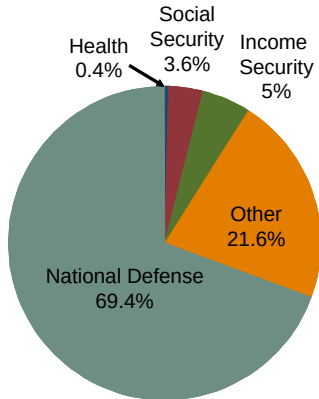
Harvard University
Fall 2012

- ① Motivations for Social Insurance
- ② Unemployment Insurance
- ③ Workers' Compensation
- ④ Disability Insurance
- ⑤ Health Insurance

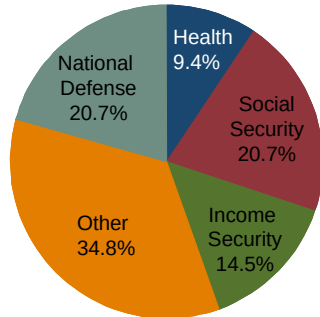
Definition of Social Insurance

- 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.



1953



2008

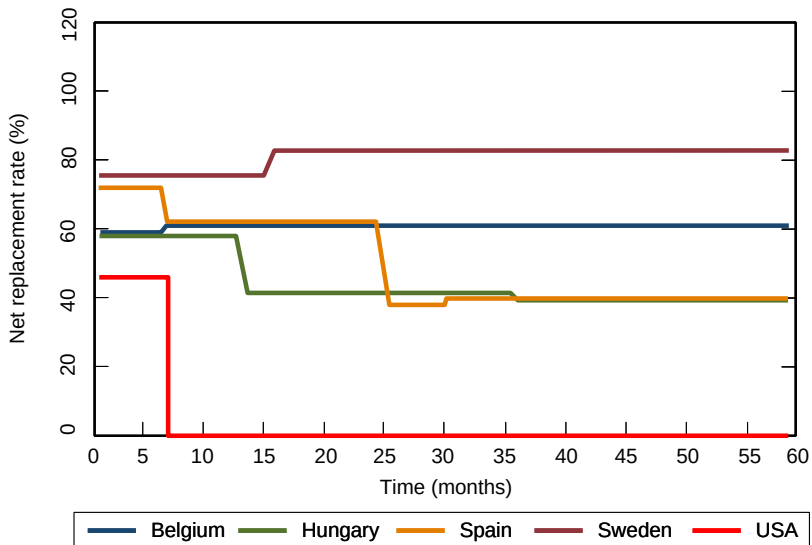
Source: Office of Management and Budget, historical tables, government outlays by function

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

Unemployment Benefit Systems in Developed Countries



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

- ① Institutional details: see handout posted on course website
- ② Expected utility theory: See MWG or other graduate texts
- ③ Survival analysis: Kiefer (1988 JEL)
- ④ Surveys: Krueger and Meyer (2002) Handbook chapter
- ⑤ Chetty and Finkelstein (2012) Handbook chapter

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
- More recent “sufficient statistic” version that can be connected to data: Einav, Finkelstein, and Cullen (QJE 2010)
- 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.

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

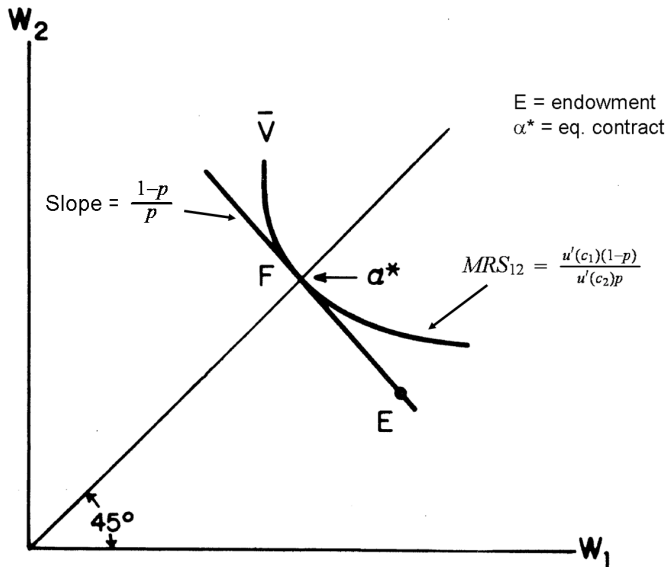
$$\max_{\alpha_1} (1-p_i)u(w-\alpha_1) + p_i u(w + \frac{1-p_i}{p_i}\alpha_1).$$

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

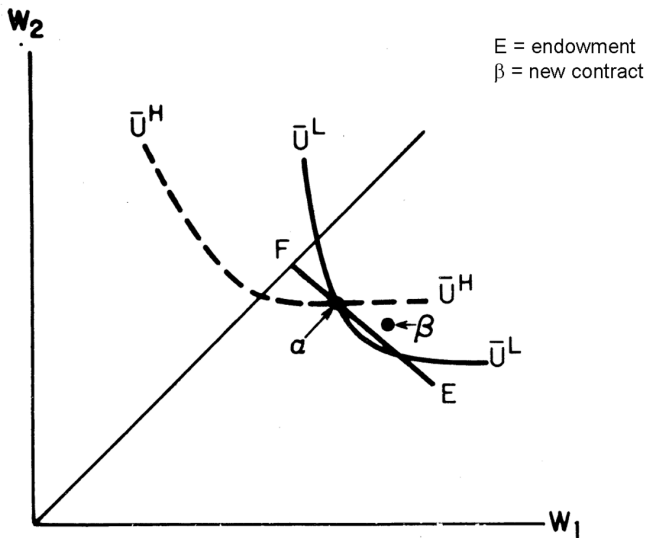
Rothschild-Stiglitz: Second Best Problem

- 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 $\alpha_2 = \frac{1-\bar{p}}{\bar{p}}\alpha_1$ but $\bar{p} > p_L$.
 - 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

$$\left(\frac{9}{10} \frac{3}{4} + \frac{1}{10} \frac{1}{4}\right)100 = \frac{7}{10}100 = 70.$$

- 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
- Standard test: “positive correlation” property in equilibrium (Chiappori and Salanie 2000)
 - 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.

- 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		ESTIMATES FROM LINEAR PROBABILITY MODEL OF PROBABILITY OF DYING WITHIN FIVE YEARS	
	Compulsory Market (1)	Voluntary Market (2)	Compulsory Market (3)	Voluntary Market (4)
Index-linked	-.839*** (.217)	-.894** (.358)	-.053*** (.019)	-.185*** (.050)
Escalating	-1.085*** (.113)	-1.497*** (.253)	-.072*** (.010)	-.152*** (.030)
Guaranteed	.019 (.029)	.216*** (.060)	.007* (.004)	.046*** (.016)
Capital-protected	...	.056 (.051)	...	.064*** (.016)
Payment (£100s)	-.003*** (.0006)	.001** (.0004)	-.0003*** (.0001)	.0003*** (.0001)
Male Annuitant	.640*** (.039)	.252*** (.051)	.044*** (.005)	.044*** (.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

Limitations of Positive Correlation Test

- ❶ Does not account for other dimensions of heterogeneity that may confound the correlation
 - Literature on “advantageous selection” (e.g., Finkelstein and McGarry 2006)
- ❷ Correlation does not clearly map into parameters that control welfare costs of selection
 - Einav, Finkelstein, and Cullen (2010) develop “cost curve” tests that map to measures of welfare costs
- ❸ Only applicable in markets that exist, i.e. those that have not totally unravelled
 - Hendren (2012) uses subjective expectations data to bound welfare costs in markets that have unraveled

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

- 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 moral hazard
 - Individuals will not work if they have perfect unemp insurance
 - Must take this distortion into account to find optimal level of social insurance

- **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 on the job

- ① Optimal level of UI benefits ignoring firm responses [Bailey-Chetty model]
 - Theory applies to all 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

- 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

$$eu(A + w_h - t(b)) + (1 - e)u(A + w_l + b) - \psi(e)$$

First Best Problem

- In first best, there is no moral hazard problem
- To solve for FB, suppose government chooses b and e jointly to maximize agent's welfare:

$$\begin{aligned} \max_{b,e} & e(A + w_h - t) + (1 - e)u(A + w_l + b) - \psi(e) \\ \text{s.t. } & t = \frac{1 - e}{e}b \end{aligned}$$

- Solution to this problem is $u'(c_e) = u'(c_u) \Rightarrow$ full insurance

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 $w_h - w_l$
 - Private marginal product is $w_h - w_l - b - t$
 - Agents therefore search too little from a social perspective, leading to efficiency losses

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:

$$\max_{b,t} V(b, t)$$

$$\text{s.t. } e(b)t = (1 - e(b))b$$

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.
- ② **Sufficient Statistic:** derive formulas for b^* as a fn. of reduced-form elasticities
 - Baily-Chetty formula is one example

Chetty (2006) Sufficient Statistic Formula

- 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:

$$\frac{dV(b)}{db} = (1 - e)u'(c_l) - \frac{dt}{db}eu'(c_h)$$

- Can ignore $\frac{\partial e}{\partial b}$ terms because of agent optimization

- 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

- Suppose moving home raises consumption by H and has a cost $g(H)$:

$$\begin{aligned} V(b) = & \max_{e, H} eu(A + w_h - t(b)) \\ & + (1 - e)[u(A + w_l + b + H) - g(H)] - \psi(e) \end{aligned}$$

- 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)

Chetty (2006) Sufficient Statistic Formula

- The government's UI budget constraint implies

$$\begin{aligned}\frac{dt}{db} &= \frac{1-e}{e} - \frac{b}{e^2} \frac{de}{db} = \frac{1-e}{e} \left(1 + \frac{\varepsilon_{1-e,b}}{e}\right) \\ \implies \frac{dV(b)}{db} &= (1-e) \left\{ u'(c_l) - \left(1 + \frac{\varepsilon_{1-e,b}}{e}\right) u'(c_h) \right\}\end{aligned}$$

- 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 Formula

$$\frac{u'(c_l) - u'(c_h)}{u'(c_h)} = \frac{\varepsilon_{1-e,b}}{e}$$

- 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

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

$$\begin{aligned} \frac{u'(c_l) - u'(c_h)}{u'(c_h)} &\approx -\frac{u''}{u'} c_h \frac{\Delta c}{c} \\ &= \gamma \frac{\Delta c}{c} \end{aligned} \tag{1}$$

- Gap in marginal utilities is a function of curvature of utility (risk aversion) and consumption drop from high to low states

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

$$\frac{\Delta c}{c} = \frac{c_h - c_l}{c_h} = \text{consumption drop during unemployment}$$

$$\gamma = -\frac{u''(c_h)}{u'(c_h)} c_h = \text{coefficient of relative risk aversion}$$

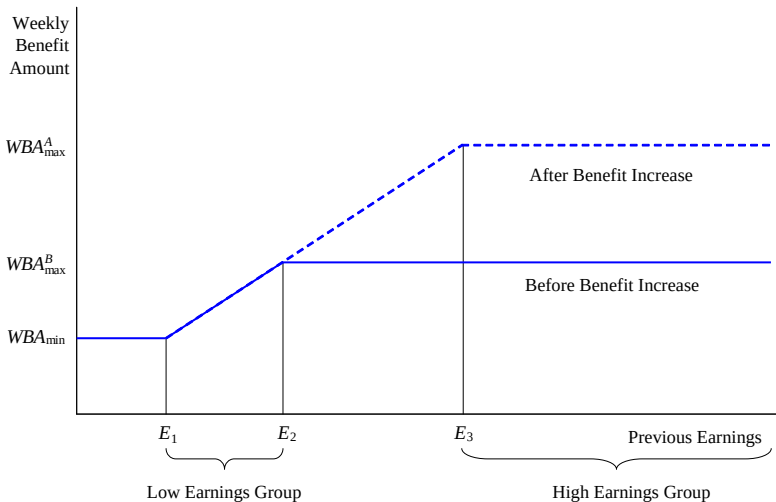
$$\varepsilon_{1-e,b} = \frac{d \log 1 - e}{d \log b} = \text{elast. of probability of unemp. w.r.t. benefits}$$

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 hold
- Empirical work on UI provides estimates of the three key parameters $(\gamma, \frac{\Delta c}{c}, \varepsilon)$.

- 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

- 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 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 (0.0169)	−.0422 (0.0171)	−.0416 (0.0168)	−.0386 (0.0239)	−.0386 (0.0242)
1 = married, spouse present	.1302 (0.0508)	.1221 (0.0515)	.1315 (0.0507)	.1006 (0.0722)	.1001 (0.0730)
1 = white	.2097 (0.0572)	.2230 (0.0579)	.2171 (0.0568)	.2337 (0.0834)	.2364 (0.0841)
Years of schooling	−.0276 (0.0083)	−.0275 (0.0084)	−.0272 (0.0083)	−.0177 (0.0123)	−.0176 (0.0124)
Log UI benefit level	−.8782 (0.1091)	−.8157 (0.1096)	−.8478 (0.1088)	−.8685 (0.2042)	−.8757 (0.2065)
Log pre-UI after tax wage	.5630 (0.0855)	.5651 (0.0860)	.5530 (0.0848)	.7289 (0.1415)	.7411 (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 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
- Suggests that ins. markets are not perfect and UI does play a consumption smoothing role, but estimates are imprecise
 - Key area for future work: admin. consumption data

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:

$$\frac{b^*}{w} = \frac{\varepsilon_{1-e,b}/e}{\beta_2} \left(\frac{1}{\gamma} \right) - \frac{\beta_1}{\beta_2}$$

- Plugging in $\varepsilon_{1-e,b} = .43$ as in Gruber (1997) and $e = .95$ (5% unemployment rate) yields:

$$\frac{b^*}{w} = - \left(\frac{.43/.95}{.28} \cdot \frac{1}{\gamma} \right) - \frac{(-.24)}{.28}$$

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 $\gamma < 2$
- Implies optimal benefit level is much lower than observed

Measurement of Risk Preferences

- Parameter that is most poorly identified: γ
- Barseghyan et al. (2012) and Einav et al. (2012) estimate risk preferences across different domains
 - Do individuals who choose more insurance for health also choose more insurance for long-term disability?
 - Find positive correlation in risk preferences but substantial heterogeneity across domains
- Suggests that appropriate value of γ is highly context-dependent
 - Could be due to behavioral factors and framing but also due to neoclassical factors

Chetty and Szeidl (2007): Consumption Commitments

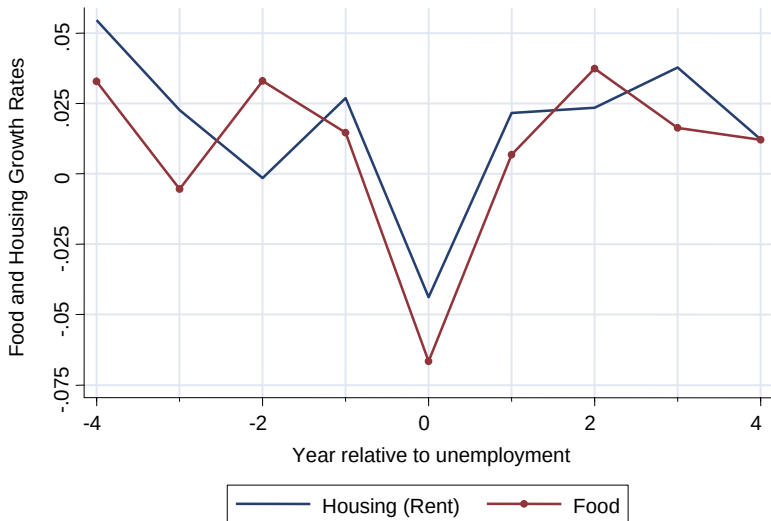
- Standard expected utility model: one composite consumption good c
- Composite commodity assumes that people can cut back on all consumption goods at all times freely.
- E.g. when unemployed, cut consumption of food, housing, cars, furniture, etc.
- In practice, difficult to adjust many elements of consumption in short run because of fixed adjustment costs

Homeowners' Consumption around Unemployment Shocks



Source: Chetty and Szeidl 2007

Renters' Consumption around Unemployment Shocks



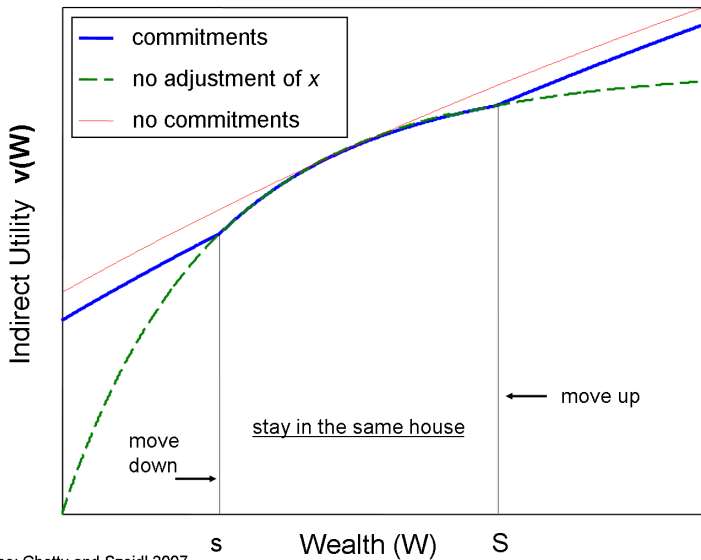
Source: Chetty and Szeidl 2007

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

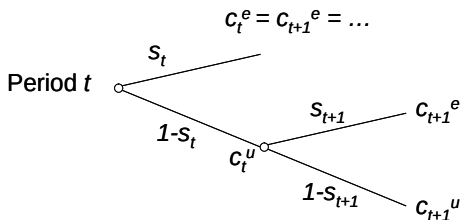
- ① Assets prior to job loss exogenous
 - ② No heterogeneity
 - ③ Fixed wages: choose only search intensity, not reservation wage
- Each of these is relaxed in paper, so model nests search models used in structural literature (e.g. Wolpin 1987)

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: $\psi(s_t)$



- 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

$$J_t(A_t) = \max_{s_t} s_t V_t(A_t) + (1 - s_t) U_t(A_t) - \psi(s_t)$$

- 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

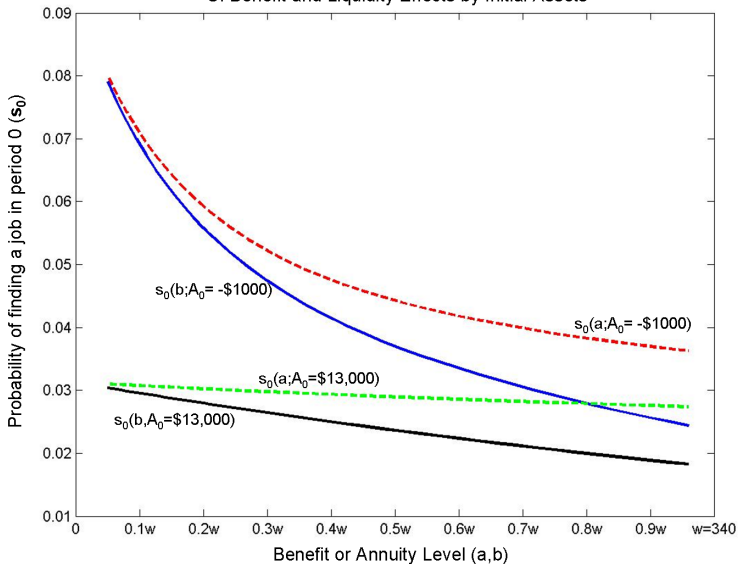
- Benefit effect can be decomposed into two terms:

$$\begin{aligned}\partial s_t / \partial A_t &= \{u'(c_t^e) - u'(c_t^u)\} / \psi''(s_t) < 0 \\ \partial s_t / \partial w_t &= u'(c_t^e) / \psi''(s_t) > 0 \\ \Rightarrow \partial s_t / \partial b_t &= \partial s_t / \partial A_t - \partial s_t / \partial w_t\end{aligned}$$

- $\partial s_t / \partial A_t$ 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

Figure 1

UI Benefit and Liquidity Effects by Initial Assets



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

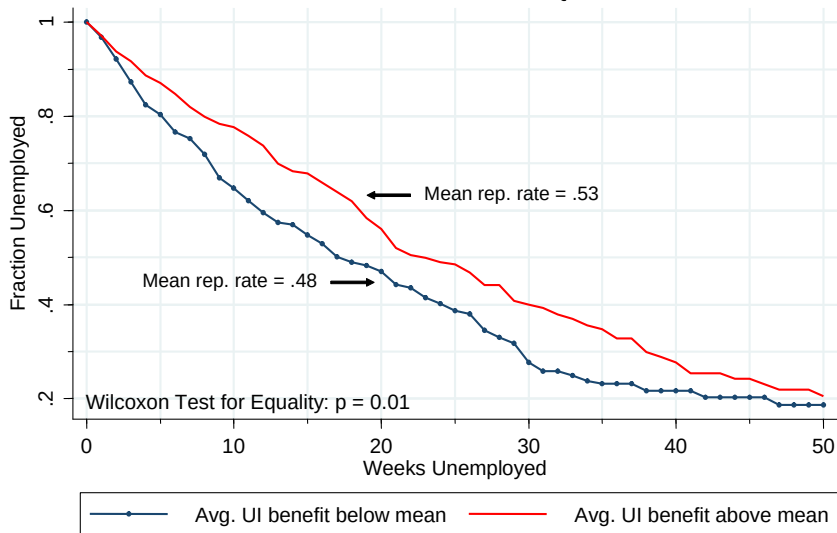
	Net Liquid Wealth Quartile			
	1 (< -\$1,115)	2 (-\$1,115-\$128)	3 (\$128-\$13,430)	4 (>\$13,430)
Median Liq. Wealth	\$466	\$0	\$4,273	\$53,009
Median Debt	\$5,659	\$0	\$353	\$835
Median Home Equity	\$2,510	\$0	\$11,584	\$48,900
Median Annual Wage	\$17,188	\$14,374	\$18,573	\$23,866
Mean Years of Education	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

Effect of UI Benefits on Durations: Lowest Quartile of Net Wealth



Source: Chetty 2008

Figure 3b

Effect of UI Benefits on Durations: Second Quartile of Net Wealth

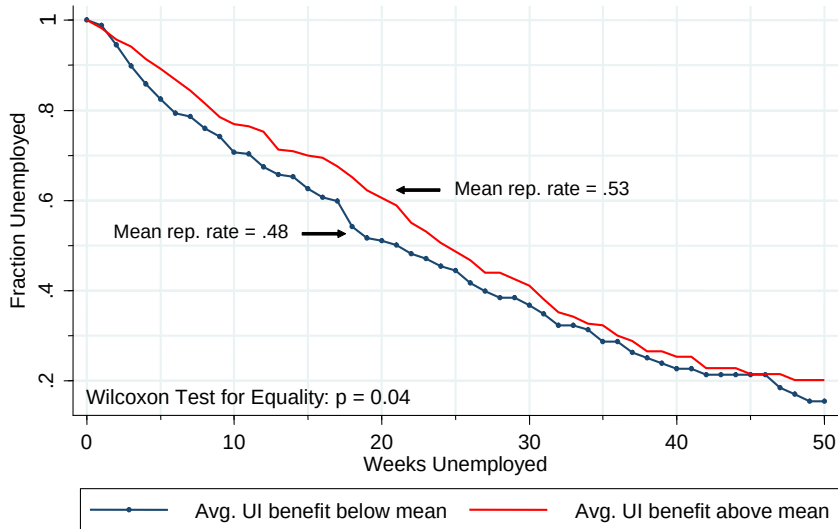
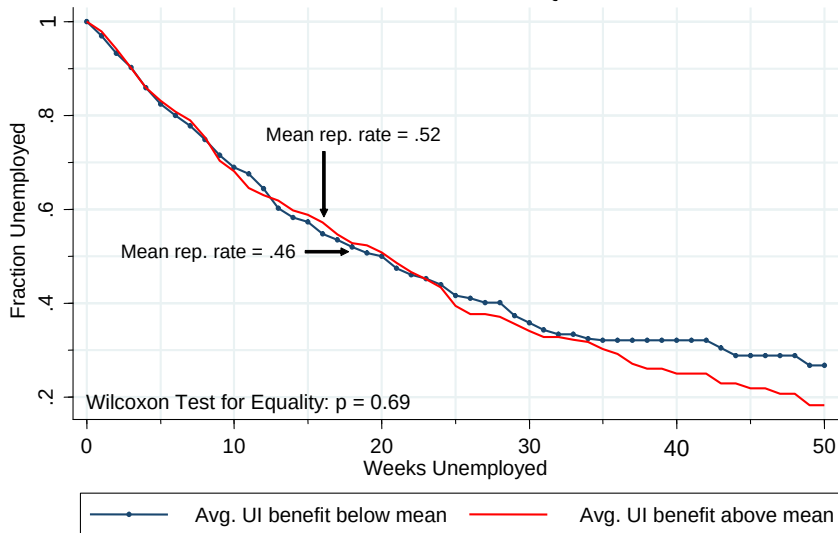


Figure 3c

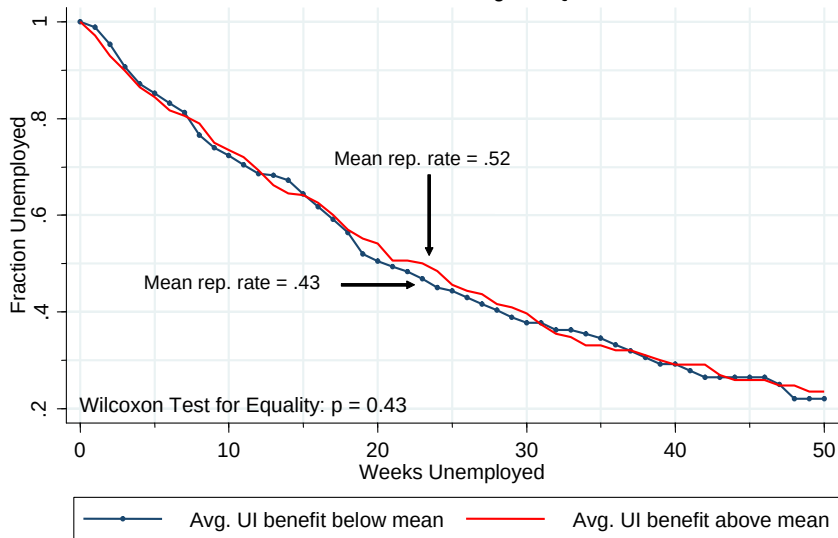
Effect of UI Benefits on Durations: Third Quartile of Net Wealth



Source: Chetty 2008

Figure 3d

Effect of UI Benefits on Durations: Highest Quartile of Net Wealth



Source: Chetty 2008

TABLE 2
Effect of UI Benefits: Cox Hazard Model Estimates

	(1)	(2)	(3)	(4)	(5)
	Pooled Full cntrls	Stratified No cntrls	<u>Stratified with Full Controls</u>		
		Avg WBA	Max WBA	Ind. WBA	
log UI ben	-0.527 (0.267)				
Q1 x log UI ben		-0.721 (0.304)	-0.978 (0.398)	-0.727 (0.302)	-0.642 (0.241)
Q2 x log UI ben		-0.699 (0.484)	-0.725 (0.420)	-0.388 (0.303)	-0.765 (0.219)
Q3 x log UI ben		-0.368 (0.309)	-0.476 (0.358)	-0.091 (0.370)	-0.561 (0.156)
Q4 x log UI ben		0.234 (0.369)	0.103 (0.470)	0.304 (0.339)	0.016 (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

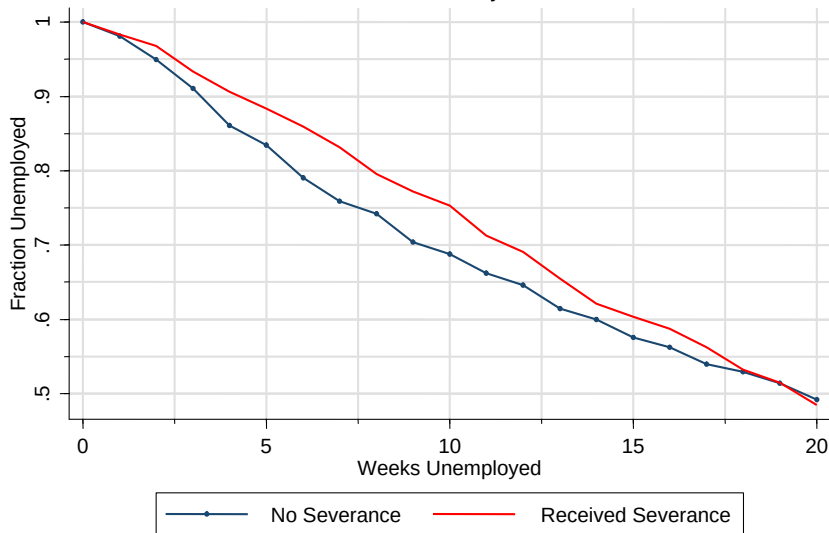
TABLE 3
Summary Statistics for Mathematica Data

	<u>Pooled</u>	<u>No Severance</u> (0.83)	<u>Severance</u> (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

Figure 5

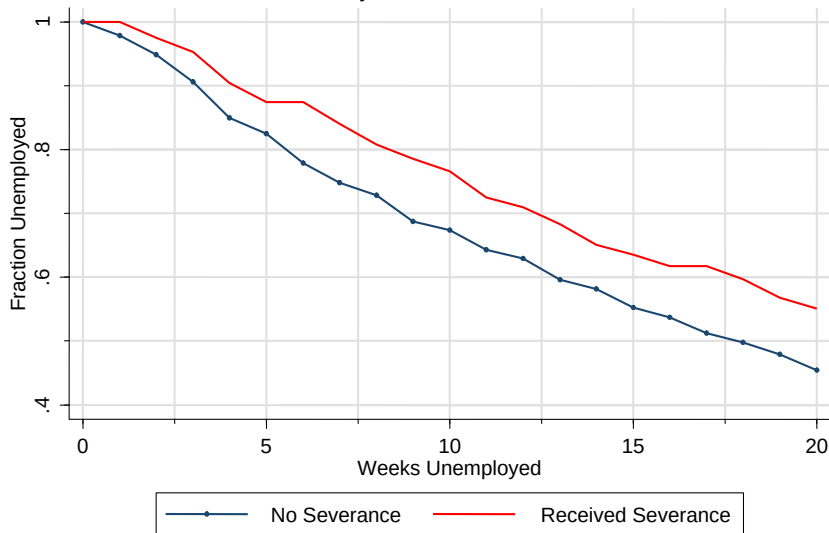
Effect of Severance Pay on Durations



Source: Chetty 2008

Figure 6a

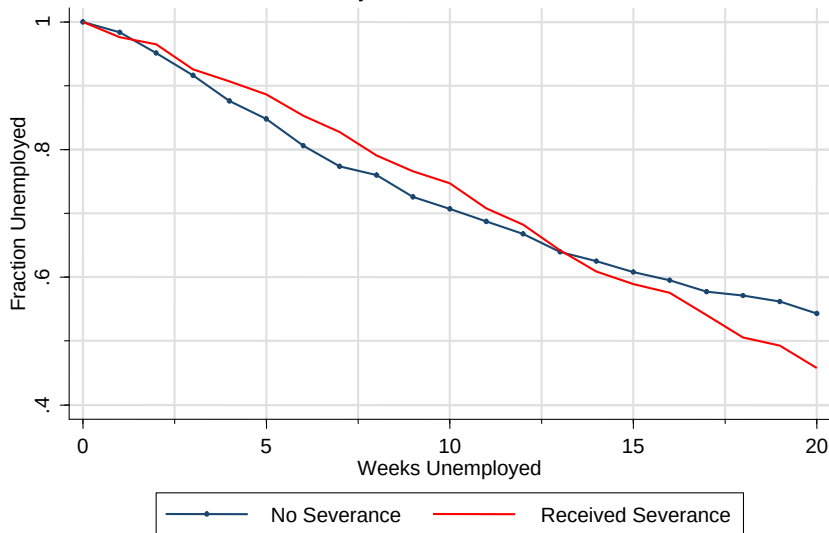
Effect of Severance Pay on Durations: Below Median Net Wealth



Source: Chetty 2008

Figure 6b

Effect of Severance Pay on Durations: Above Median Net Wealth



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.

Chetty 2008: Implications for Optimal UI

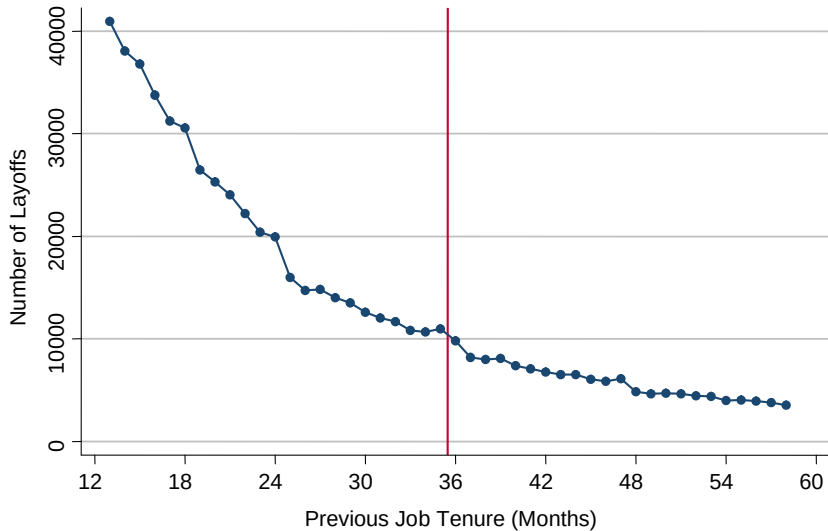
- Plug reduced-form estimates of de/dA and de/db into formula to calculate dW/db
- Welfare gain from raising benefit level by 10% from current level in U.S. (50% wage replacement) is \$5.9 bil = 0.05% of GDP
 - Small but positive
- In structural models calibrated to match sufficient statistics, dW/db falls rapidly with b
 - Small dW/db suggests we are currently near optimal benefit level

- 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

Frequency of Layoffs by Job Tenure



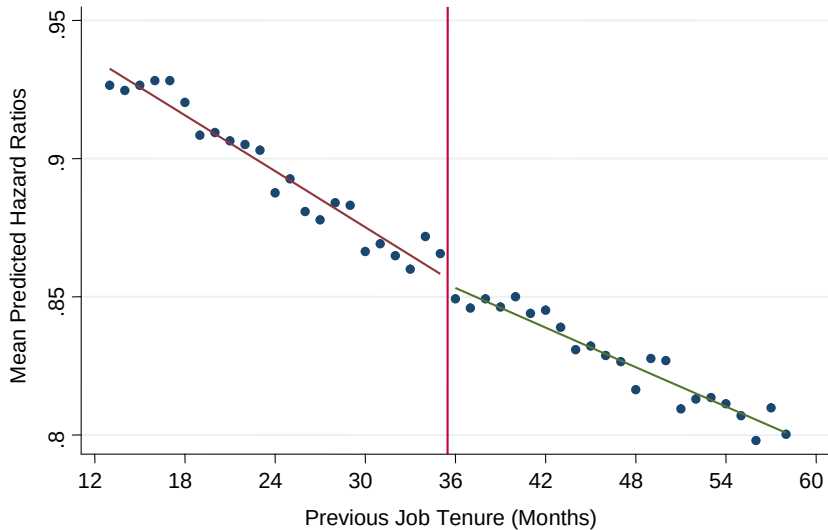
Source: Card, Chetty, and Weber 2007



Source: Card, Chetty, and Weber 2007

Figure 4

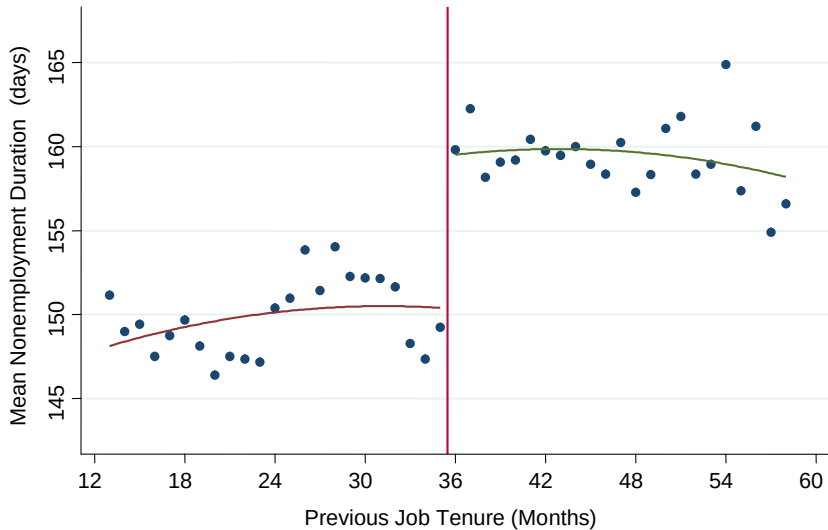
Selection on Observables



Source: Card, Chetty, and Weber 2007

Figure 5a

Effect of Severance Pay on Nonemployment Durations



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 (0.019)		-0.125 (0.017)
Extended benefits		-0.084 (0.018)	-0.093 (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.

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

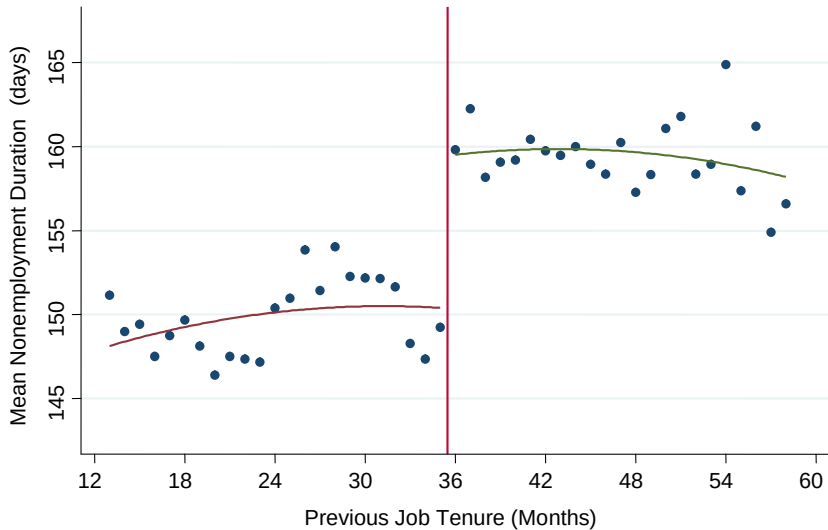
$$\begin{aligned}\frac{dW}{db} &= \frac{d\bar{w}_0}{db} - \frac{dt}{db} \\ &= \frac{d\bar{w}_0}{db} - \frac{1-e}{e} \cdot \left(1 + \frac{1}{e} \cdot \varepsilon_{1-e,b}\right)\end{aligned}$$

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

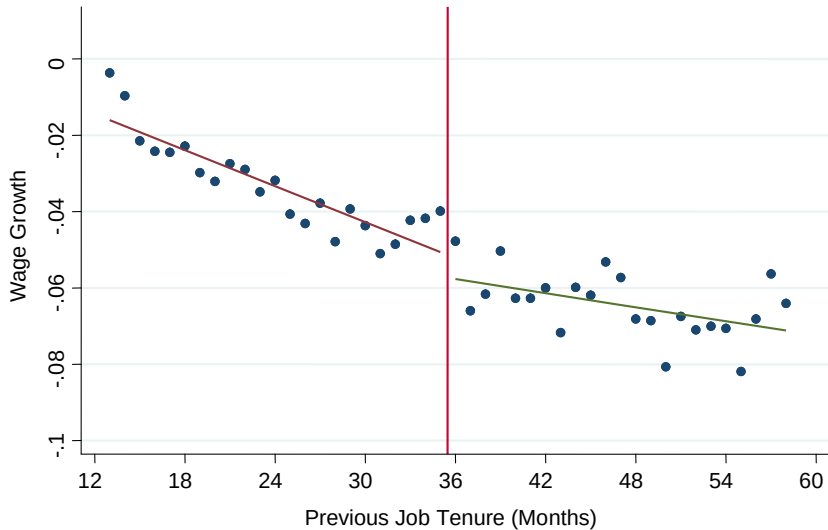
Effect of Severance Pay on Nonemployment Durations



Source: Card, Chetty, and Weber 2007

Figure 10a

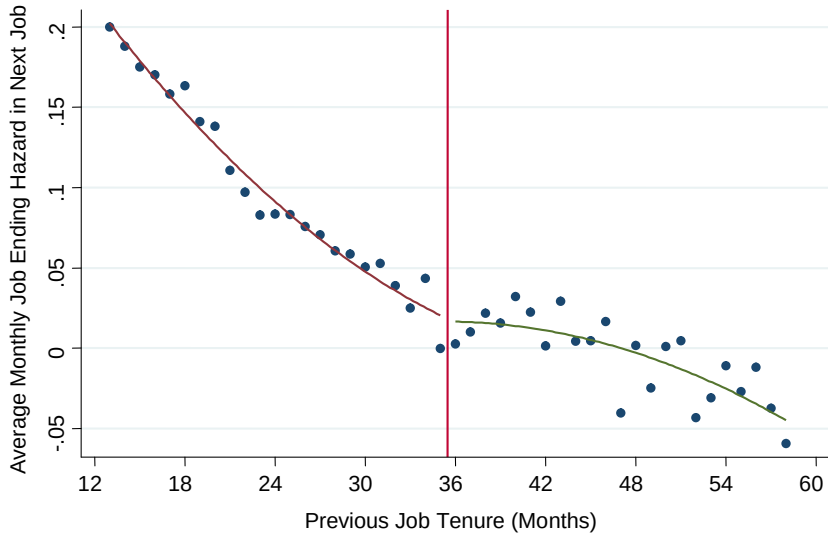
Effect of Severance Pay on Subsequent Wages



Source: Card, Chetty, and Weber 2007

Figure 10b

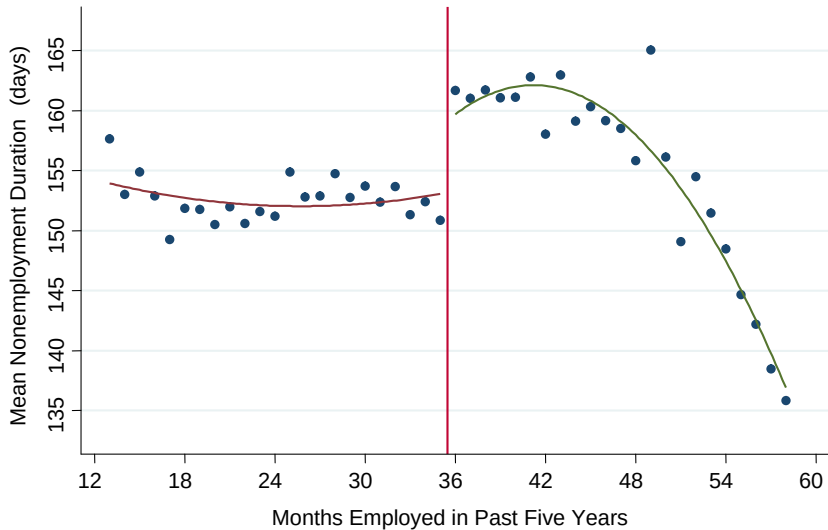
Effect of Severance Pay on Subsequent Job Duration



Source: Card, Chetty, and Weber 2007

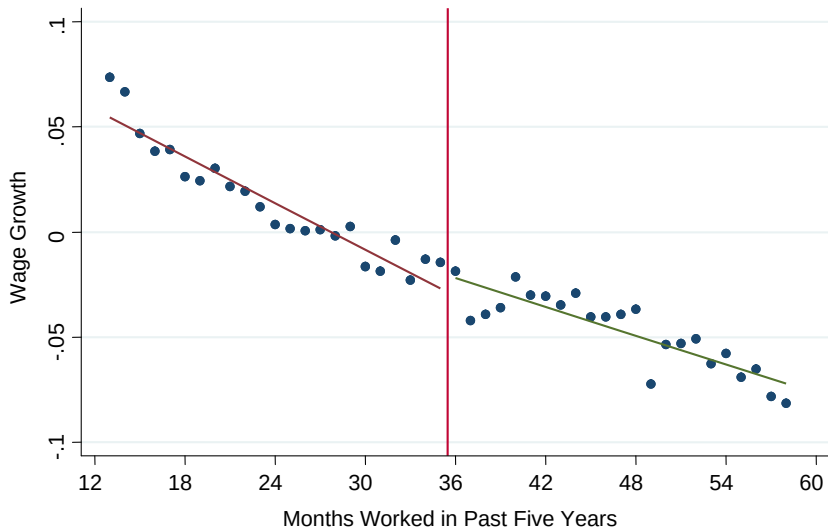
Figure 9a

Effect of Benefit Extension on Nonemployment Durations



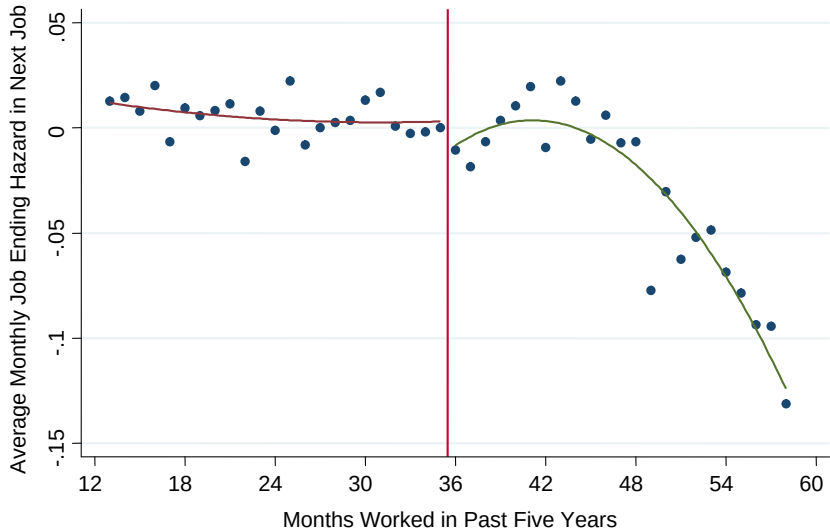
Source: Card, Chetty, and Weber 2007

Effect of Extended Benefits on Subsequent Wages



Source: Card, Chetty, and Weber 2007

Effect of Extended Benefits on Subsequent Job Duration

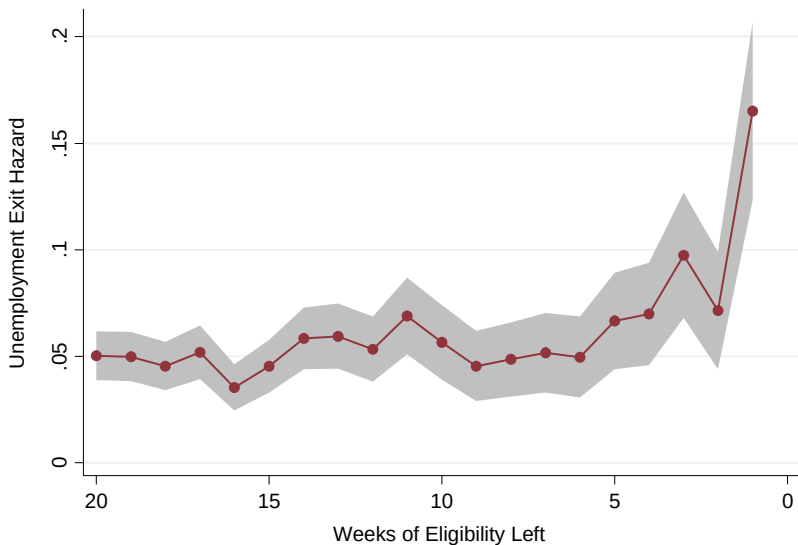


Source: Card, Chetty, and Weber 2007

Spike at Benefit Exhaustion

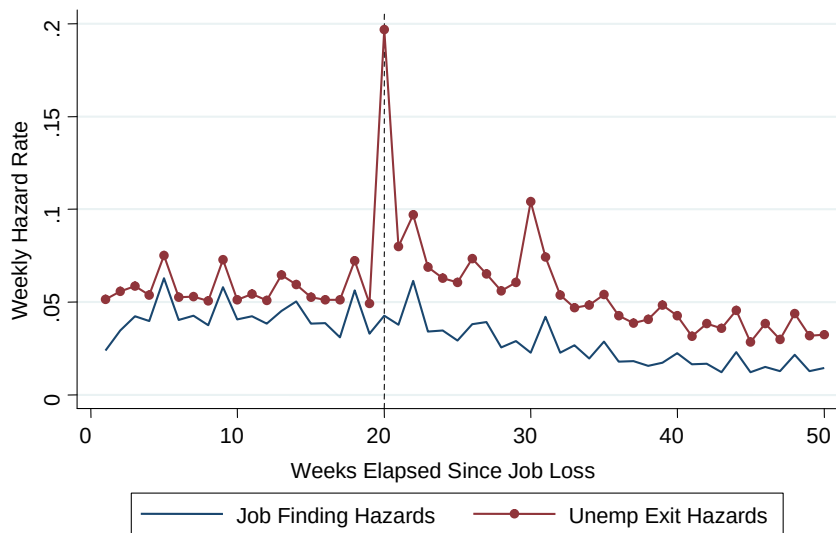
- Most striking evidence for distortionary effects of social insurance: “spike” in hazard rate at benefit exhaustion
 - Katz and Meyer (1990), Meyer (1990), ...
- Traditional measure of hazard: exiting UI system
- Preferred measure based on theory: finding a job
- The two could differ if workers transit 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

Time Until Benefits Lapse Empirical Hazard



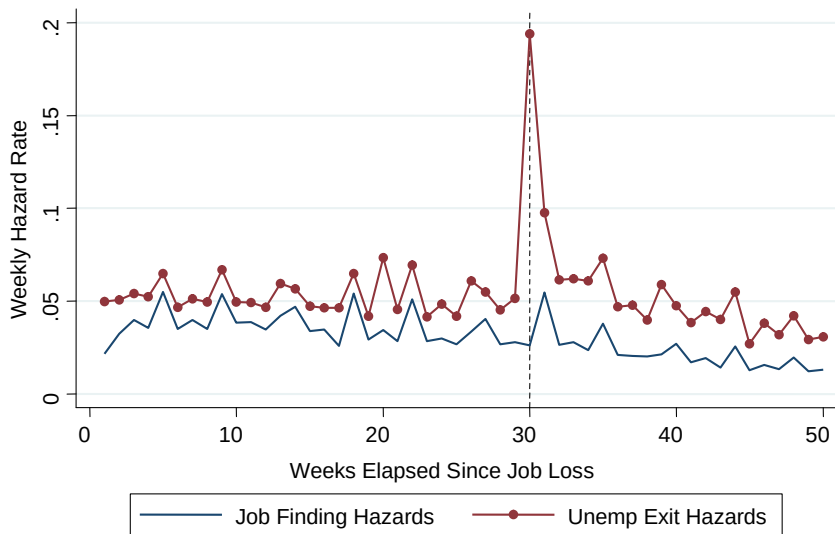
Source: Meyer 1990

Job Finding vs. Unemployment Exit Hazards: 20 Week UI



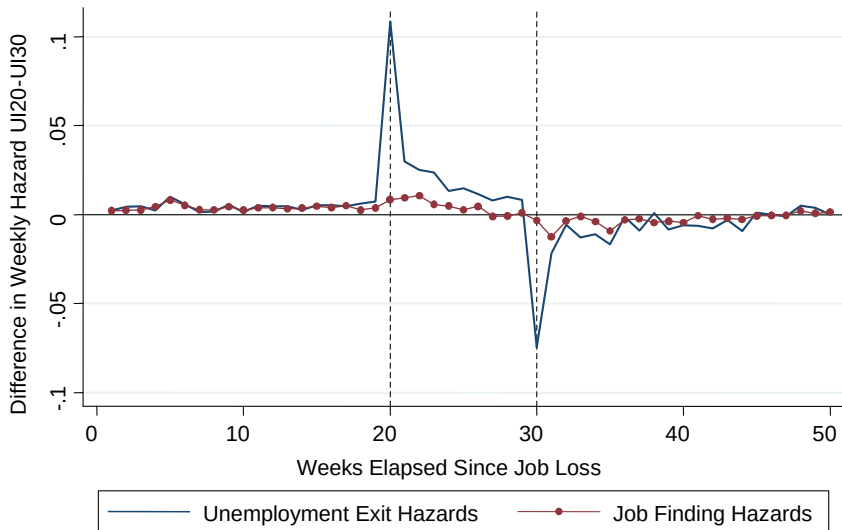
Source: Card, Chetty, Weber 2007b (AER P&P)

Job Finding vs. Unemployment Exit Hazards: 30 Week UI



Source: Card, Chetty, Weber 2007b (AER P&P)

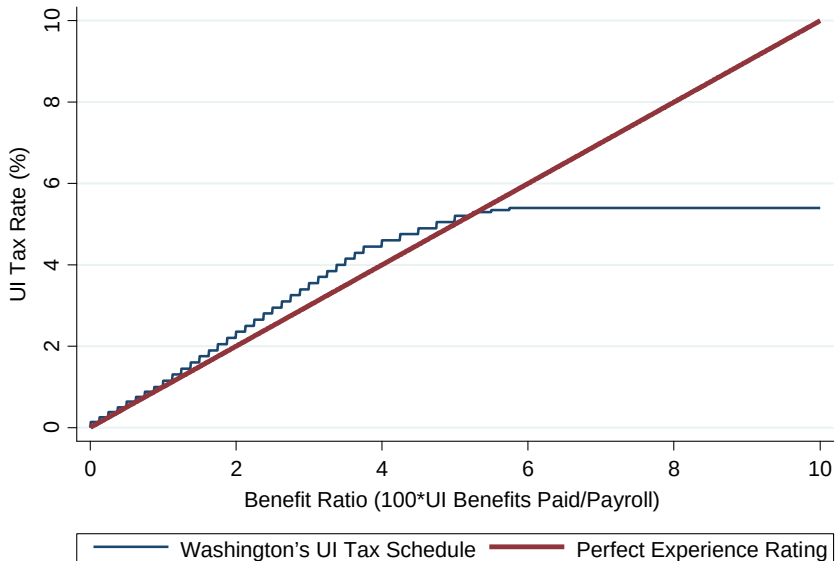
Effect of Benefit Expiration on Hazard Rates



Source: Card, Chetty, Weber 2007b (AER P&P)

- 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.

- 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

- 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.

- 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.

- 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

- 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.

- 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

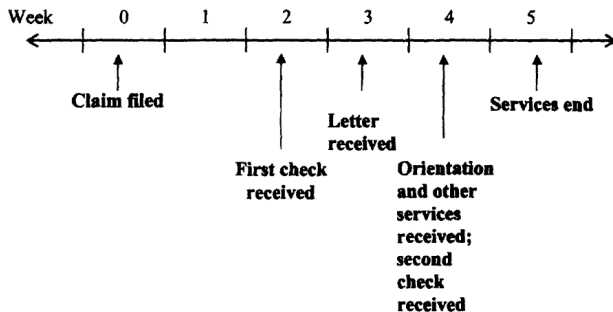


FIGURE 1. TIMELINE FOR TYPICAL UI CLAIMANT IN KENTUCKY WPRS PROGRAM

Source: Black, Smith, Berger, and Noel 2003

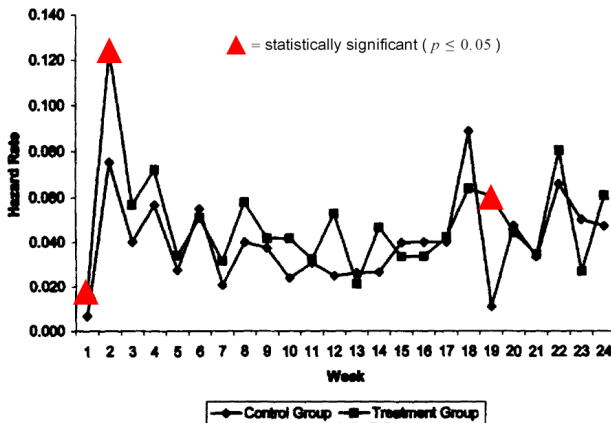


FIGURE 2. HAZARD FUNCTIONS OF THE TREATMENT AND CONTROL GROUPS, KENTUCKY WPRS EXPERIMENT, OCTOBER 1994 TO JUNE 1996

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 (0.509) [0.000]	-2.045 (0.411) [0.000]
UI benefits received	-143.18 (100.3) [0.077]	-81.44 (81.6) [0.159]
Fraction exhausting benefits	-0.024 (0.023) [0.152]	-0.030 (0.0019) [0.0055]
Earnings in the year after the start of the UI claim	1,054.32 (588.0) [0.037]	1,599.99 (475.2) [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 in 2003

State	Type of permanent impairment					Temporary Injury (10 weeks)
	Arm	Hand	Index finger	Leg	Foot	
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

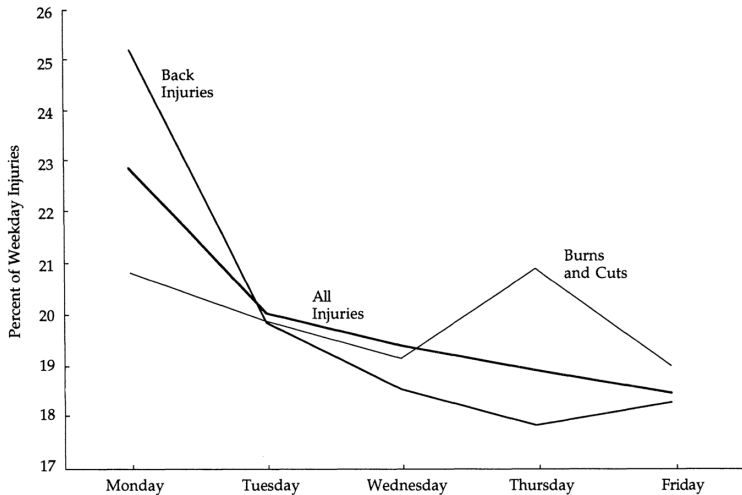


Figure 1. Distribution of Weekday Injuries.

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 explanations:
 - 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

Variable	High earnings		Low earnings		Differences		Difference in differences
	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 (0.83)	12.89 (0.83)	6.25 (0.30)	7.01 (0.41)	1.72 (1.17)	0.76 (0.51)	0.96 (1.28)
Michigan	14.76 (2.25)	19.42 (2.67)	10.94 (1.09)	13.64 (1.56)	4.66 (3.49)	2.70 (1.90)	1.96 (3.97)
Median duration (weeks)							
Kentucky	4.00 (0.14)	5.00 (0.20)	3.00 (0.11)	3.00 (0.12)	1.00 (0.25)	0.00 (0.16)	1.00 (0.29)
Michigan	5.00 (0.45)	7.00 (0.67)	4.00 (0.22)	4.00 (0.28)	2.00 (0.81)	0.00 (0.35)	2.00 (0.89)
Median medical cost (dollars)							
Kentucky	393.51 (19.29)	411.49 (22.72)	238.96 (8.48)	254.40 (9.11)	17.98 (29.80)	15.44 (12.44)	2.55 (32.30)
Michigan	689.73 (77.30)	765.00 (134.53)	390.63 (32.80)	435.00 (33.09)	75.27 (155.16)	44.38 (46.59)	30.89 (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 could be biased by selection

Effect on Equilibrium Wage

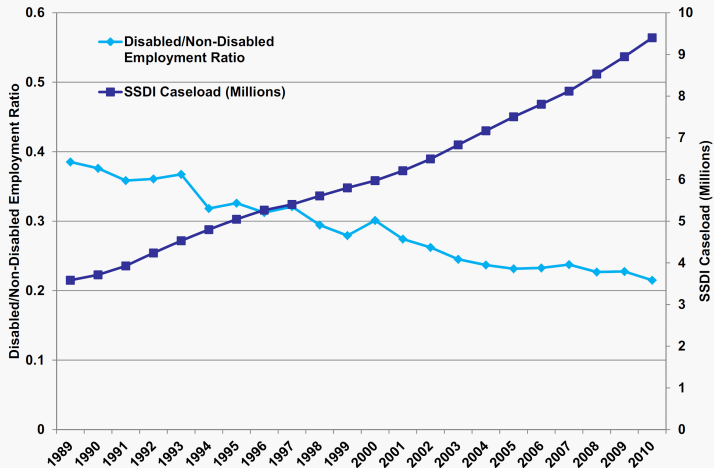
- Workers' compensation is a mandated benefit
 - When firms hire, should adjust wage downwards if workers value benefit (Summers 1989)
- Gruber-Krueger (1991) test this using changes in WC laws
 - 85% of WC cost is shifted to workers, no significant employment effect
- Fishback-Kantor (1995) study initial implementation of program
 - Find 100% shift to workers' wages
- Both studies suggest 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

- 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

Figure 1. The Rise of SSDI and the Decline in Employment of the Disabled



Source: Authors' tabulations of Current Population Survey; Annual Statistical Report on the Social Security Disability Insurance Program, 2010, Table 3.

Source: Mullen, Maestas, Strand (2012)

Two Views on the Rise in DI

- One perspective: 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 disentangle these two views

TABLE 1—REASSESSMENTS OF INITIAL
SOCIAL SECURITY DETERMINATIONS

*A. Bureau of Disability Insurance Review One Year
After Initial Determination (Percentages):*

BDI assessment	Initial determination	
	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

- 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 1980's is of methodological interest
- Parsons (1980)
 - Uses cross-sectional variation in replacement rates
 - Data on men aged 45-59 in 1966-69 NLSY
 - OLS regression:
$$LFP_i = \alpha + \beta DI_{retrate}_i + \varepsilon_i$$
where $DI_{retrate}$ 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

- Bound highlights key econometric problem in Parson's specification
 - Dltreprate variation correlated with wage
 - Identification assumption: LFP rates equal across wage groups
- Parson's solution: "control" for wage rate

$$LFP_i = \alpha + \beta Dltreprate_i + f(wage_i) + \varepsilon_i$$

- Does this resolve the problem?

Identification by Functional Form

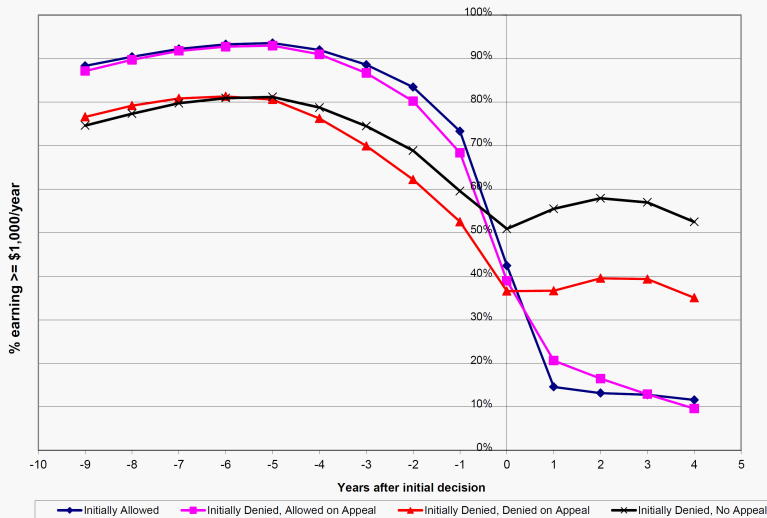
$$LFP_i = \alpha + \beta DI_{reprice_i} + f(\text{wage}_i) + \varepsilon_i$$

- This is an example of identification by “functional form”
- As f is made increasingly flexible, standard error on β goes to infinity
- Problem is that only source of variation is due to wages
- To illustrate practice importance, Bound replicates Parson's regression on sample that never applied to DI and obtains a similar elasticity
- Motivates literature that focuses on quasi-experiments other sources of non-parametric identification
 - Key idea of non-parametric identification: with sufficiently large samples, estimate is identified without parametric assumptions on f
 - Impose functional forms only for computational convenience and precision in finite samples

Empirical Evidence: Bound-Parsons Debate

- Bound (1990) 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
- Illustrate using better data from Mullen, Maestas, Strand (2012)

Figure 2. Employment Before and After Initial Decision
2005 Decisions



Source: Mullen, Maestas, Strand (2012)

- 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

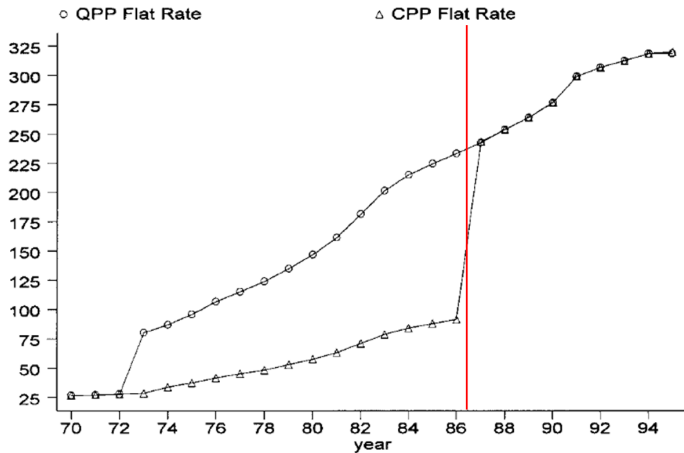


FIG. 1.—Flat-rate portion in Quebec and the rest of Canada

Source: Gruber 2000

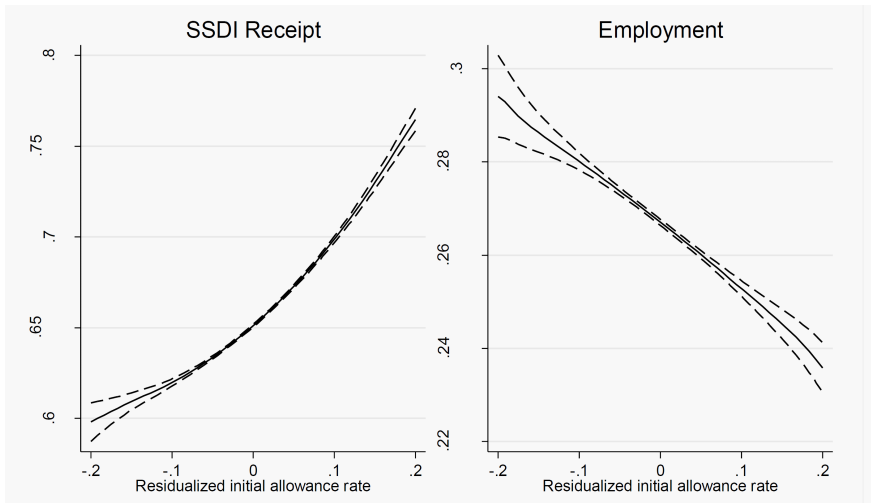
TABLE 1
MEANS

	CPP		QPP		DIFFERENCE IN DIFFERENCE (5)
	Before (1)	After (2)	Before (3)	After (4)	
Benefits	5,134	7,776	6,878	7,852	1,668 (17)
Replacement rate	.245	.328	.336	.331	.088 (.003)
Not em- ployed 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

- Implied elasticity of non-employment rate w.r.t. DI benefit level: 0.25-0.3
- Agrees more with Bound than Parsons
- But estimates are imprecise and only capture short-run effects

- Maestas, Mullen, and Strand (2012) use random variation in examiner assignment to identify effects of DI
- Disability cases randomly assigned by computer to examiners at state board
- Substantial discretion generates significant variation across examiners in allowance rates
- Instrument for receipt of DI w/ examiner's conditional allowance propensity
 - Another approach: use set of examiner f.e.'s as instruments
- Use administrative data on DI decisions and earnings from SSA
 - 1 million observations



Source: Mullen, Maestas, Strand (2012)

- Estimate causal effects of DI on employment rates using IV regressions
- First stage:

$$DI_i = a + \phi_i \text{examiner}_i + v_i$$

- Second stage:

$$y_i = \alpha + \beta DI_i + \varepsilon_i$$

- Note that first stage coeffs. ϕ_i are average DI allowance rates by examiner
- Therefore IV regression is equivalent to examiner-level OLS regression

$$\bar{y}_e = a + \beta \overline{DI}_e + \varepsilon_e$$

- Weighting this OLS regression by number of individuals per examiner will yield identical estimate of β

Table 4. Effects of SSDI Receipt on Employment and Earnings

Outcome	2005 Decisions					
	2 Years After Decision		3 Years After Decision		4 Years After Decision	
	OLS	IV	OLS	IV	OLS	IV
1) Earn $\geq$ \$1,000/year						
Mean dependent variable allowed	0.148		0.128		0.106	
Mean dependent variable denied	0.522		0.515		0.471	
Coeff. on ALLOW	-0.347*** (-322.48)	-0.279*** (-8.64)	-0.361*** (-336.60)	-0.227*** (-6.99)	-0.345*** (-321.81)	-0.158*** (-4.83)
R-squared	0.200	0.195	0.218	0.200	0.209	0.171

Source: Mullen, Maestas, Strand (2012)

- Conclude that DI receipt reduces probability of employment by 28% for *marginal applicants*
- Important to recognize that this is a LATE for people who are at the margin of getting DI
- Severely disabled individuals would be granted DI by all examiners and are not captured in this LATE
- Hence should be interpreted as an upper bound on ATE
- Maestas et al. confirm this by studying heterogeneity in treatment effects by disease severity

Table 6. Heterogeneity: Effects of SSDI Receipt on Employment (2 Years Later), 2005 and 2006 Combined

	No. obs.	Mean Empl. Allowed	Mean Empl. Denied	OLS		IV	
				Coeff. On ALLOW	t-Stat.	Coeff. On ALLOW	t-Stat.
Body system code							
Musculoskeletal system	839,847	0.125	0.520	-0.361	-300.48	-0.173	-3.42
Mental disorders	455,433	0.171	0.535	-0.333	-210.25	-0.366	-11.73
Cardiovascular system	185,063	0.116	0.462	-0.328	-134.18	-0.335	-5.81
Neurological	181,162	0.119	0.538	-0.386	-144.34	-0.359	-7.40
Endocrine system	94,156	0.130	0.467	-0.316	-100.41	-0.494	-2.14
Respiratory system	88,578	0.099	0.462	-0.309	-84.61	-0.061	-0.51
Special/other	85,587	0.132	0.322	-0.222	-57.78	2.000	0.59
Malignant neoplastic diseases	77,021	0.212	0.621	-0.390	-94.75	-0.347	-6.18
Immune system disorders	59,188	0.170	0.531	-0.330	-65.02	-0.048	-0.61
Digestive system	53,224	0.158	0.523	-0.353	-80.64	-0.324	-3.08
Special senses and speech	48,396	0.158	0.529	-0.344	-77.59	-0.158	-1.37
Genitourinary impairments	31,837	0.185	0.540	-0.327	-41.38	-0.084	-0.74
Skin disorders	7,483	0.156	0.560	-0.377	-32.25	-0.916	-3.66
Hematological disorders	7,091	0.219	0.590	-0.336	-26.35	0.116	0.44

Source: Mullen, Maestas, Strand (2012)

- Results highlight importance of estimating relevant LATE for policy
- If policy question is raising DI benefit, relevant treatment effect is for people on margin w.r.t. benefit change
- May not be the same people as those who are on the margin with respect to examiner decision
- This is the advantage of *directly* studying the policy of interest
 - Does not require extrapolations from estimated LATE to the policy-relevant sufficient statistic
- Also important to note that paper estimates uncompensated elasticities
 - Critical to distinguish moral hazard vs. liquidity for normative purposes
 - Might still have people with disutility $\varphi_i > w_i$ working when rejected

Methodological Note: Weak Instruments

- Random assignment instrument to judges, classes, etc. now popular in many applications
- Common problem: weak instruments
- Arises when each examiner has few cases
 - Not an issue in Maestas et al. but common e.g. in education with 20 kids per class
- In this case, IV estimate of β will be biased toward OLS

Weak Instruments Problem

- Recall examiner-level OLS regression

$$\bar{y}_e = a + \beta \overline{DI}_e + \varepsilon_e$$

- Same observation appears on LHS and RHS of this regression
 - With few individuals, this biases β toward OLS
 - Extreme case: one individual per examiner equivalent to OLS
- If an examiner gets a draw of particularly sick people, they will both get DI and have low employment rates even if DI has no causal effect

Weak Instruments Problem

- Problem vanishes as number of individuals *per examiner* grows large
 - Only remaining variation in $\overline{DI}_e$ is due to examiner effects
- Key question: does number of instruments grow at same rate as sample size?
- If sample gets bigger by adding more people per examiner, then instruments are asymptotically strong
- If sample gets bigger by adding more examiners, then instruments and asymptotically weak

- How to obtain unbiased estimates with weak instruments?
- Traditional recommendation: LIML
- Currently preferred alternative: Jackknife IV
 - Leave out own observation for each i when estimating $\overline{DI}_e$ in first stage
 - Second stage regression becomes

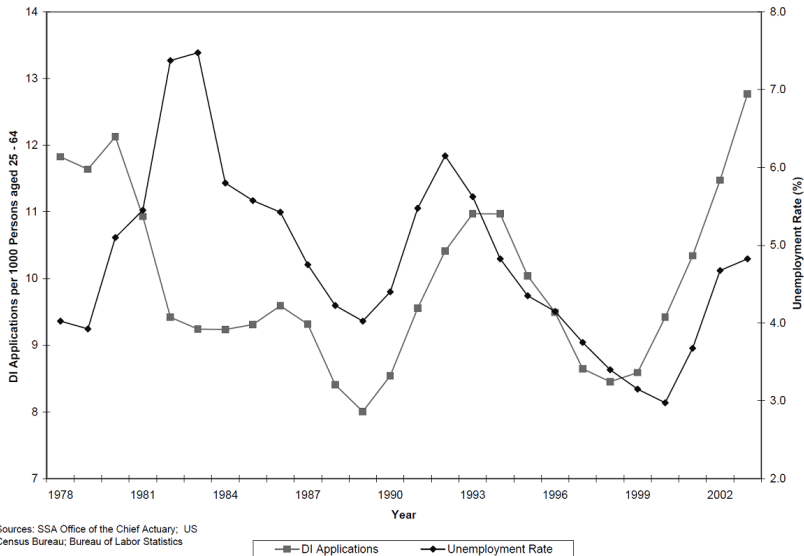
$$y_i = a + b\overline{DI}_{j \neq i} + \varepsilon_i$$

- Directly fixes own-observation problem and is more robust than LIML (Kolesar et al. 2012, Kolesar 2012)

- 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 $\rightarrow$ unemployment rate lower
- But economic situation is the same: one less person working
- Test this hypothesis using cross-state variation in employment shocks

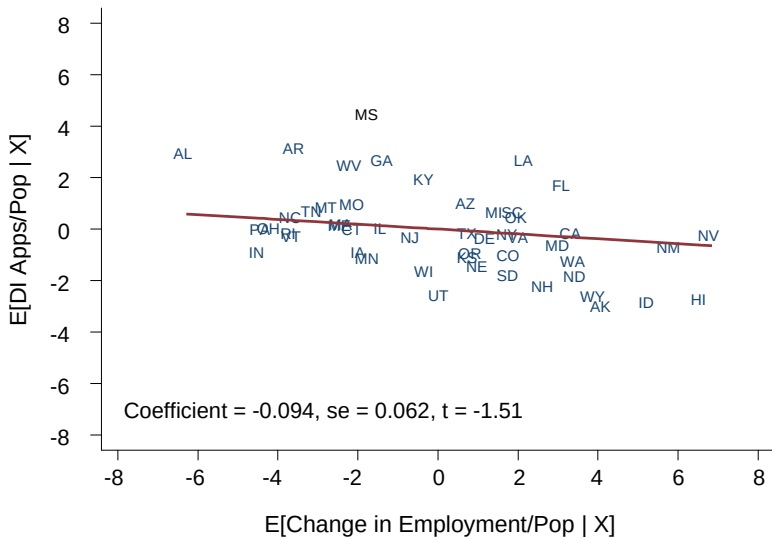
- Standard technique to construct state-level employment shocks over a five year window:
 - Calculate industry shares in a given state in base year
 - Calculate employment changes over five year period by industry using data on national employment (excluding state in question)
 - Project changes in each 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

Figure 5: DI Applications and Unemployment Rate



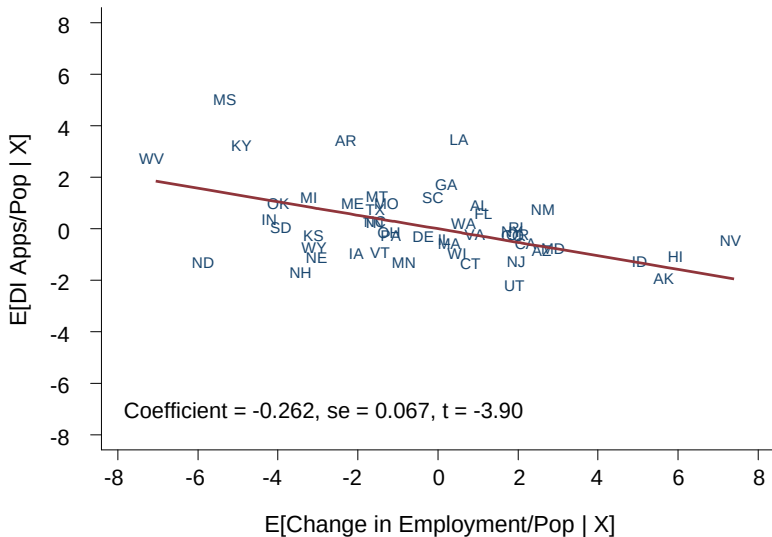
Source: Duggan and Imberman 2005

Employment Shocks and DI Applications: 1979-1984



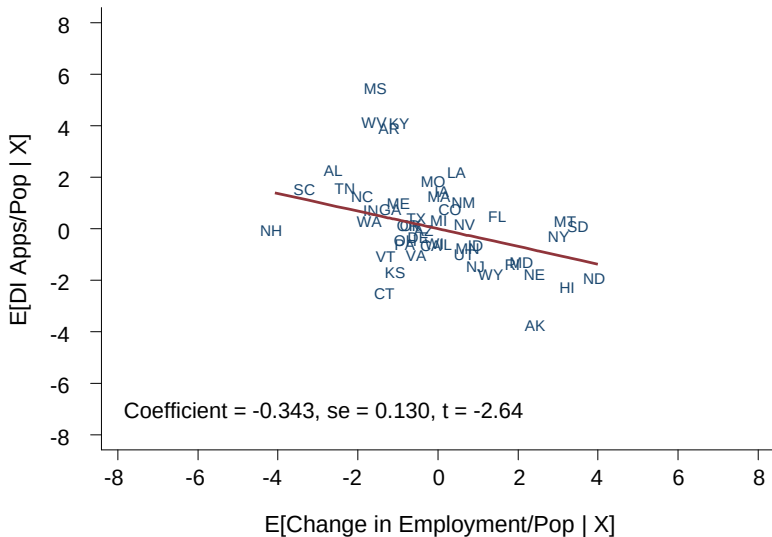
Source: Autor and Duggan 2003

Employment Shocks and DI Applications: 1984-1989



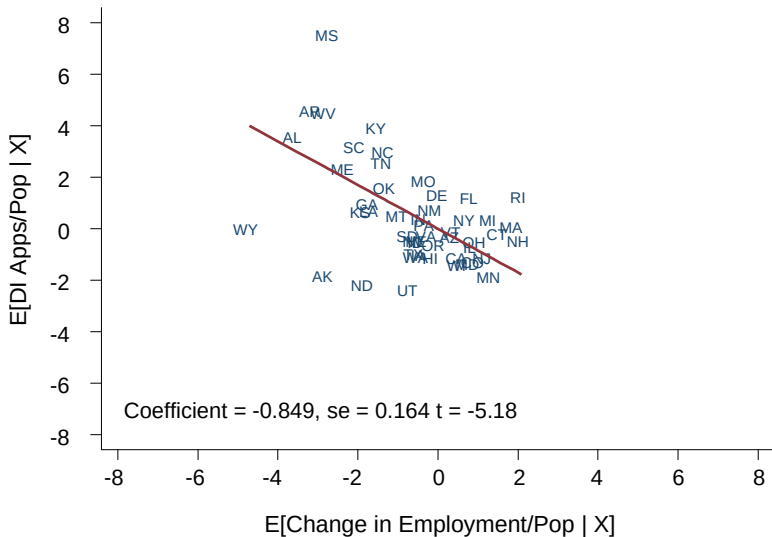
Source: Autor and Duggan 2003

Employment Shocks and DI Applications: 1989-1994



Source: Autor and Duggan 2003

Employment Shocks and DI Applications: 1993-1998



Source: Autor and Duggan 2003

- 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 progressivity 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

Public Economics Lectures

Part 7: Public Goods and Externalities

Raj Chetty and Gregory A. Bruich

Harvard University
Fall 2012

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

- 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 revealed-preference

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

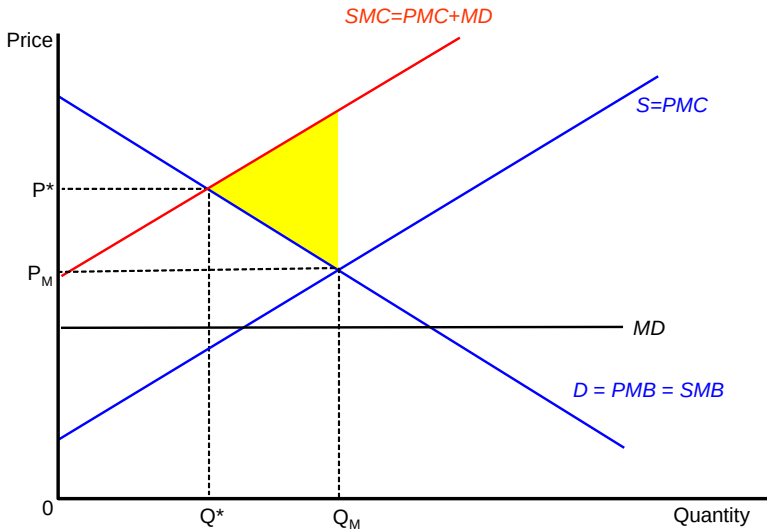
$$c'(x^S) = p$$

- PMB equals PMC in equilibrium:

$$u'(x^D) = c'(x^S)$$

- But this is not Pareto efficient

Negative Production Externalities: Pollution



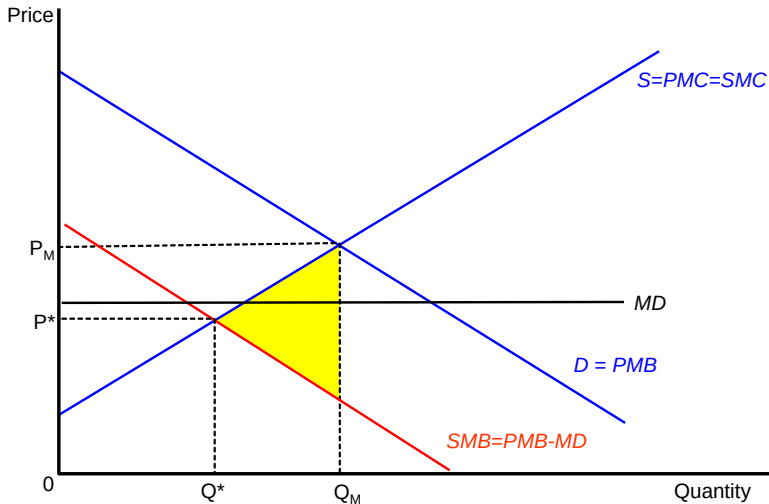
Model of Externalities: Deadweight Loss

- 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

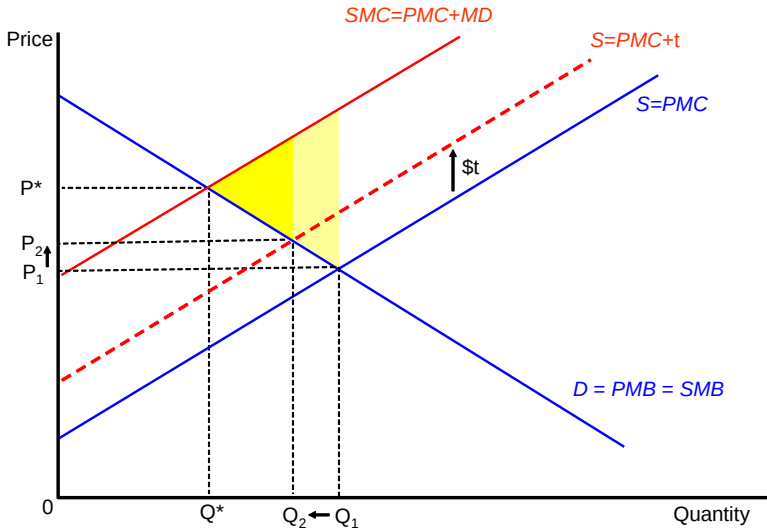
② Asymmetric information: competitive equilibrium can break down

- 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



- Quantity-based restriction: reduce pollution to fixed level or face legal sanctions
- Same outcome as Pigouvian taxation: move people to x_2
- Disadvantages: no marginal incentives
 - Allocative inefficiency with heterogeneity in costs of pollution reduction
 - Dynamic inefficiency: no incentive to innovate
- These problems can be solved by cap and trade system

Permits: Cap-and-Trade

- Cap total amount of pollution and auction permits to firms
- Then allow firms to trade permits to pollute
- Hybrid of regulation and Coasian solution: create the market
- 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

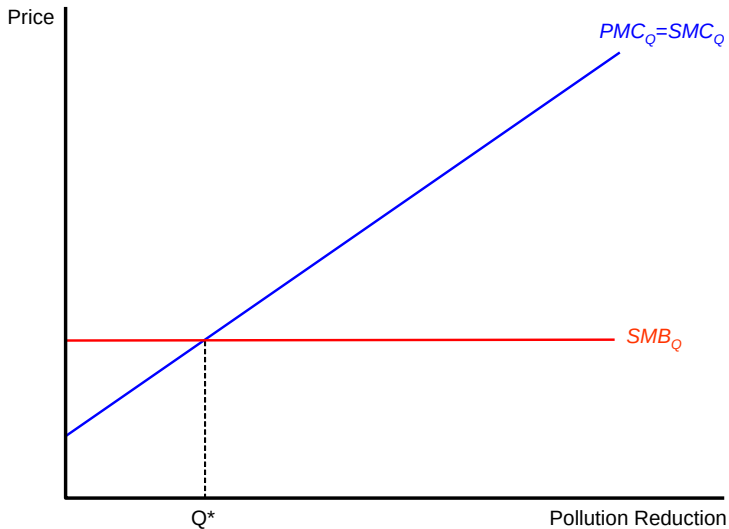
Weitzman 1974: Prices vs. Quantities

- 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: Prices vs. Quantities

- 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



Optimal 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 $p^* = C'(Q^*)$.

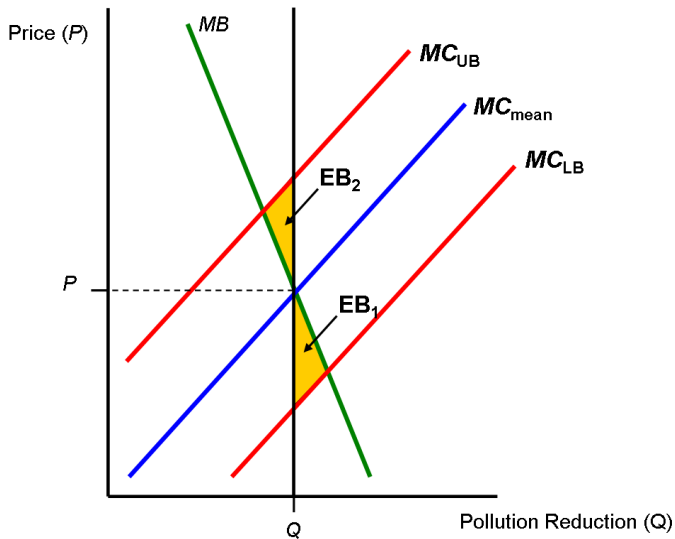
Optimal Policy with Uncertainty

- 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:

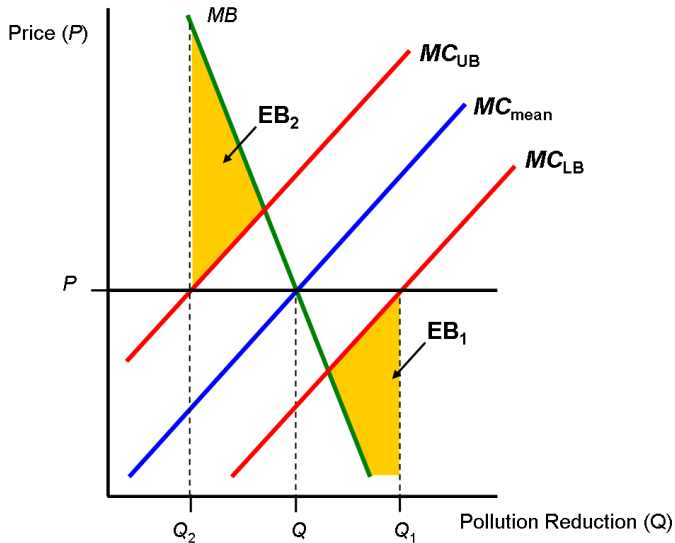
$$E_{\theta}[B(Q^*) - C(Q^*, \theta)] >? E_{\theta}[B(Q(p^*)) - C(Q(p^*), \theta)]$$

- Optimal choice depends on $B''(Q)/C''(Q)$
- Quantity regulation preferred if MB steep relative to MC

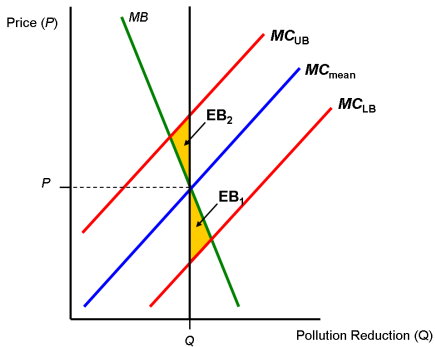
MB steep, Quantity regulation



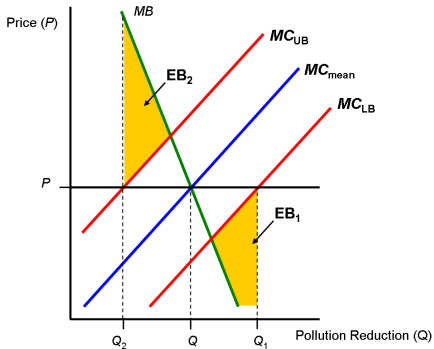
MB Steep, Price Regulation



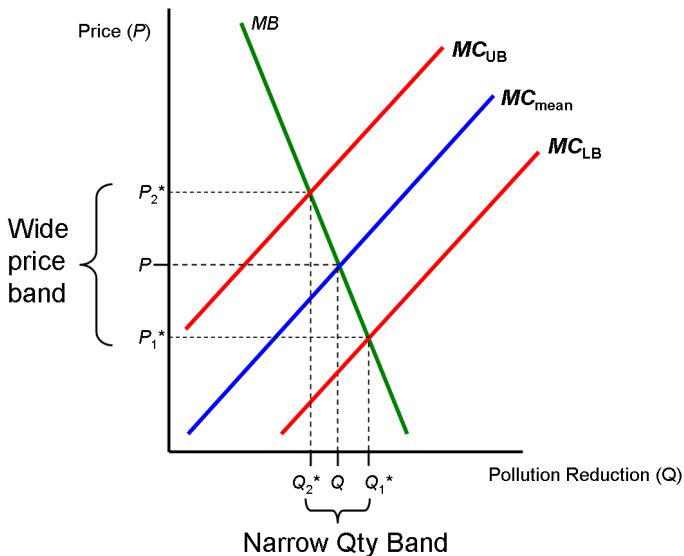
Quantity Regulation



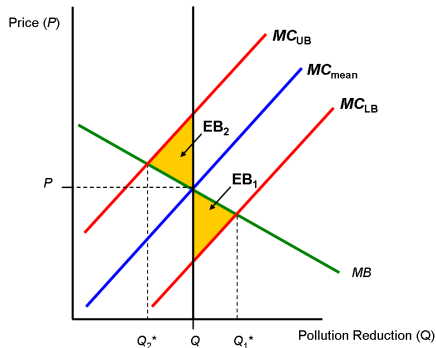
Price Regulation



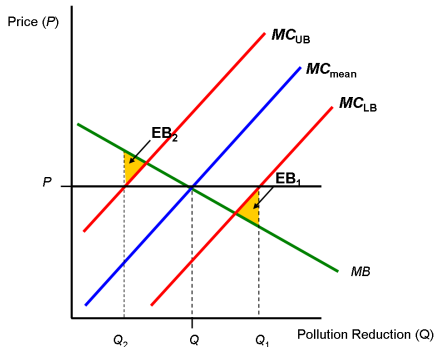
Price Band vs. Quantity Band with Steep MB



MB Flat, Quantity regulation



MB Flat, Price Regulation



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:

$$u(x_1, \dots, x_N, l) - d(x_N)$$

- Utility is maximized subject to:

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

Sandmo 1975: Setup

- Individual maximization program

$$L = u(x_1, \dots, x_N, l) + \lambda(wl + Z - (q_1x_1 + \dots + q_Nx_N))$$

- 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

- 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

$$\begin{aligned}\frac{\tau_i - \tau_{ip}}{1 + \tau_i} &= (\theta/\lambda)/\epsilon_{ii}^c \\ \Rightarrow \tau_i &= \frac{\theta x_i^c}{\lambda} / \frac{dx_i^c}{dp_i} + \tau_{ip} \\ &= \tau_{ip} + \tau_{ir}\end{aligned}$$

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

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

Double Dividend Debate

- Suppose we discover that production of computers generates negative externality.
- 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

- Accident externalities from driving automobiles.
- If I drive, I increase probability you will get into an accident → 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.)

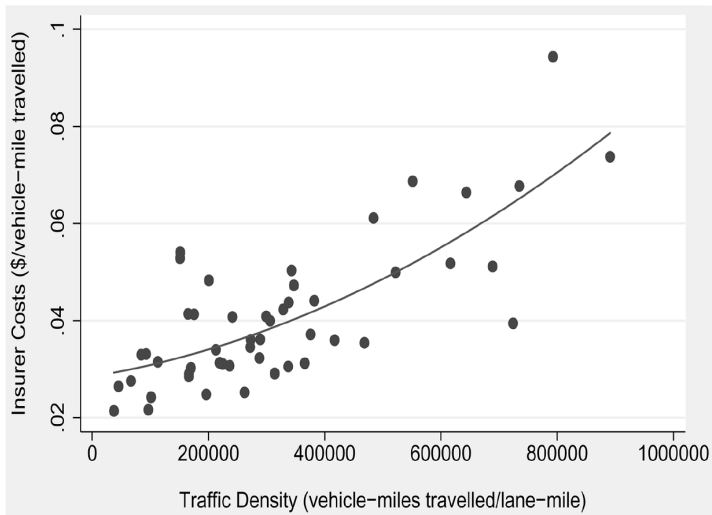


FIG. 1.—Traffic density and insurance costs (1996 dollars)

Source: Edlin and Karaca-Mandic 2006

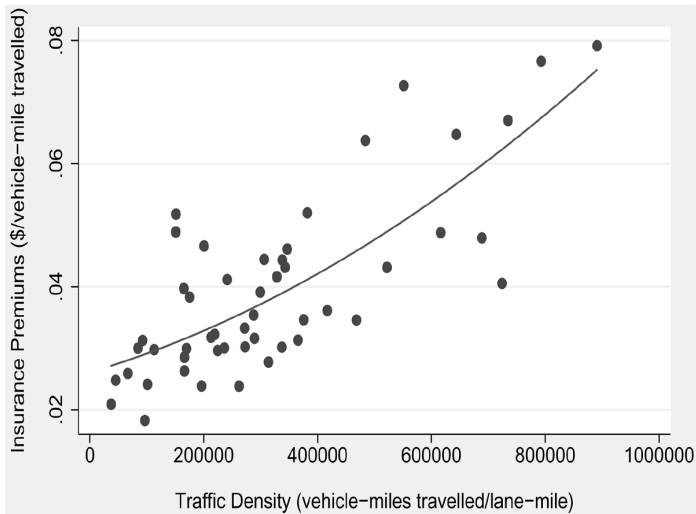


FIG. 2.—Traffic density and insurance premiums (1996 dollars)

Source: Edlin and Karaca-Mandic 2006

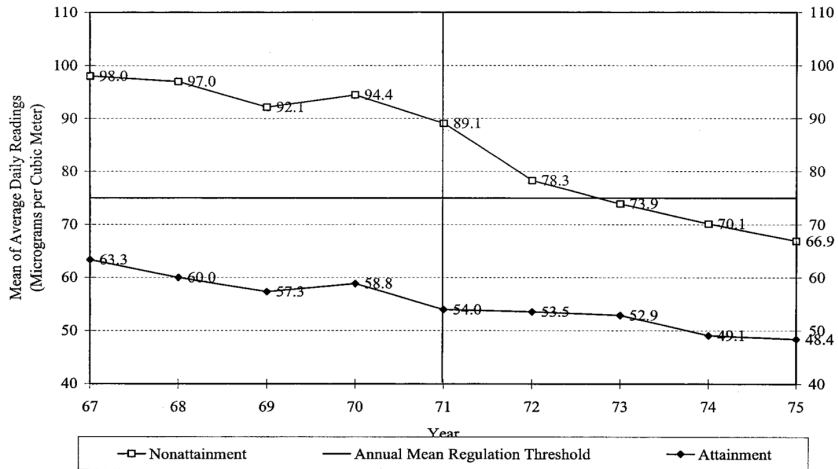
- Traffic density substantially increases marginal insurer costs
- Insurers set premia to cover average costs in market equilibrium
- Individual's marginal cost on other driver's not internalized
- Externality is convex
 - Increase in traffic density from average driver has external cost of \$2,000 per 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

- 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.

$$P_i = \alpha + \text{Pollution}_i + X_i\beta + \epsilon_i$$

- Econometric problems
 - Omitted variables: polluted neighborhoods worse on many dimensions
 - Deeper problem: sorting
 - Recover marginal WTP rather than average WTP
 - People with allergies avoid polluted areas

- 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)
<i>F</i> -statistic TSPs nonattainment*	31.3 (1)	29.9 (1)	24.4 (1)	21.5 (1)
<i>R</i> ²	.04	.10	.19	.20
B. Log Housing Changes				
TSPs nonattainment in 1975 or 1976	.036 (.012)	.022 (.009)	.026 (.008)	.019 (.008)
<i>F</i> -statistic TSPs nonattainment*	8.5 (1)	6.2 (1)	9.3 (1)	6.4 (1)
<i>R</i> 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

- 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

- Without tradeable permits, efficiency costs of regulation can be very high because of allocation distortions
- Study allocation of apartments under rent control
- Standard model assumes that with price controls, still have allocative efficiency
 - Those who value the apartments most get them
- But regulation will generally lead to allocative inefficiency that generates first-order welfare losses
 - For small price caps, allocation inefficiency dwarfs undersupply inefficiency

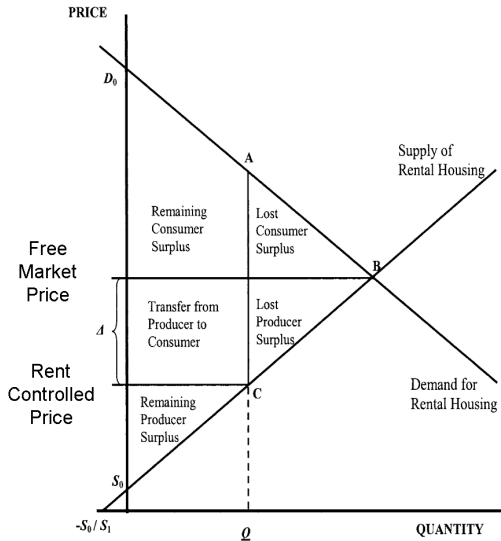


FIGURE 1. CLASSICAL ANALYSIS OF WELFARE LOSSES FROM RENT CONTROL

Source: Glaeser and Luttmer 2003

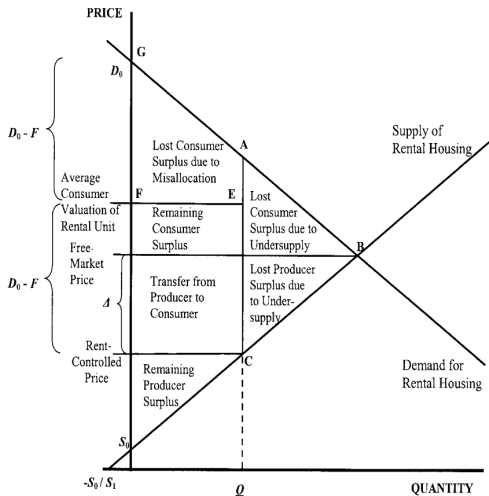


FIGURE 2. THE WELFARE LOSSES FROM RENT CONTROL WHEN APARTMENTS ARE RANDOMLY ALLOCATED ACROSS CONSUMERS

Source: Glaeser and Luttmer 2003

- 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

Allocation of New York City households across apartments (fraction of households)						
Actual apartment size (number of rooms):	Efficient apartment size (number of rooms):					Marginal
	1	2	3	4	5	
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

- 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?

- 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

- 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

- Show that addictive goods can be modeled in perfectly rational framework
- Dynamic model with habit formation
- Current consumption of the addictive good decreases utility in future periods 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

- Hyperbolic discounting preferences for smokers

$$U_0 = u(c_0) + \beta \left(\sum_{t \geq 1} \gamma^t u(c_t) \right) \text{ with } \beta < 1.$$

$$U_1 = u(c_1) + \beta \left(\sum_{t \geq 2} \gamma^t u(c_t) \right)$$

- 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.

- 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

- 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/subsidies.
 - For rational individuals, excess burden due to taxation is second-order (Harberger triangle).
 - For irrational individuals, welfare gains from correction of internalities is first-order (Harberger trapezoid)
 - Therefore always optimal to have a positive tax; calibrations suggest fairly large corrective taxes

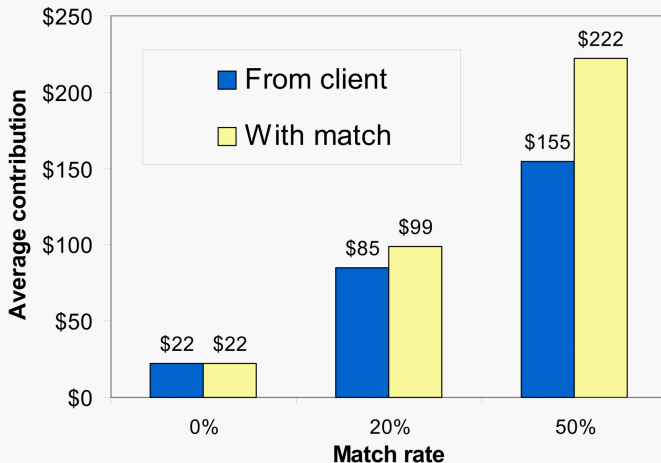
Application: Retirement Savings

- Many believe that people do not save enough for retirement because of myopia, self-control problems, etc.
- What are the best corrective policies to increase savings?
 - Price subsidies: 401(k)'s, IRA's (Duflo et al. 2006)
 - Nudges: defaults and automatic enrollment in pension plans (Madrian and Shea 2001)
 - Commitment devices (Bernartzi and Thaler 2004; Ashraf, Karlan, and Yin 2006)
 - Information provision and financial literacy (Lusardi and Mitchell 2011)
- Focus on the first two here

Duflo et al. (2006): Price Subsidies

- Duflo et al. conduct a randomized experiment providing matches for IRA contributions
- Subject pool: H&R Block tax filers
- Main finding: provision of a non-zero subsidy significantly increase IRA participation rates
- Equivalent government Saver's Credit program has no impact, suggesting that salience matters

Effects on contributions (unconditional)



Source: Duflo et al. (2006)

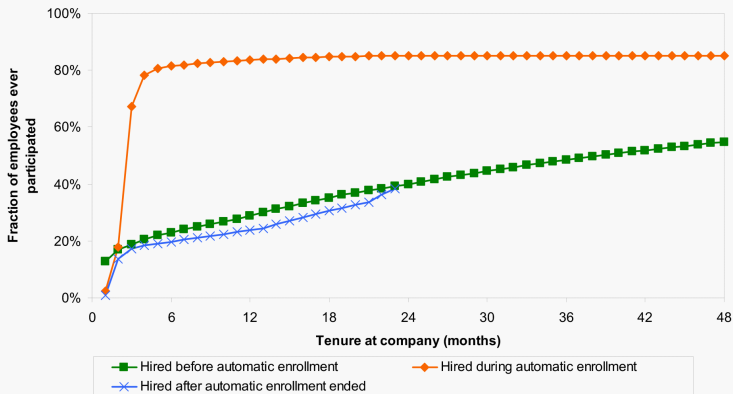
Madrian and Shea (2001): Defaults

- Madrian and Shea show that defaults have powerful effects on savings behavior even though they do not change budget set
- Clearly violates neoclassical model
 - Utility consequences of changing retirement savings rate are large
 - Therefore difficult to explain with optimization costs
- Carroll et al. (2009) propose a model with optimization costs and hyperbolic discounting to explain the pattern
 - Hyperbolic discounters keep postponing plans to set up retirement account because cost of delaying by one period is small
 - Interesting implication: short deadlines may help *improve* decision making

Automatic enrollment effect

Automatic enrollment dramatically increases participation.

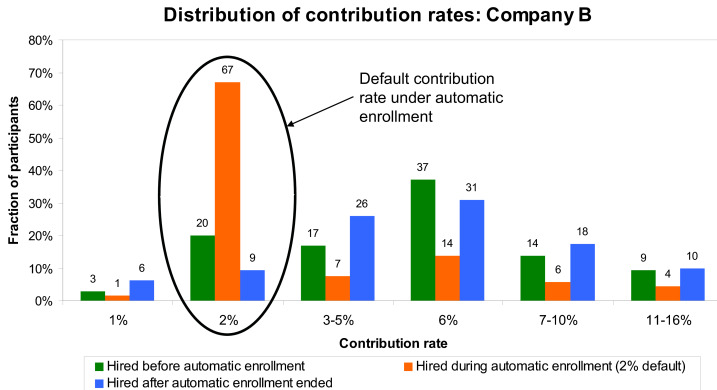
401(k) participation by tenure at firm: Company B



Source: Madrian and Shea (2001)

Automatic enrollment effect

Employees enrolled under automatic enrollment cluster at the default contribution rate.



Source: Madrian and Shea (2001)

7

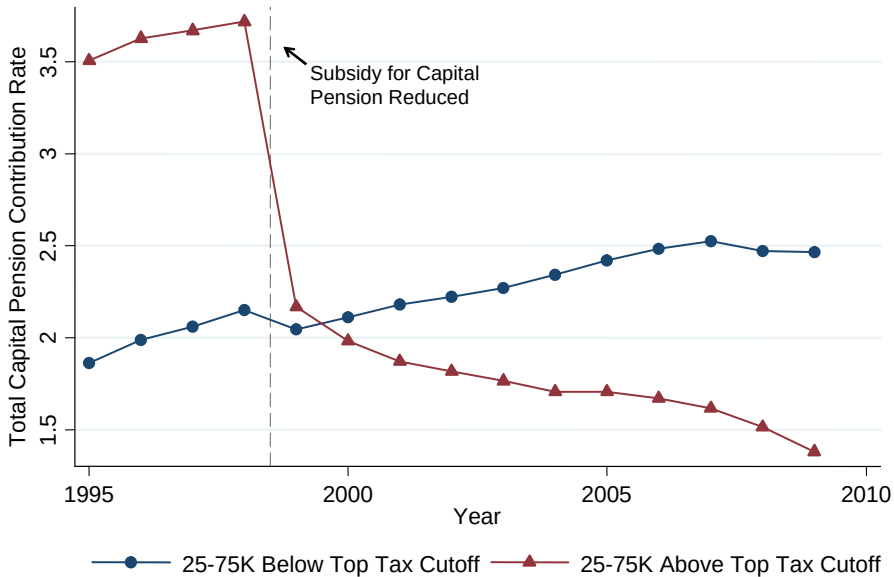
Chetty et al. (2012): Crowdout in Retirement Savings

- Do defaults increase total savings or just lead to shifting of assets from non-retirement to retirement accounts?
 - Even inattentive individuals still have to satisfy budget constraint by cutting consumption or savings in non-retirement accounts
- Do price subsidies raise total savings or induce shifting across accounts?
 - Large literature on “crowdout” in retirement savings accounts (Engen, Gale, Scholz 1996; Poterba, Venti, Wise 1996)
 - Data in U.S. on wealth outside retirement accounts very limited
- Chetty et al. (2012) analyze this question using third-party reported data on all financial wealth for population of Denmark

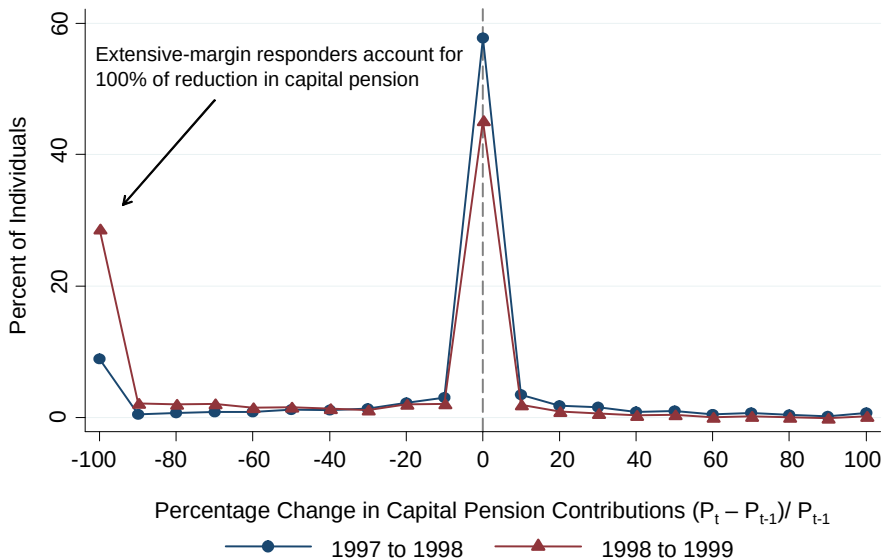
Chetty et al. (2012): Price Subsidies

- Denmark has two types of pension accounts: capital pensions and annuity pensions
- Reform in 1999 in Denmark lowered subsidy for saving in capital pensions by 12 cents per DKr
- Question 1: how did this affect contributions to capital pensions?
- Question 2: how much money was shifted to annuity pensions and to non-retirement taxable accounts?

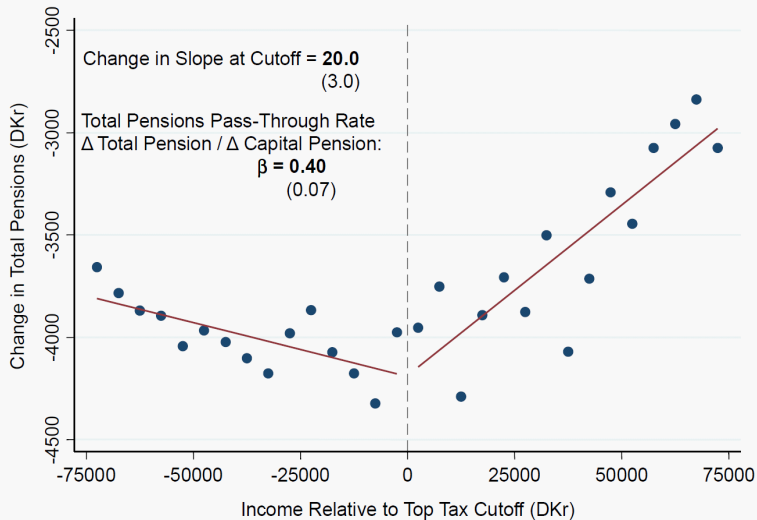
Impact of Capital Pension Subsidy Reduction On Capital Pension Contributions



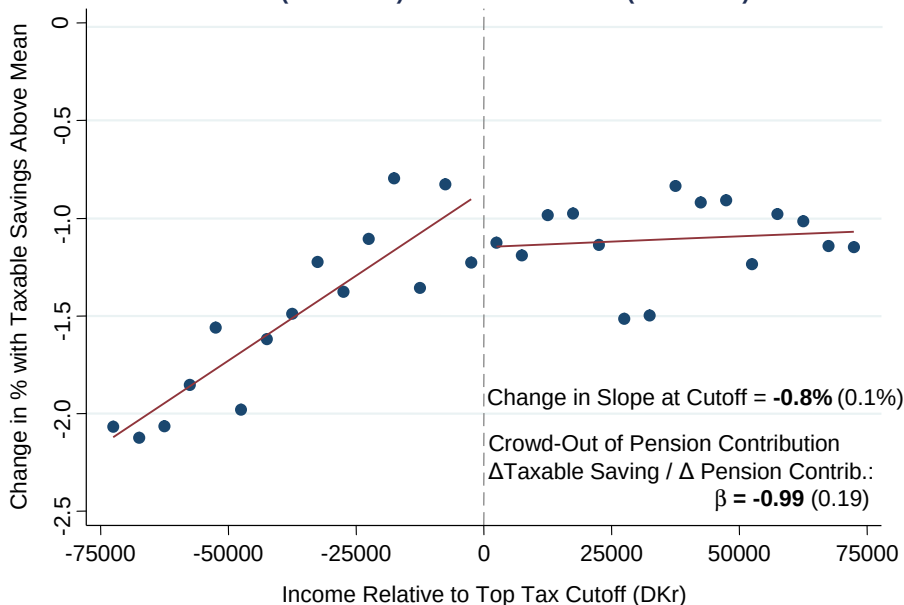
Impact of 1999 Capital Pension Subsidy Reduction on Distribution of Capital Pension Contributions for Prior Contributors



c) Change in Total Pensions Around Subsidy Reduction: Pre-Reform (96-98) Minus Post-Reform (99-01)

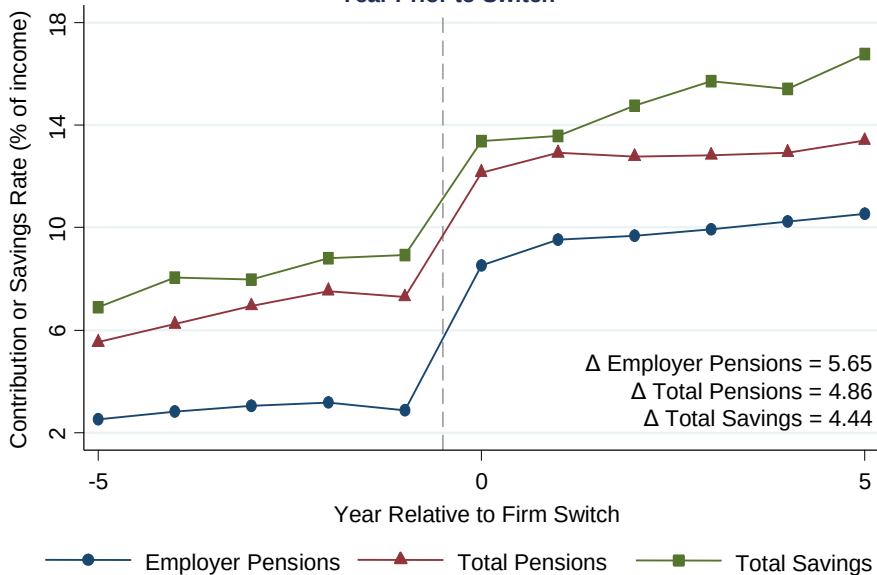


Difference in Fraction of High Savers Before vs. After Subsidy Change,
Pre-Reform (1996-1998) minus Post-Reform (1999-2001)

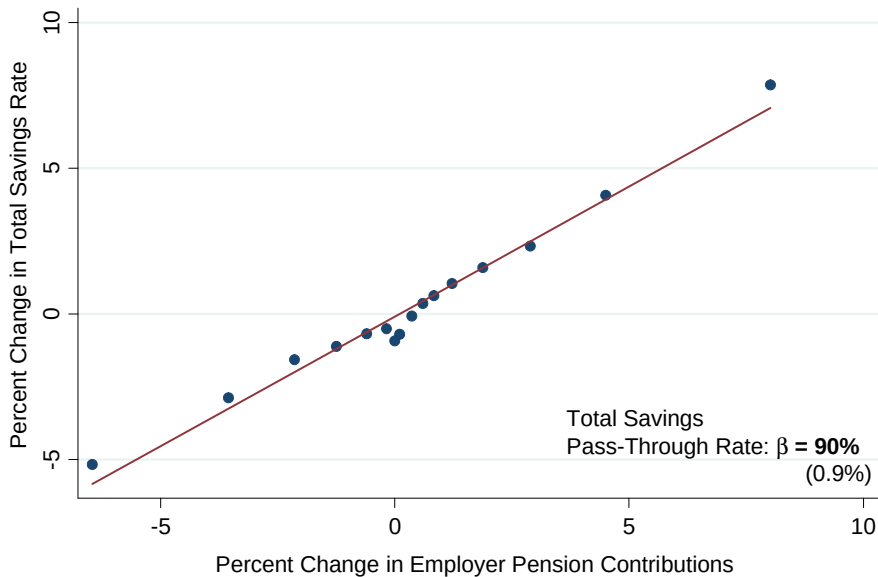


- Next, study impacts of automatic contributions
- Employers make pension contributions on workers behalf automatically
- Research design: event study when workers switch firms
 - Retirement savings rate can change sharply when workers switch firms
 - Do workers offset these changes in private savings as neoclassical model predicts?

Event Study around Switches to Firm with >3% Increase in Employer Pension Contrib.: Switchers with Positive Individual Pensions and Savings in Year Prior to Switch



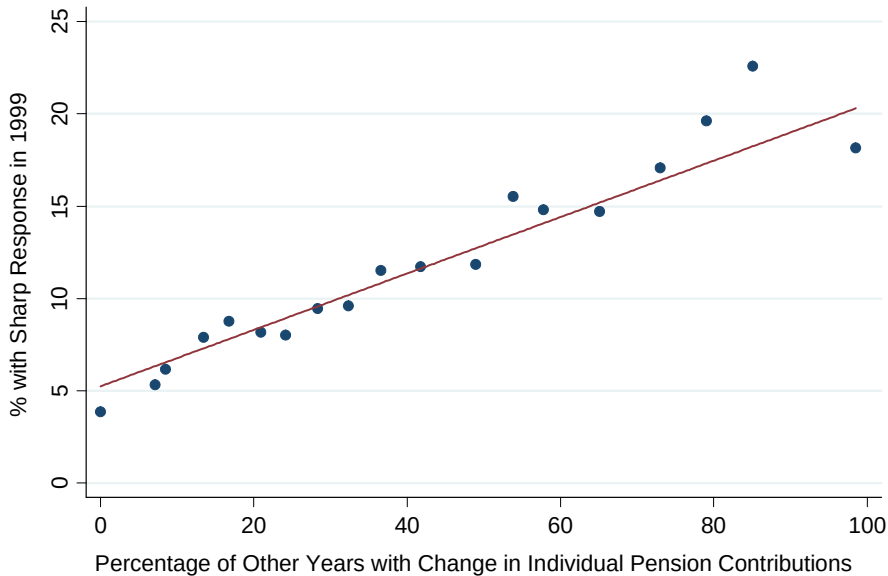
Changes in Total Savings Rates vs. Changes in Employer Pension Rates for Firm Switchers, Cond. on Lagged Savings



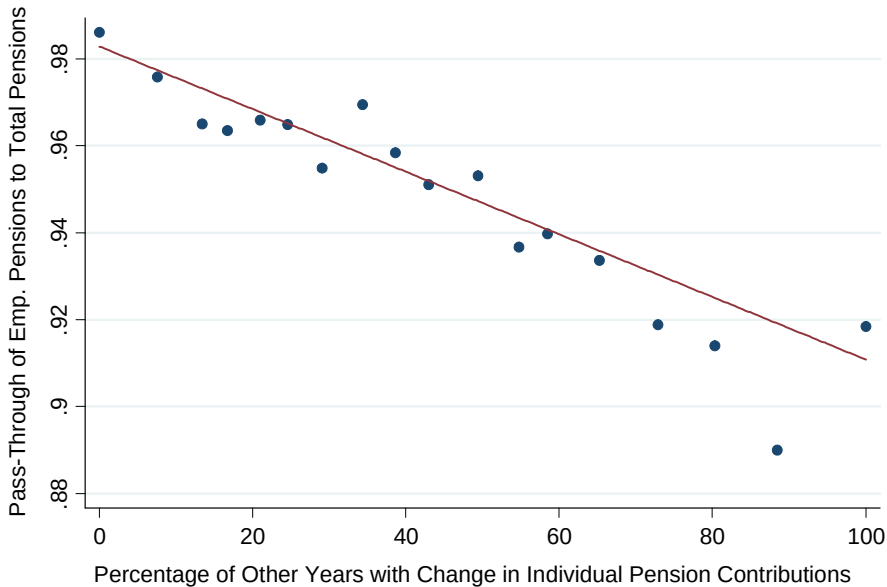
Chetty et al. (2012): Active vs. Passive Savers

- Why do automatic contributions have much larger impacts on total savings than price subsidies?
- Hypothesis: active vs. passive choice, as in Carrol et al. (2009)
- Analyze heterogeneity of responses to test key predictions of this model
 - ① Price subsidies affect active savers
 - ② Automatic contributions affect passive savers
 - ③ Active savers save more for retirement to begin with

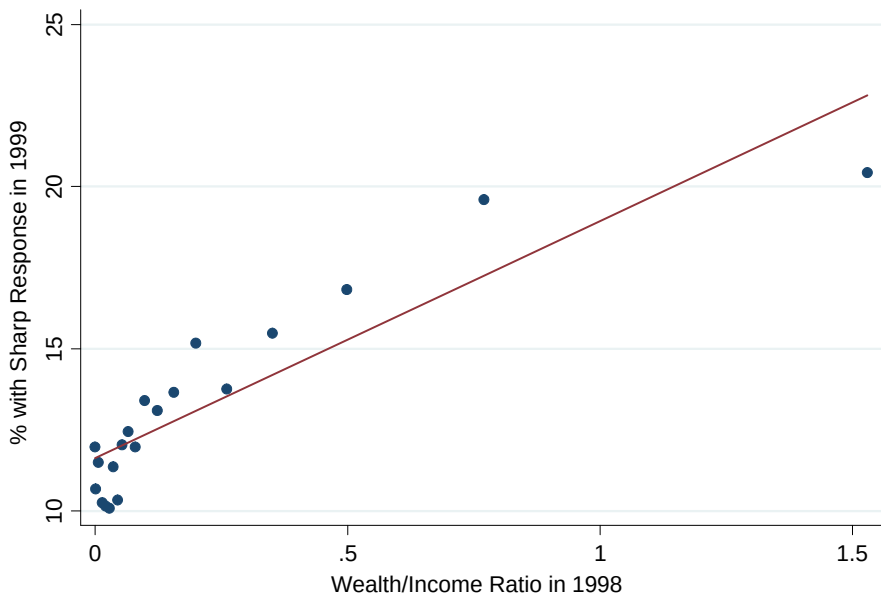
Percent Responding to Capital Pension Subsidy Change in 1999 by Frequency of Active Changes in Other Years



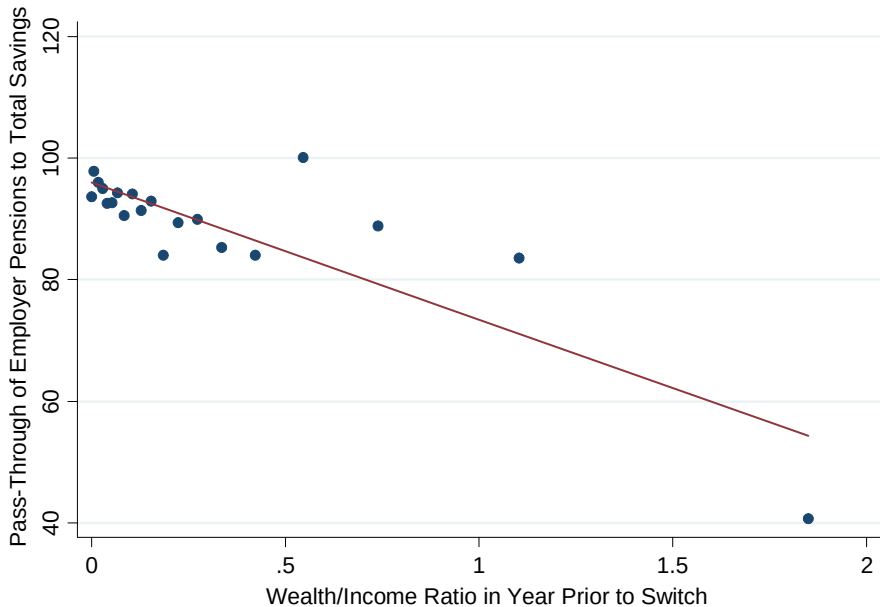
Pensions Pass-Through of Employer Pension Changes for Firm-Switchers by Frequency of Active Changes in Other Years



Heterogeneity in Response to Capital Pension Subsidy by Wealth/Income Ratio



Heterogeneity in Pass-Through of Employer Pensions by Wealth/Income Ratio



Chetty et al. (2012): Correcting “Internalities”

- Tax subsidies tend to influence the behavior of those who are already saving
- Need to be an active saver to pay attention and respond to subsidies
- More general lesson: economic tools (prices) may not be the best way to change the behavior of non-optimizing agents
- Non-traditional tools such as “nudges” may be more effective

Public Goods: Outline

- 1 Definitions and Baseline Model
- 2 Samuelson Rule
- 3 Public Goods with Endogenous Private Provision
- 4 Public Goods with Distortionary Taxation
- 5 Alternative Instruments

Public vs. Private Goods

- **Private goods** benefit one individual h

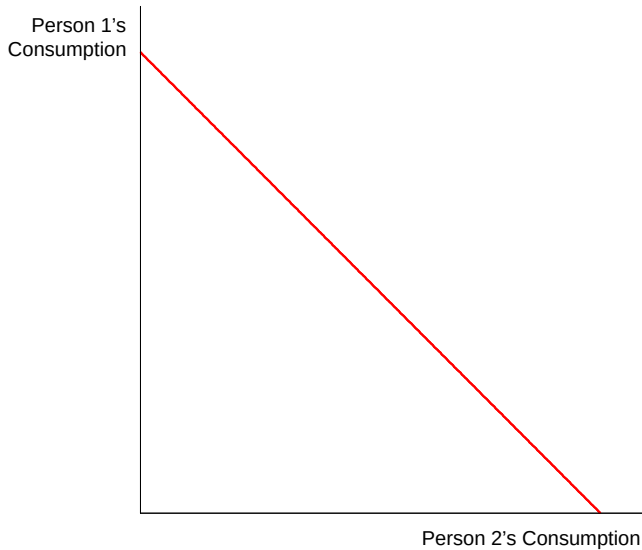
$$\sum_h X_h \leq X$$

- **Public goods** benefit several individuals simultaneously

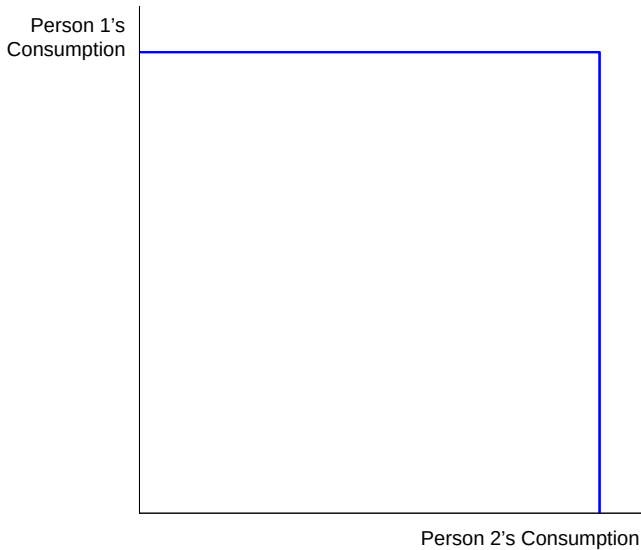
$$X_h \leq X \quad \forall h$$

- 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 $h = 1, \dots, H$
- 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 $U^h = U^h(X^h, G)$

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:

$$\begin{aligned} \max \quad & \sum_h \beta^h U^h(X^h, G^h) \\ \text{s.t.} \quad & F\left(\sum_h X^h, \sum_h G^h\right) \leq 0 \quad [\lambda] \end{aligned}$$

- Lagrangian:

$$L = \sum \beta^h U^h - \lambda F$$

- First order conditions

$$\begin{aligned} [X^h] \quad & : \quad \beta^h U_X^h = \lambda F_X \\ [G^h] \quad & : \quad \beta^h U_G^h = \lambda F_G \end{aligned}$$

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:

$$MRS_{GX}^h = MRT_{GX} \quad \forall h$$

- Decentralized market equilibrium will implement such an allocation (1st Welfare Thm).

First Best if G is a Pure Public Good

- To identify Pareto efficient outcomes, now solve:

$$\begin{aligned} \max \quad & \sum_h \beta^h U^h(X^h, G) \\ \text{s.t.} \quad & F(\sum_h X^h, \sum_h G^h) \leq 0 \quad [\lambda] \end{aligned}$$

- FOC's:

$$\begin{aligned} [X^h] \quad & : \quad \beta^h U_X^h = \lambda F_X \\ [G] \quad & : \quad \sum_h \beta^h U_G^h = \lambda F_G \end{aligned}$$

- Using $\beta^h = \lambda F_X / U_X^h$ from f.o.c. for X^h we obtain:

$$\sum_h \left[\frac{U_G^h}{U_X^h} \right] = \frac{F_G}{F_X}$$

Samuelson (1954) Rule

- Condition for Pareto efficiency: **sum of** MRS is equal to MRT:

$$\sum_h MRS_{GX}^h = MRT_{GX}$$

- 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

Decentralized Private Provision is Suboptimal

- Private good X and a pure public good G as above
- Price of each good is normalized to 1
- 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 is Suboptimal

- Individual h solves

$$\begin{aligned} \max U^h(X^h, G^1 + \dots + G^h + \dots + G^H) \\ \text{s.t. } X^h + G^h = Y^h. \end{aligned}$$

- 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

Optimal Second Best Provision of PG's

- Now suppose government provides public good to rectify under-provision in market equilibrium
- Two complications arise
 - ① Crowdout of private sector provision
 - Private contributions to charity exceed \$250 bn. per year
 - Key model: Bergstrom, Blume, and Varian (1986)
 - ② Government cannot finance PGs through lump sum taxation
 - Need to modify Samuelson rule to account for distortionary taxation?
 - Related to Sandmo (1975) analysis of externalities

Bergstrom, Blume, Varian (1986): Setup

- Individual h solves:

$$\begin{aligned} \max_{X^h, G^h} U^h(X_h, G_h + G_{-h}) \\ \text{s.t. } X_h + G_h = Y_h \end{aligned}$$

- 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:

$$\begin{aligned} \max & U(X^h, G_h + G_{-h} + T) \\ \text{s.t.} & X^h + G^h = Y^h - t^h \end{aligned}$$

Bergstrom-Blume-Varian Model: Crowd-out

- Let $Z_h = G_h + t_h$ denote total contribution of individual h .
- Can rewrite this as:

$$\begin{aligned} \max U(X^h, Z_h + Z_{-h}) \\ \text{s.t. } X^h + Z^h = Y^h \end{aligned}$$

- This is isomorphic to original problem $\Rightarrow Z^* = G^*$
- Total public good provision is unchanged!
- Each person simply reduces voluntary provision by t_h

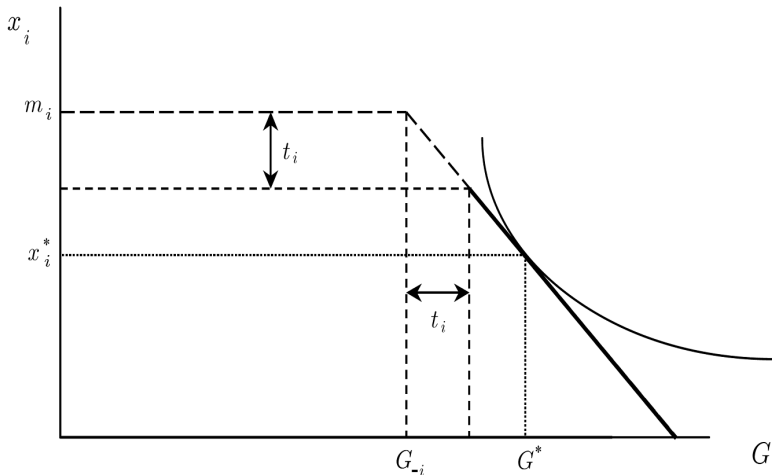


Figure 5: Complete Crowding Out

Source: Andreoni 2006

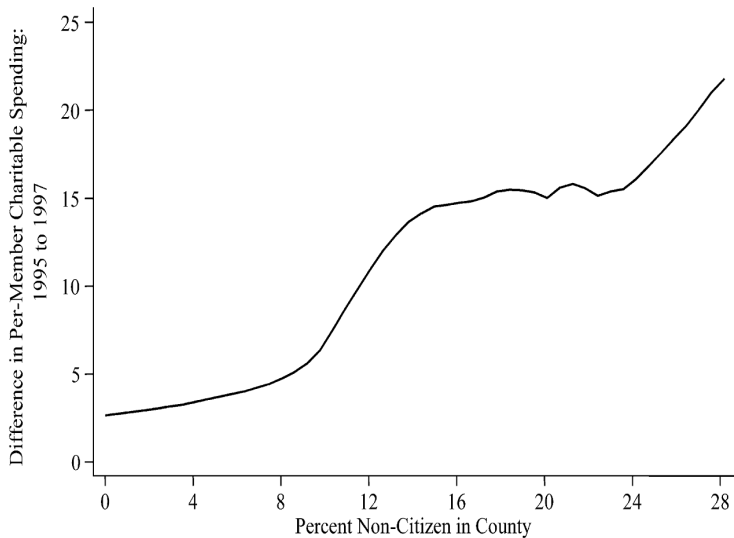
BBV Model: Key Assumptions

- ❶ 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.
- ❷ 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
- ❸ 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

- 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
- Estimates imply that total church expenditures in a state go up by 40 cents when welfare spending is cut by \$1



Source: Hungerman 2005

Andreoni and Payne (2003, 2008): Fundraising Mechanism

- Government spending crowds-out private donations through two channels: willingness to donate + fundraising
- Use tax return data on arts and social service organizations
- Instrument for government spending using changes in state budget due to federal grants
- Key findings:
 - \$1000 increase in government grant leads to \$250 reduction in private fundraising
 - \$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

- 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

- 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?”
- Andreoni (1988, 1993) implements experiments with repeated contributions
 - Shows that contributions to public goods fall over time but remain positive

Expanding the Policy Set: Social Prices

- Traditional public goods and externalities literature focuses on economic incentives
- Induce public goods provision by changing relative prices of goods
- Another potential policy tool: manipulation of social prices
 - Exploit concerns for perception by peers to encourage pro-social behavior
 - E.g. have researchers compete on publications and help society in the process

Social Pressure: Existing Evidence

- Recent examples from psychology and political science
 - Cialdini (2003) on energy conservation: telling people how their energy use compares with averages reduces energy use
 - Gerber et al. (2008): using social pressure to increase voter turnout

Neighbors mailing

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have votes in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not

DO YOUR CIVIC DUTY – VOTE!

	Aug 04	Nov 04	Aug 06
MAPLE DR			
9995 JOSEPH JAMES SMITH	VOTED	VOTED	_____
9995 JENNIFER KAY SMITH	VOTED		_____
9997 RICHARD B JACKSON	VOTED		_____
9999 KATHY MARIE JACKSON		VOTED	_____
9987 MARIA S. JOHNSON	VOTED	VOTED	_____
9987 TOM JACK JOHNSON	VOTED	VOTED	_____

Source: Gerber, Green, and Larimer (2008)

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

Source: Gerber, Green, and Larimer (2008)

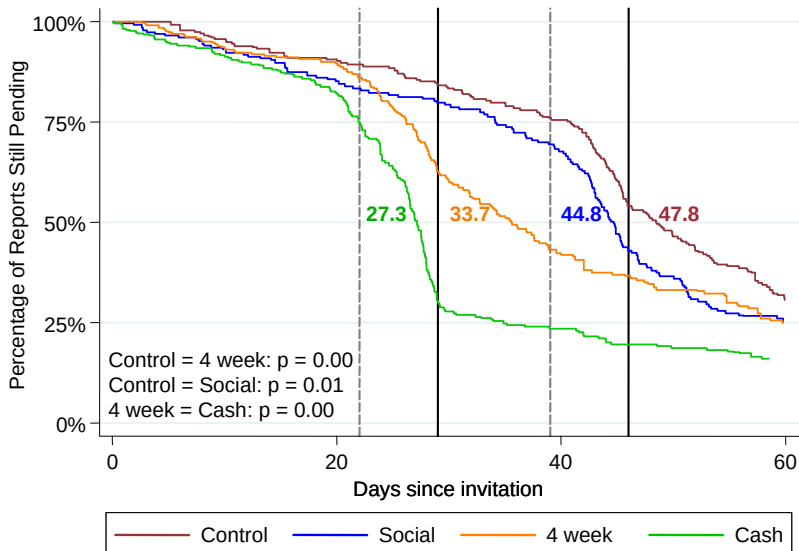
Social Prices as a Policy Tool: Research in Progress

- ① [Perez-Truglia and Cruces 2012] Social incentives in campaign contributions and political polarization
 - ② [Chetty, Saez, Sandor 2012] Comparing cash and social incentives to reduce referee times at JPubE
 - ③ [Chetty, Mobarak, Singhal] Increasing tax revenue in Bangladesh using social incentives
- Theoretical question: optimal social prices and policy design

		Referee	Manuscript(s)	
	Name	Affiliation	Reviewed	Avg Days
1	Acharya, Viral V.	New York University	1	28
2	Aggarwal, Reena	Georgetown University	2	29
3	Ait-Sahalia, Yacine	Princeton University	2	30
4	Albuquerque, Rui	Boston University	3	31
5	Almeida, Heitor	New York University	1	28
	⋮			
9	Asquith, Paul	MIT	1	0
254	Scharfstein, David	Harvard Business School	1	173
255	Scheinkman, Jose A.	Princeton University	1	308
257	Schneider, Martin	Stanford University	2	200

Source: *Journal of Financial Economics* Ad Hoc Referees, <<http://jfe.rochester.edu/colab.htm>>

Figure 2b: Turnaround times, DFL reweighted on pre-experiment turnaround



Source: Chetty, Saez, Sandor 2010

Public Economics Lectures

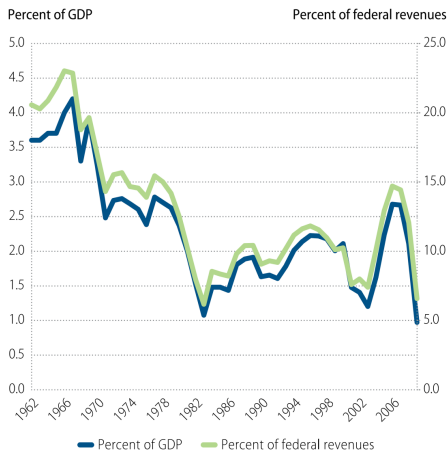
Part 5: Corporate Taxation

Raj Chetty

Harvard University
Fall 2013

FIGURE 1

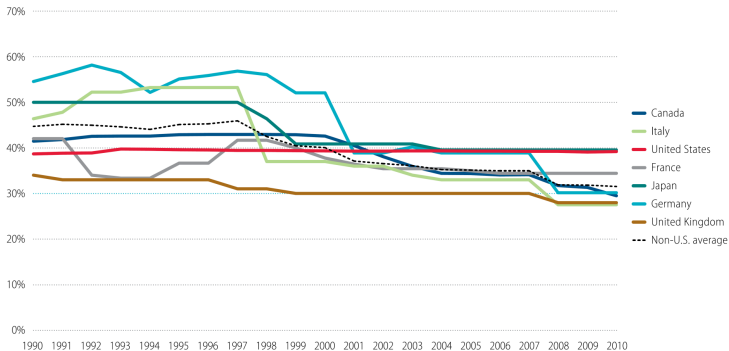
Corporate tax revenues, percent of GDP and of federal revenues



Source: Auerbach (2010)

FIGURE 3

G-7 corporate tax rates, 1990-2010

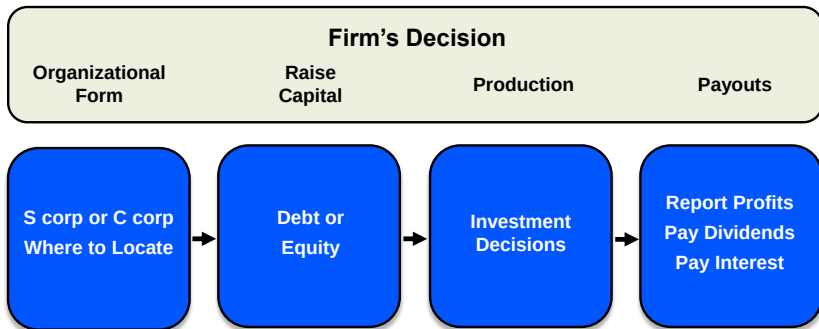


Source: Auerbach (2010)

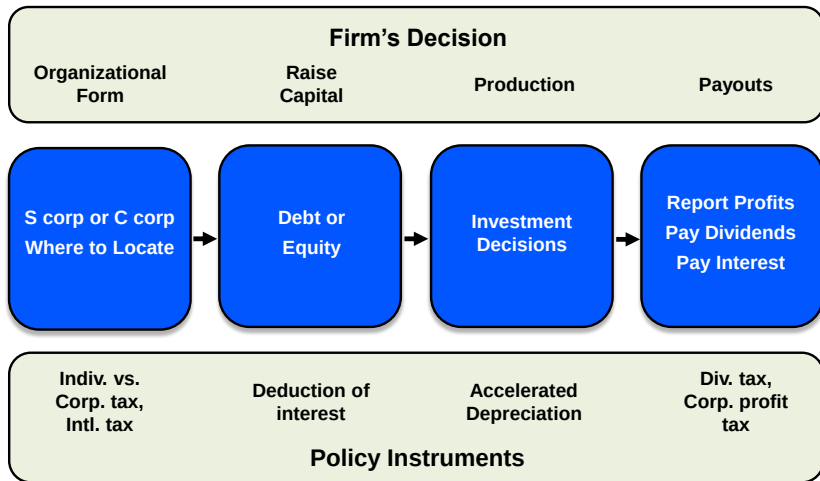
Corporate Tax Structure in the U.S.

- Taxes on firms in the U.S. consist of several elements
 - ① Individual-level taxes on payouts (capital gains, dividends, interest income)
 - ② Tax on corporate profits (earnings minus expenses) at flat rate of 34%
 - Expenses include wages+materials, depreciation, and interest payments
 - Acceleration of depreciation used to stimulate investment
 - ③ Firms that are not publicly traded (“S corporations”) subject to individual income tax system
 - ④ International tax provisions (transfer pricing, tax havens)
- Goal: characterize consequences of this tax system and optimal design of corporate taxation

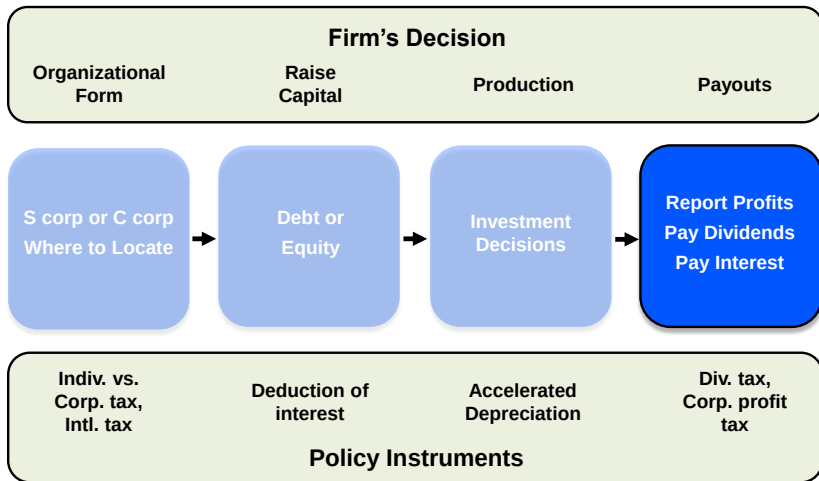
Corporate Decisions and Tax Policies



Corporate Decisions and Tax Policies



Corporate Decisions and Tax Policies



Neoclassical Model of Firm Behavior

- Structure analysis using stylized two period model of firm behavior (Chetty and Saez 2010)
- Results generalize to continuous time model (Auerbach 2001)
- Firm has cash holdings of X at $t = 0$
- Can raise more funds by issuing equity E , so total cash is: $X + E$
- Chooses level of investment, I , with concave payoff $F(I)$ at $t = 1$

Neoclassical Model of Firm Behavior

- Pays out remaining cash as a dividend in period 0 :

$$D = X + E - I$$

- Rule out share repurchases for now, return to this below
- Tax τ_d levied on dividend payments in all periods
- Tax τ_c on corporate profits
- Investors can also purchase a govt. bond that pays fixed rate r

Neoclassical Model: Manager's Objective

- Manager maximizes value of the firm:

$$\max_{E,D} V = (1-\tau_d)D - E + \frac{(1-\tau_d)[(1-\tau_c)f(X+E-D) + X - D] + E}{1+r}$$

where $f(I) = F(I) - I$ denotes net profit from investing I

- No tax benchmark: invest up to point where $f'(I) = r$
- To characterize behavior with taxes, divide firms into two types:
 - 1. Cash-Rich [new view]: X s.t. $(1-\tau_c)f'(X) < r$
 - 2. Cash-Constrained [old view]: X s.t. $(1-\tau_c)f'(X) > r$

Cash-Rich Firms: “New View”

- Marginal value of issuing equity is negative for cash rich firm (e.g., Microsoft)
 - Even pre-tax return on investment is below interest rate
- Therefore $E = 0$ and firm splits cash between D and I according to:

$$(1 - \tau_c)f'(X - D) = r$$

- Invest to point where after-tax marginal product $(1 - \tau_c)f'(I)$ equals bond return r
- Higher corporate tax rate lowers investment

Cash-Rich Firms: “New View”

- Change in dividend tax rate has no effect on dividend or investment behavior (Auerbach 1979, Bradford 1981, King 1977)
 - τ_d factors out of V b/c investment financed from retained earnings
- \$1 of investment + dividend tomorrow yields $(1 - \tau_d)(1 - \tau_c)f'(I)$
- \$1 of dividend yields $(1 - \tau_d)$ today
- Relative price of investment tomorrow vs. today unaffected by τ_d

Cash-Constrained Firms: “Old View”

- Marginal value of paying dividends is negative for cash-constrained firm (e.g., Twitter)
 - Pre-tax return on investment is above interest rate r
- Therefore $D = 0$ and $I = X + E$. Optimal equity issue E satisfies:

$$(1 - \tau_d)(1 - \tau_c)f'(X + E) = r$$

- Invest to point where marginal net-of-tax return $(1 - \tau_d)(1 - \tau_c)f'(I)$ equals interest rate

Cash-Constrained Firms: “Old View”

- Key result: E , I fall with both τ_d and τ_c
- Dividend tax cuts stimulate equity issues and investment, and dividend payout in period 1 (Poterba and Summers 1985)
- τ_d does not factor out of value function because marginal investment is financed from external capital

Efficiency Analysis: Neoclassical Model

- Denote total dividend payout over 2 years by

$$P_d = D + [(1 - \tau_c)f(I) + X - D]/(1 + r)$$

- Total surplus in the economy is firm value plus tax revenue is

$$W = V + \tau_d P_d$$

- Using envelope theorem (manager optimization), deadweight cost is

$$\frac{dW}{d\tau_d} = -P_d + P_d + \tau_d \frac{dP_d}{d\tau_d} = -\frac{\tau_D}{1 - \tau_D} \cdot \epsilon_{P_d} \cdot P_d$$

Efficiency Analysis: Neoclassical Model

- Under old view, $\epsilon_{P_d} > 0$, so dividend taxes reduce efficiency
- New view: Dividend tax has no efficiency cost and simply takes money from wealthy shareholders, which may be desirable for redistribution
- Old view and new view concur that taxes on corporate profits are distortionary
- Sinn (1991): lifecycle view. Old view applies to young firms (entrants) while new view applies to mature firms
- Distinguishing between competing views important for policy

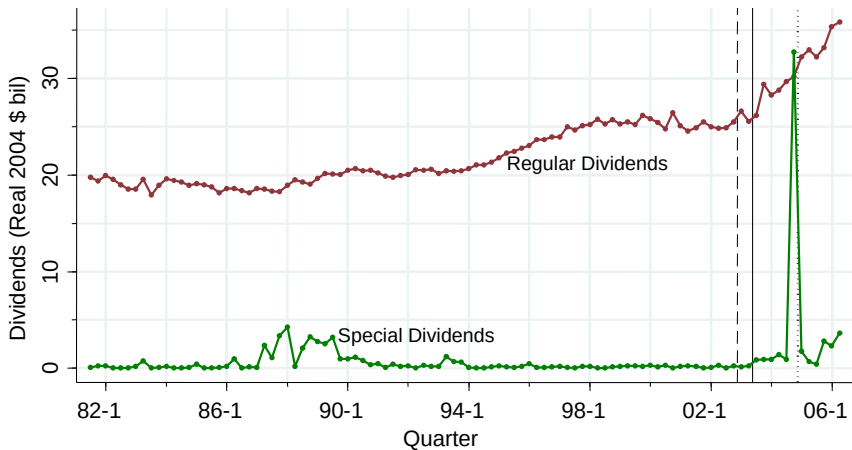
Empirical Evidence on Dividend Taxation

- Several studies examine effects of tax cuts on dividend payouts to test between old view and new view
 - Poterba and Summers (1985) find a positive link between div payout ratio and $(1 - \tau_d)$ in U.K. time series
 - Poterba (2004) reports time series evidence in the U.S.
 - Mixed results from studies of Tax Reform Act of 1986

Evidence from 2003 Dividend Tax Cut

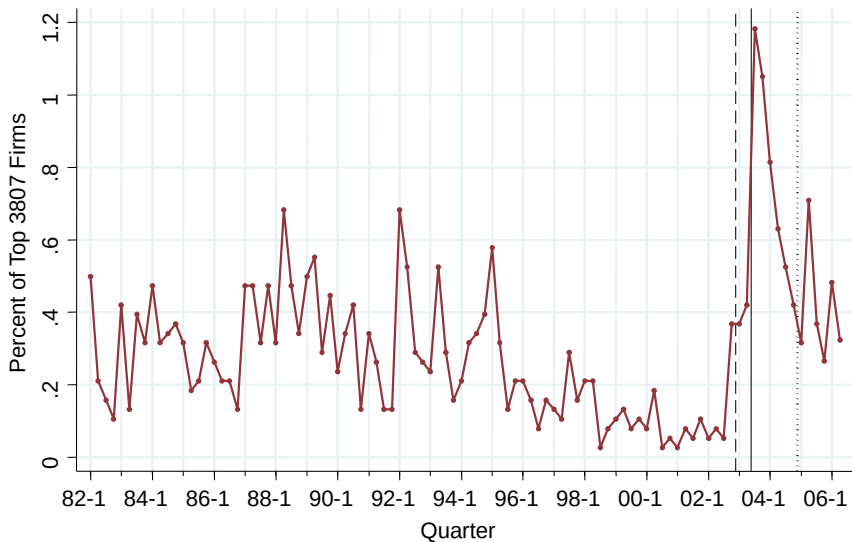
- Chetty and Saez (2005) use a quasi-experimental research design
- Study effect of the 2003 dividend tax cut on initiations of and overall levels of dividend payments
 - 2003 tax cut reduced tax rates on dividend from normal income tax rates to 15% flat rate
 - Simple diff-in-diff design; results directly visible in raw time series
- Key challenge: difficult to analyze mean effects due to outliers (typical problem with firms)

Dividend Payments: Aggregate Time Series

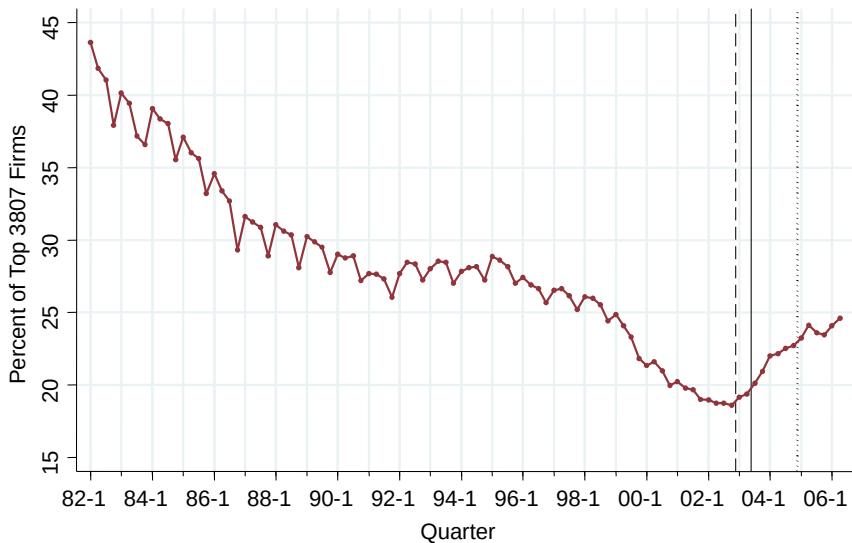


Source: Chetty and Saez (2005), using data through 2006Q2.

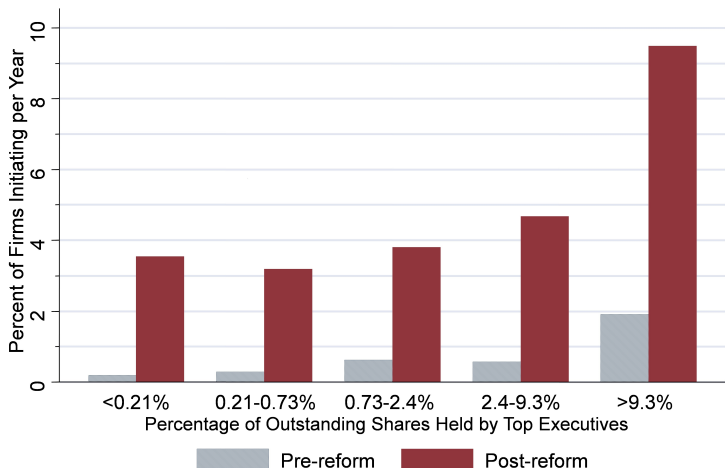
Regular Dividend Initiation Time Series



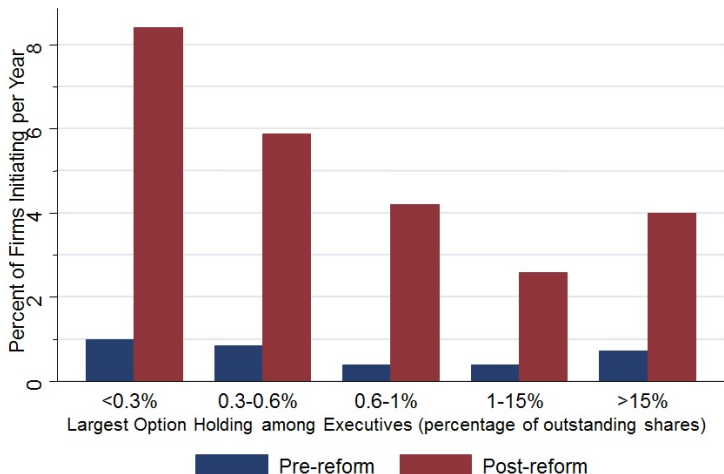
Fraction of Dividend Payers



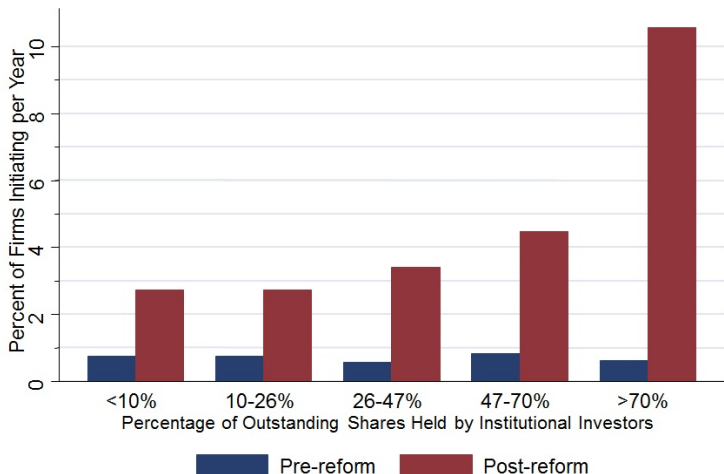
Effect of Tax Cut on Initiations by Executive Shareholding



Effect of Tax Cut on Initiations by Executive Option Holdings

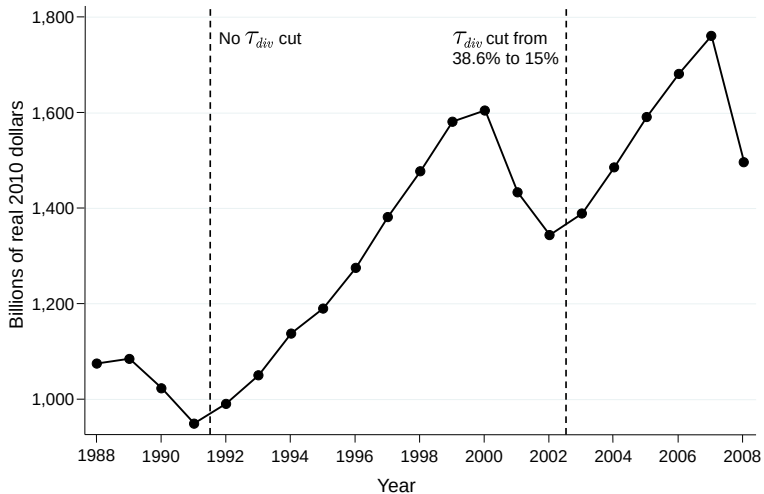


Effect of Tax Cut on Initiations by Institutional Ownership



- Yagan (2013) analyzes effect of dividend tax cut on real outcomes: investment and employee compensation
- If the dividend tax cut reduces the cost of capital as in the old view, we should expect investment to increase
- Research design: diff-in-diff using unaffected S corporations as control

U.S. Corporate Investment in National Accounts

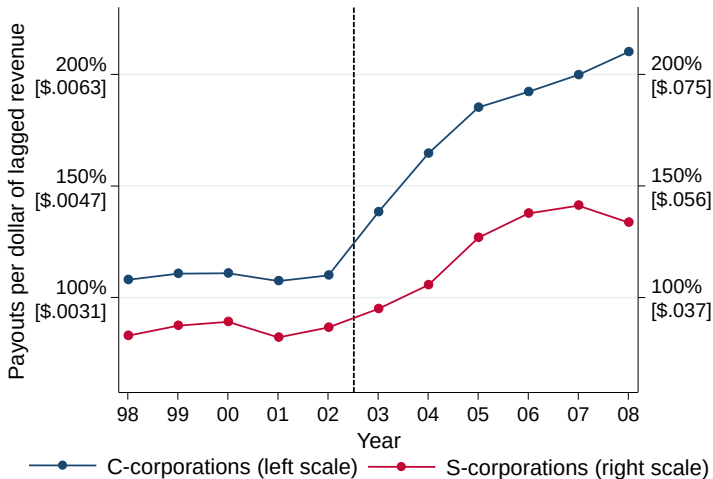


- After incorporating, a corporation elects either C or S tax status:

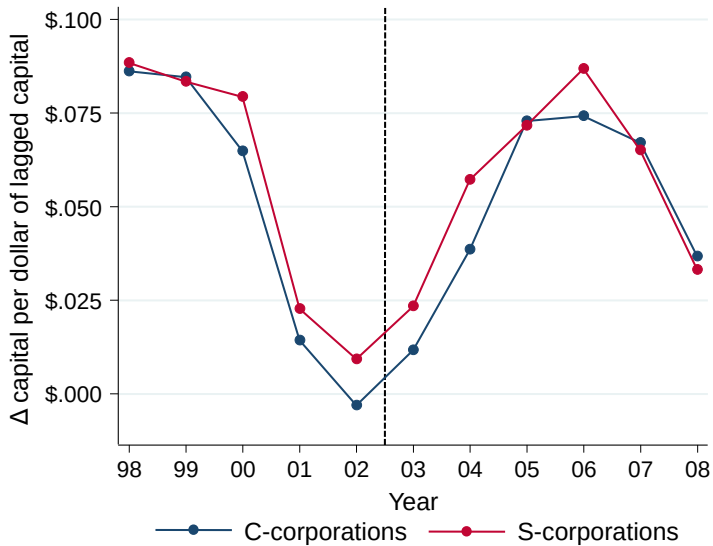
	Tax rate on annual profits	Tax rate on dividends
C-corporations	35%	Cut in 2003
S-corporations	35%	0%

- S-corporations: < 100 non-institutional shareholders, one stock class
- Compete in the same narrow industries and at the same scale throughout the United States (common trends)
- Ex: Home-Depot is a C corp, Menards is an S corp

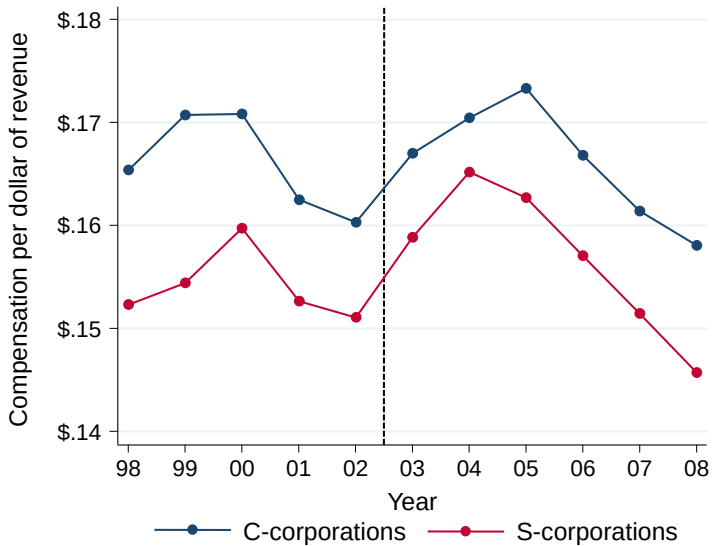
Payouts to Shareholders



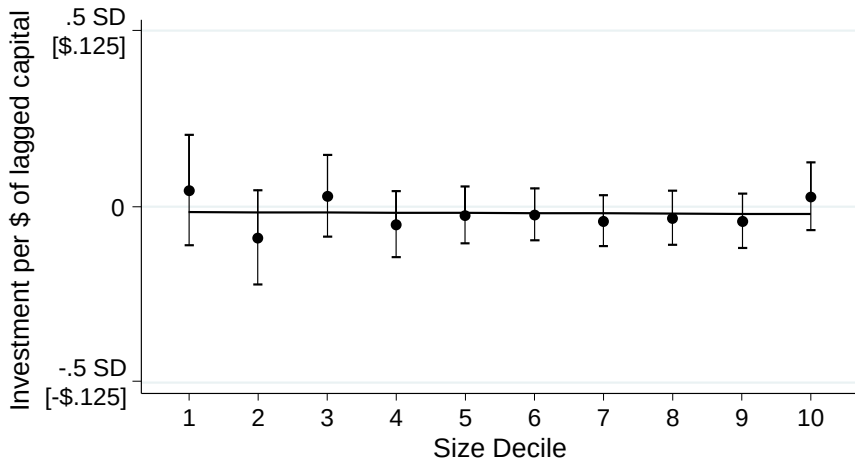
Net Investment



Employee Compensation



Investment Response by Firm Size Decile



Corporate Taxation in an Agency Model

- Chetty and Saez (2010) develop an agency model of corporate taxation consistent with the findings in Chetty and Saez (2005)
- Key idea in corporate finance models of firm behavior: divergence between the interests of the owners and managers of a firm
 - Shareholders: want firm to maximize profits
 - But CEO may get utility from running a big company, flying in private jets, or investing in pet projects

Corporate Taxation in an Agency Model

- To model agency problem, generalize two period model above
- Firm is financed from retaining earnings (new view)
- Now CEO has three options instead of two:
 - 1 Productive investment: I
 - 2 Pet project investment: J
 - 3 Dividends: $D = X - I - J$
- CEO owns a fraction $\alpha < 1$ of the shares of the company

Corporate Taxation in an Agency Model

- Shareholder's payout = firm's profits:

$$\pi = (1 - \tau_d) \left(D + \frac{(1 - \tau_c)f(I) + X - D}{1 + r} \right)$$

- CEO chooses I and J to maximize his personal payoff

$$\max_{I,J} \alpha\pi + \frac{g(J)}{1 + r}$$

Corporate Taxation in an Agency Model

- Optimality conditions in interior:

$$(1 - \tau_c)f'(I) = r$$
$$g'(J) = \alpha(1 - \tau_d)r$$

- Comparative statics:
 - Increase in τ_d raises J , leaves I unaffected, and lowers D
 - Increase in τ_c lowers I leaves J unaffected

Efficiency Cost of Dividend Tax in Agency Model

- Social welfare is

$$\begin{aligned} W &= \pi + \frac{g(J)}{1+r} + \tau_d P_d \\ &= \left\{ \alpha(1 - \tau_d)P_d + \frac{g(J)}{1+r} \right\} + (1 - \alpha)(1 - \tau_d)P_d + \tau_d P_d \end{aligned}$$

- Using envelope theorem for term in curly brackets (objective maximized by private sector),

$$\begin{aligned} \frac{dW}{d\tau_d} &= -\alpha P_d - (1 - \alpha)P_d + (1 - \alpha)(1 - \tau_d) \frac{dP_d}{d\tau_d} + P_d + \tau_d \frac{dP_d}{d\tau_d} \\ &= -\left(\frac{\tau_D}{1 - \tau_D} + (1 - \alpha) \right) \cdot \epsilon_{P_d} \cdot P_d \end{aligned}$$

- Main result: dividend tax has a first-order efficiency cost

Efficiency Cost of Dividend Tax in Agency Model

- Analogous derivation shows that corporate tax has standard second-order deadweight cost

$$\frac{dW}{d\tau_c} = \tau_c \frac{dP_c}{d\tau_c}$$

- Dividend tax has higher than DWL from corporate tax
 - Contrary to prediction of both old and new view

Efficiency Cost of Dividend Tax in Agency Model

- Why does a small dividend tax have a first-order efficiency cost in agency model?
- Intuition: Dividend tax makes Microsoft over-invest relative to Twitter
- Dividend tax magnifies an existing inefficiency even in new view (retained earnings) case
- Distorts allocation of investment even if leaving aggregate level of investment $(I + J)$ unchanged
 - Is this true empirically?

Dividends vs. Share Repurchases

- Preceding models assume that firms can only repay money through dividends
- But firms can also return money to shareholders by buying back shares
- Repurchases are taxed at lower capital gains rate rather than dividends
- Why pay dividends given tax disadvantage?

Explanations for Dividends

- ① Agency model (Jensen 1986): firms have free cash flow that they would otherwise waste on perks
 - Dividends are a commitment (why?) and therefore tie firms hands and reduce agency costs
- ② Signaling model (Bernheim 1994): taxation of dividends makes them a way to “burn money”
 - Those who have a lower marginal cost of burning money will do more of it in equilibrium
 - The firms with lower marginal costs are the more profitable ones, and so dividends signal profitability

- Tax-based test of agency vs. signalling models of dividends
- Dividend taxes should affect announcement premium in opposite directions in agency and signalling theory
 - In agency theory, higher taxes mean agents get less per dollar paid out $\implies$ lower premium
 - In signalling theory, higher taxes mean higher signalling cost $\implies$ higher premium.
- Use data from 1962-86 and find that premia are higher in higher tax regimes using time series regressions, supporting signalling model

Effective Tax Rate on Dividends

- What is the effective tax rate on dividends with variable tax rates?
 - Dividend tax rates (used to) vary by individuals' tax bracket
 - And non-taxable institutions such as pensions are completely exempt
- Miller (1977) predicts that non-taxable shareholders hold dividend paying stocks at the time of payment, so effective tax rate is 0
- We know this does not occur: dividend taxes generate revenue
 - Could be because of transaction costs
- What is the effective tax rate on dividend payments?

Ex-Dividend Day Premium

- Elton and Gruber (1970) propose a method to back out implied tax rate based on price change around ex-dividend day
 - Ex-dividend day which is the day when ownership rights to dividend payout are determined
- Profit from selling just before the ex-day should equal the profit from selling on the ex-day to eliminate arbitrage

Ex-Dividend Day Premium

- With taxes the no-arbitrage condition is:

$$P_A = P_B + D(1 - \tau_d) \quad (1)$$

P_A is the stock price cum-dividend (just before the ex-dividend day starts)

P_B is the expected stock price on the ex-day

D is dividend amount per share

τ_d is the tax rate on dividend income

Ex-Dividend Day Premium

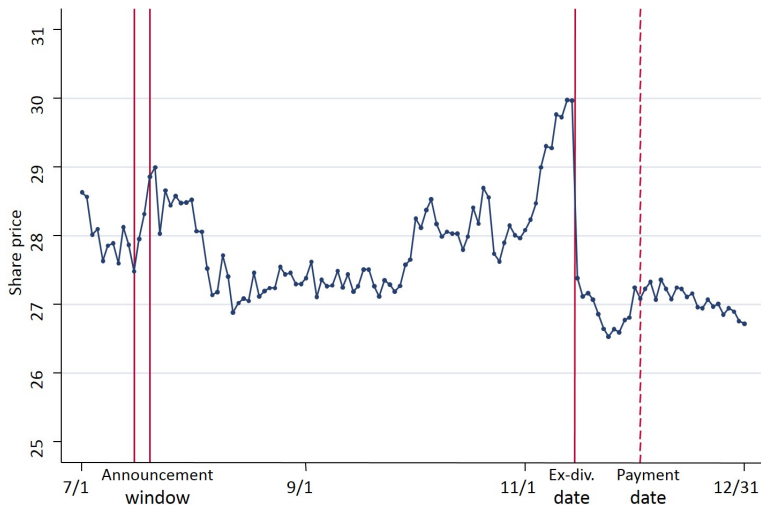
- Rearranging 1, we get:

$$\frac{P_A - P_B}{D} = 1 - \tau_d \equiv \rho$$

- Left-hand-side of this expression = **ex-day premium**
- Right-hand-side = effective net-of-tax rate on dividends for marginal investor

- Examine announcement premia and ex-day premia around 2003 tax cut and in the time series since 1962
- Do these premia vary as predicted by models of taxation and corporate finance?

Microsoft Dividend Announcement and Ex-Day Price Effects, 2004



Ex-Dividend Day Premium and Tax Ratio, 1963-2004



- Prior empirical results in ex-day and announcement effect literature are very fragile and based on tenuous parametric assumptions
- Stock returns are far too volatile to draw reliable conclusions about tax effects
- Need to understand determinants of these excess returns before being able to understand tax effects

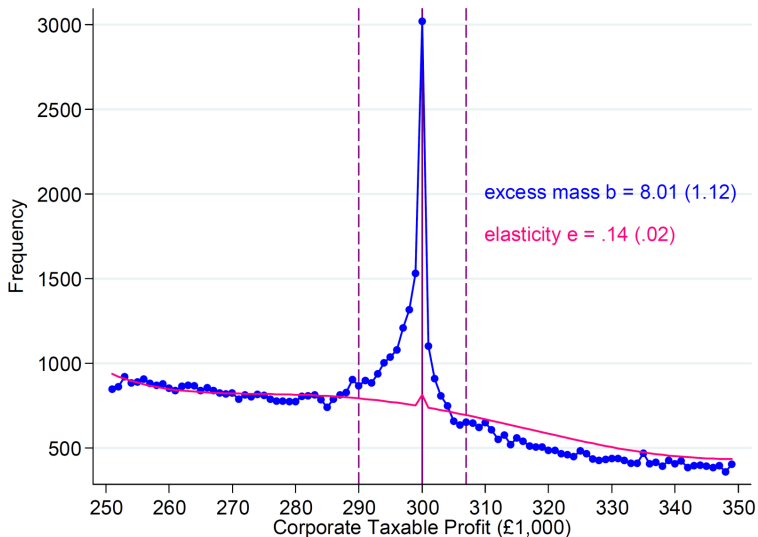
Empirical Evidence on Corporate Taxation

- Next, turn to effects of changes in corporate tax rates on investment and profits
- Here, all existing models predict that higher corporate taxes should reduce investment
- What is the magnitude of this effect?
- Traditional literature used time series regressions with changes in corporate tax rate
- Limited quasi-experimental evidence and problems that closely parallel labor supply literature

Corporate Taxable Income Elasticity

- Devereux, Liu, and Loretz (2013) use corporate tax records in the U.K. to study impacts of taxes on reported profits
- Marginal corporate tax rate jumps from 20% to 32% at 300,000 pounds
- Study bunching at kink to estimate the elasticity of corporate taxable income with respect to the tax rate (same method as Saez 2010, Chetty et al. 2011)

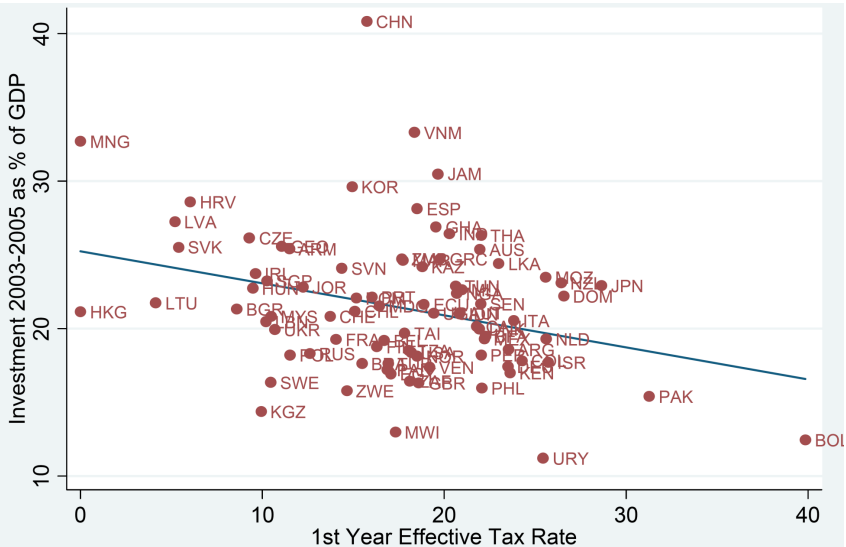
Bunching at the £300k Tax Kink: 2002-2006



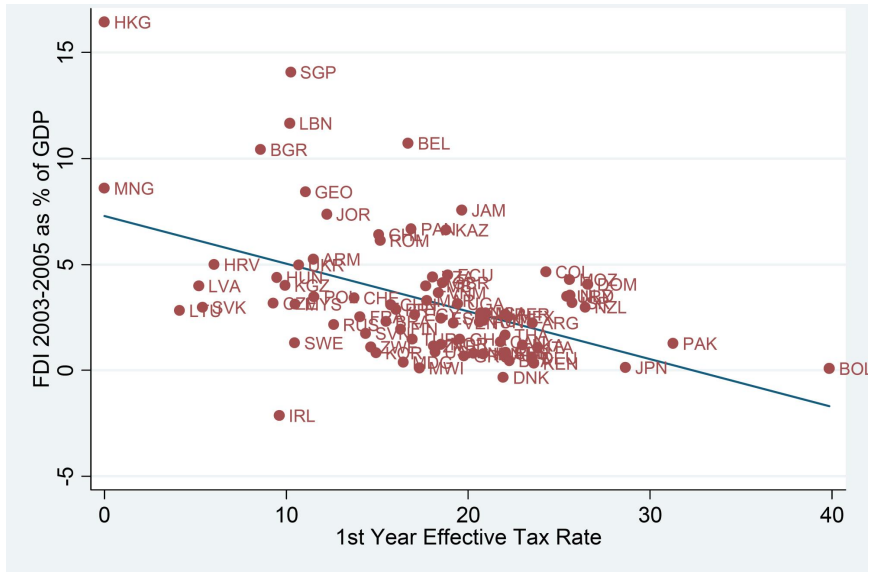
Cross-Country Comparisons

- Djankov et al. (2010) conduct a cross-country analysis of the impacts of effective corporate tax rate
- Measure corporate tax rate for an identical mid-sized firm using a survey conducted with PriceWaterhouseCoopers
- OLS regressions of investment and entrepreneurial activity on corporate tax rate
- Attempt to isolate causal effect using various controls for observables (e.g., other tax rates, levels of economic development)
- Implies substantial effect of taxes on investment: 10 pp increase in corp tax rate leads to 2 pp decrease in I/GDP

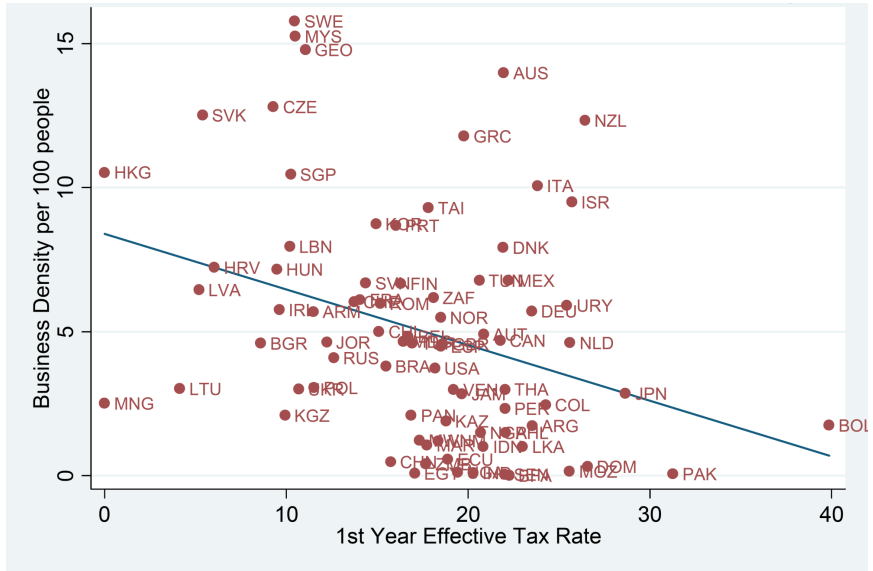
Effective Tax Rate and Investment



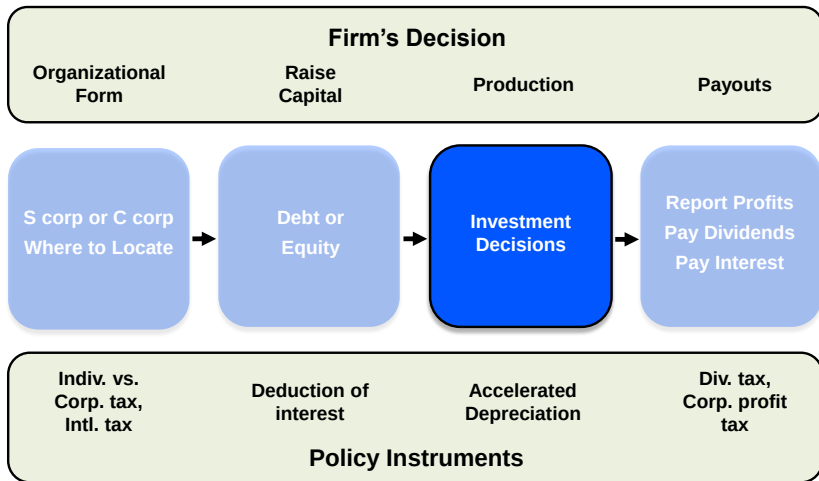
Effective Tax Rate and Foreign Direct Investment



Effective Tax Rate and Business Density



Corporate Decisions and Tax Policies



Tax Incentives for Investment: Accelerated Depreciation

- Most common policies to directly change level of investment: changes in depreciation rules and tax credits for investment
- Frequently used in recessions to attempt to stimulate investment by firms
- Begin with a simple example to understand why depreciation rules matter
 - Suppose a firm buys a machine for \$1000, which wears down by \$100 a year

Tax Incentives for Investment: Accelerated Depreciation

- Consider two tax treatments of the machine
 - ① Expensing: subtract the full \$1000 from profits in the year you buy machine
 - ② Economic depreciation: subtract \$100 per year from your profits
- Expensing reduces effective tax rate for firm given interest rate $r > 0$
- Current policy in U.S.: approximate economic depreciation using linear or exponential rules by asset class

Recovery Periods and Depreciations Methods by Type of Capital

Type of capital	Recovery period, R (years)	Tax depreciation rate, δ (percent)	Method
Tractor units for over-the-road use, horses over 12 years of age or racehorses with over 2 years in service	3	66.7	200 DB
Computers and office equipment; light vehicles, buses and trucks	5	40.0	200 DB
Miscellaneous equipment, office furniture, agricultural equipment	7	28.6 or 21.4	200 DB or 150 DB
Water transportation equipment (vessels and barges); single-purpose agricultural structures	10	20.0 or 15.0	200 DB or 150 DB
Radio towers, cable lines, pipelines, electricity generation and distribution systems, "land improvements," e.g., sidewalks, roads, canals, drainage systems, sewers, docks, bridges, engines and turbines	15	10.0	150 DB
Farm buildings (other than single purpose structures), railroad structures, telephone communications, electric utilities, water utilities structures including dams, and canals	20	7.5	150 DB
Nonresidential real property (office buildings, storehouses, warehouses, etc.)	39	2.6	SL

Note: Tax depreciation methods are 200 percent declining balance (200 DB), 150 percent declining balance (150 DB), and straight line (SL).

Effects of Depreciation Rules on Investment

- How do depreciation rules affect firm behavior?
- Let z = fraction of the machine's cost that can be subtracted as an up-front expense
- Use two-period model with following simplifications
 - ① Investment is financed purely from retained earnings (new view model)
 - ② Dividend tax rate $\tau_d = 0$
 - ③ All profits are returned as dividends in period 2
- Let π_0 denote profits earned in period 0 from previous activity, so that $X = (1 - \tau_c)\pi_0$ denotes cash-on-hand

Effects of Depreciation Rules on Investment

- Firm's problem

$$\max_I V = (1 - \tau_c)\pi_0 - I + \tau_c z I + \frac{[(1 - \tau_c)(f(I) + I) + \tau_c(1 - z)I]}{1 + r}$$

- First order condition for investment:

$$f'(I) = c = r \frac{(1 - \tau_c z)}{(1 - \tau_c)}$$

- Right hand side is “user cost of capital” (Hall and Jorgenson 1967)
- User cost c is sufficient to determine investment rule
 - No need to separately identify effect of r , τ_c , z , etc.

Expensing and the Corporate Cash-Flow Tax

- No expensing ($z = 0$): corresponds to baseline case analyzed above
- Full expensing ($z = 1$): called “corporate cash-flow tax” and is completely non-distortionary in simple model
 - Investment subsidized at the same rate on the margin as profits
 - Firm only taxed on rents earned above “normal” rate of return r
- Strong push for “fundamental” reform to corporate cash-flow tax (e.g. Auerbach 2010)
 - May generate some distortions with other non-deductible inputs (e.g. entrepreneur’s labor), agency problems, and scope for avoidance, but likely to improve efficiency

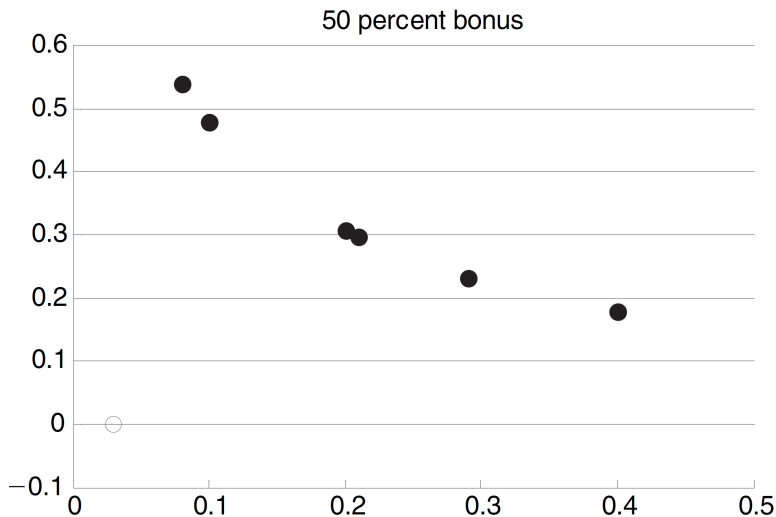
Accelerated Depreciation and Investment Stimulus

- Even if we don't switch to cash-flow tax, it is clear that increasing z reduces user cost of capital and increases incentive to invest
- In a dynamic model, particularly strong effect from temporary accelerated depreciation because of intertemporal substitution effect
- How sensitive is investment to accelerated depreciation provisions in practice?
- Can study this empirically by using enactment of bonus depreciation during recessions

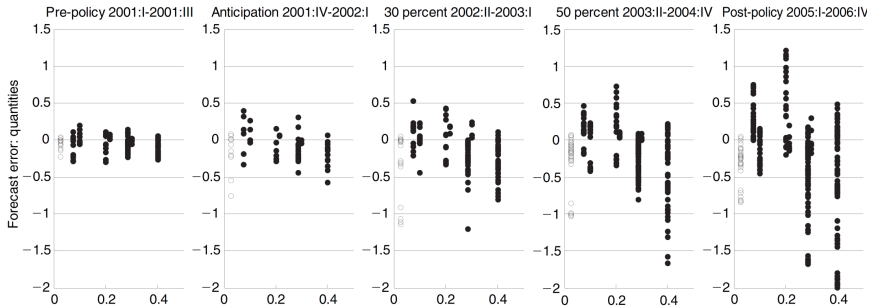
Accelerated Depreciation: Empirical Evidence

- House and Shapiro (2008) study impacts of accelerated depreciation in 2002 and 2003 tax bills
- Research design: diff-in-diff
 - Control group: Assets depreciated over more than 20 years (e.g. buildings) not granted accelerated depreciation
 - Variable intensity of treatment: bigger gains for investments with longer depreciation horizons (below 20 years)

Predicted Impact vs. Tax Depreciation Rate



A. Investment quantities

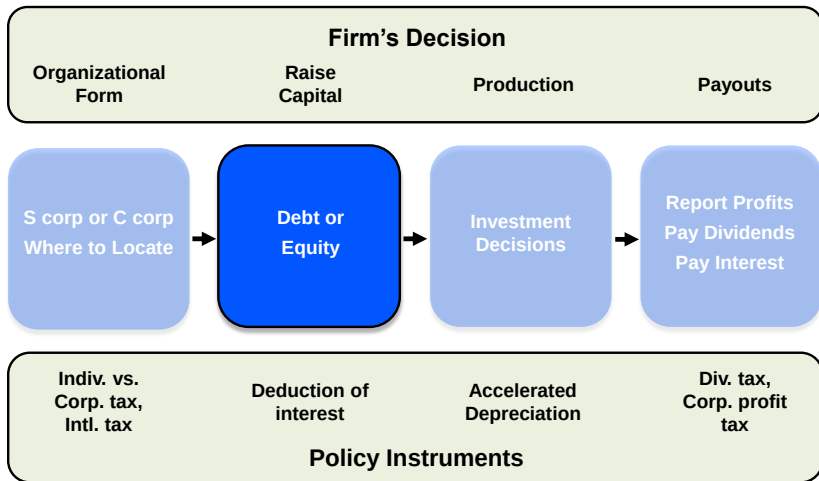


- House and Shapiro estimate an intertemporal substitution elasticity for capital investment w.r.t. user cost above 6
- Interpret the responses to tax incentives as price effects
 - Why? Corporations maximize profits (linear utility) in standard model, with no constraints
- But large literature in past 20 years in corporate finance on “investment cash-flow sensitivity” suggests there are liquidity effects

Liquidity Effects

- Ex: Lamont (1997) studies impact of changes in free cash flow in conglomerates
 - Uses shocks to oil prices as an instrument
 - Oil companies significantly increase investment in non-oil subsidiaries (relative to others in same industry) when oil prices are high
- Implies that accelerated depreciation could raise investment partly through an income effect
- Open question: how do liquidity constraints interact with impacts of corporate tax policy?
 - Should we give tax breaks to liquidity constrained firms in recessions?

Corporate Decisions and Tax Policies



Financing Investment: Debt vs. Equity

- Let τ_z denote individual income tax rate on interest income
- Debtors who lend z to firm in period 1 get paid $z + z \cdot i \cdot (1 - \tau_z)$ in period 2
- Let τ_e denote effective tax rate on equity payout (combination of capital gains and dividend)
- Let τ_c denote corporate profit tax rate

Financing Investment: Debt vs. Equity

- Let marginal return to one more dollar of investment be $\rho = f'(I)$
- Suppose we finance that dollar of investment, paying lenders an interest rate of ρ

- Debt is deductible from profits, so net return is

$$(\rho(1 - \tau_c) + \rho\tau_c)(1 - \tau_z) = \rho(1 - \tau_z)$$

- Net return to financing marginal investment from equity is

$$(1 - \tau_c)(1 - \tau_e)\rho$$

- Debt is preferred to equity if

$$(1 - \tau_z) > (1 - \tau_c)(1 - \phi)$$

Financing Investment: Debt vs. Equity

- Debt is highly tax-favored in the U.S.: τ_c is close to 40% in U.S. and highest income tax rate is below that
- So even without dividend/cap gains taxation, debt is favored
- So why don't firms use only debt finance?

Neoclassical Explanations

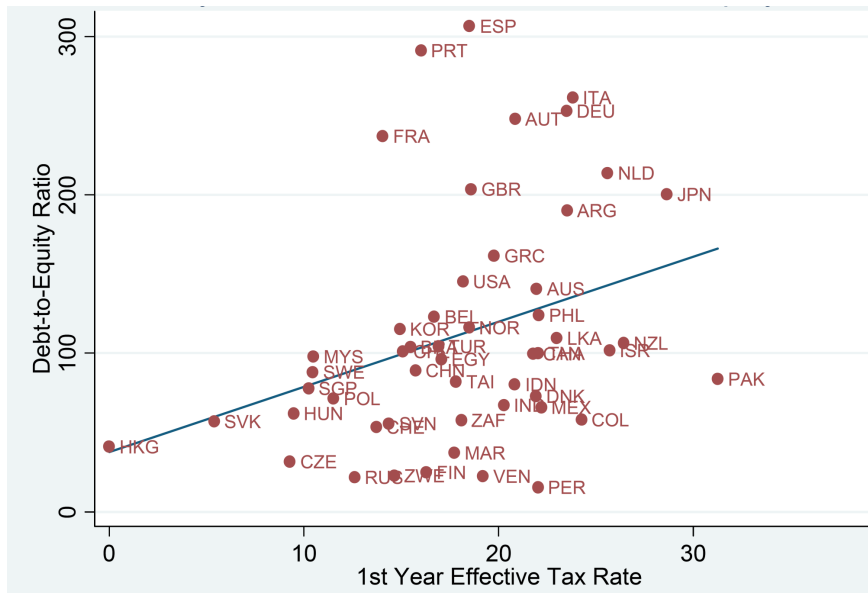
- Effective tax rate on equity may not be that much higher than debt
- First, cannot deduct interest payments if they exceed current earnings (but can carry forward)
- Altshuler and Auerbach (1991) estimate that average value of interest deductibility was 32% in practice
 - Far below statutory corporate rate of 46%
 - Suggests that constraints on deducting interest payments are binding for many firms
 - Marginal value of further debt may not be too high

- Effective tax rate on equity may not be that much higher than debt
- Second, could have a Miller (1977) equilibrium
 - If some investors face 0 tax rates on equity (e.g. pension funds) then effective tax rate on equity could be low
 - Clientele effect: pension funds buy shares before they pay dividends
 - Does not occur fully in practice because we do collect taxes from equity

Corporate Finance Explanations of Debt/Equity

- Debt and equity are not perfect substitutes in a model with uncertainty and asymmetric information
 - ① Debt requires fixed payments; with uncertain payoffs, defaulting on payment leads to potentially large bankruptcy+restructuring costs
 - ② Limited-liability of debt creates excess risk taking incentives (Myers 1977) for managers (no downside risk), increasing cost of debt finance
 - ③ Debt disciplines managers (Jensen 1986) and so they avoid it so that they are under less pressure
- With all of these factors, should still see changes in tax rates on interest income and corporate profits affecting debt/equity ratio

Effective Corporate Tax Rate and Debt to Equity Ratio



- Cross-sectional variation in marginal tax rates created by carry-forward of previous losses
 - Prior losses can be subtracted from current profits and used to reduce tax liability
 - Some firms effectively face 0 marginal corporate tax rate because of this

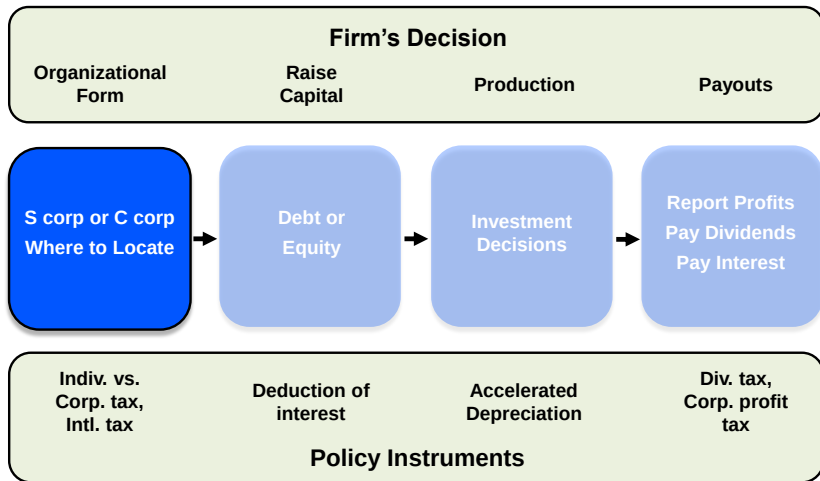
Empirical Evidence on Taxes and Debt

- Several studies use cross-sectional and panel variation in τ_c generated by such variation
- Regress debt/equity ratio on marginal tax rates (e.g., Auerbach 1985, Graham 1996)
 - General finding: firms who have less incentive to use debt do use less of it
 - Difference in tax rates explain about 15% of the variance in D/E ratios across firms
- But aggregate responses to tax reforms such as TRA86 less clear (Gordon and Mackie-Mason 1990)

Empirical Evidence on Taxes and Debt

- Several important directions for further work
 - ① Quasi-experimental evidence on how taxes affect debt/equity ratio
 - ② Do changes in debt/equity mix generated by taxes ultimately change firms' investment and employment decisions?
 - ③ What is the net impact on social welfare? Should we be subsidizing debt as we do today?
- Derive and implement optimal policy formulas in modern corporate finance models

Corporate Decisions and Tax Policies



Choice of Organizational Form

- Firms can choose to organize themselves as C corps (subject to corporate tax) or S corps (subject to indiv income tax)
- Tradeoff between C and S is quite similar to debt vs equity choice conceptually, and similar puzzles arise
- Why aren't all firms S corporations?
 - There are legal limits to public visibility and institutional ownership if an S corp.
 - S corporations cannot have more than 100 shareholders, limiting ability to raise equity

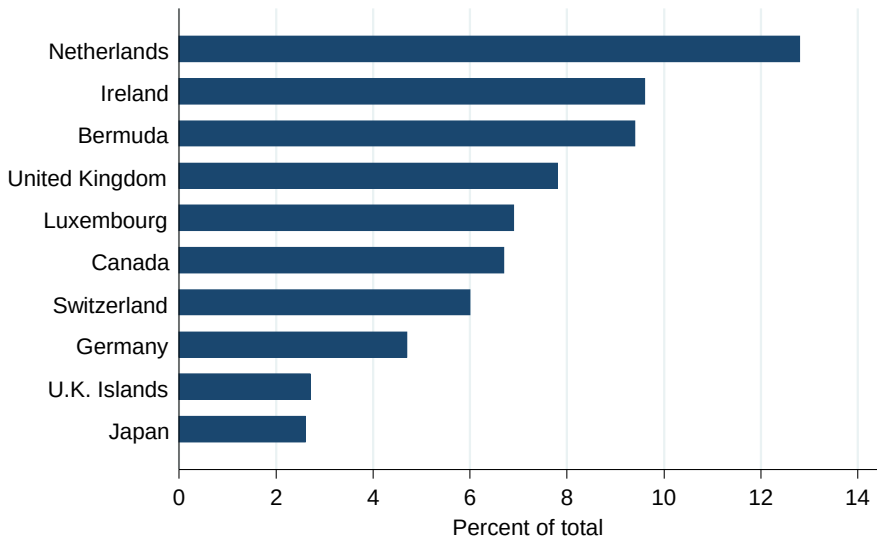
Choice of Organizational Form

- If taxes distort choice of incorporation, could lead to inefficiency
 - Ex: under-utilization of external capital for investment
- There is a correlation between corporate tax rates and fraction of assets in corporate form in time series (Mackie-Mason and Gordon 1997)
- But again quasi-experimental evidence limited and unclear what the real impacts are
- Also unclear what optimal tax policy is in modern corporate finance models
 - Should we be giving startups preferential tax treatment? Are S corps. the best way to do this?

International Taxation

- Another increasingly important dimension of firm organization that may be distorted by tax policy: location
- U.S. has a worldwide tax system
 - Pay U.S. tax only when you bring money back to the U.S. (e.g., to pay dividends)
 - Repatriations taxed on difference between tax rate where profits are earned and U.S. corporate tax of 35%
 - Ex: if company makes \$100 of profit in Ireland (4% avg corp tax), pay \$31
- Creates a strong incentive to report profits and hold money as retained earnings in foreign “tax havens.”

Figure 1. Principal Locations of Non-U.S. Profits of U.S. Multinationals, 2003
(profits as a percentage of the worldwide total)



Source: Bureau of Economic Analysis; Operations of U.S. Parent Companies and their Foreign Affiliates, various years.

Transfer Pricing and Profit Shifting

- How do firms realize such high profits in Bermuda?
- Set up a subsidiary that provides a service such as accounting in Bermuda
 - Sell this service at a very high cost to the firm's main branch in the U.S., reducing U.S. profits
 - Leads to high declared profits in Bermuda
- Government tries to limit this with “transfer pricing” rules
 - Transactions have to be priced at same rate as arms-length transactions on the market
 - But very hard to enforce in practice

Tax Havens and Shifting

- How much do firms actually shift reported profits in response to local tax rates?
- Hines and Rice (1994) estimate impacts of tax rates on declared profits
- Regress declared profits of U.S. multinationals on local tax rates in a sample of tax havens and other countries
 - Identification relies on orthogonality of tax rates to other factors (e.g., bureaucracy)
 - Conclusion: a one percentage point higher tax rate reduces reported profits by US multinational firms by 3 percent

Effect of Tax Rates on Location of Nonfinancial Profits

Dependent variable:
log (pretax nonfinancial income)

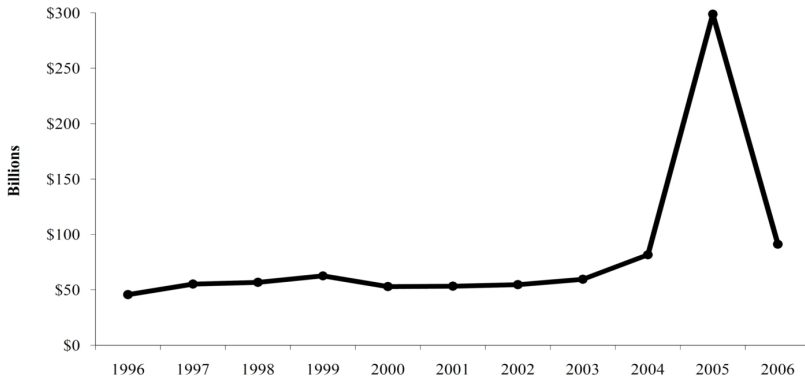
	Ordinary least squares estimation				Instrumental variable estimation ^a			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	1.36 (0.35)	0.17 (0.92)	1.30 (1.02)	-0.72 (0.88)	1.62 (0.41)	0.82 (1.61)	-0.41 (4.00)	0.60 (0.32)
Tax	-2.83 (0.59)	-2.25 (0.71)	-7.41 (2.41)		-3.65 (0.88)	-2.97 (1.63)	-24.92 (24.28)	
(Tax) ²			8.32 (3.72)				49.73 (54.32)	
(1 - Tax) ⁻²				-0.19 (0.09)				-0.41 (0.32)
log (plant, property and equipment)	0.41 (0.19)	0.48 (0.20)	0.41 (0.19)	0.56 (0.21)	0.35 (0.20)	0.42 (0.24)	0.78 (0.68)	0.43 (0.28)
log (compensation)	0.43 (0.19)	0.30 (0.21)	0.40 (0.20)	0.17 (0.21)	0.49 (0.20)	0.39 (0.28)	-0.12 (0.86)	0.35 (0.34)
log (GDP per capita)		0.16 (0.11)	0.06 (0.12)	0.26 (0.11)		0.09 (0.17)	0.34 (0.49)	0.15 (0.19)
S.E.E.	0.70	0.70	0.67	0.72	0.72	0.70	1.65	0.77
Adjusted R ²	.87	.87	.88	.86				
n = 59								

Taxes and Repatriation of Profits

- Dharmapala, Foley, and Forbes (2009) study intertemporal substitution in repatriation of profits
- Homeland Investment Act of 2004: provided a one-time tax holiday for the repatriation of foreign earnings by US multinationals
 - Goal: increase investment and create jobs in the U.S. by bringing money home
 - If US multinationals' domestic activity is financially constrained, repatriations should increase domestic employment and investment
- More generally, provides good quasi-experimental evidence on financial and real impacts of international tax laws

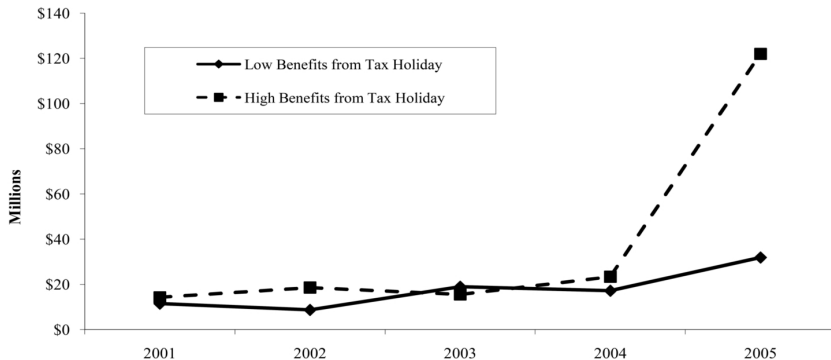
- Research design: diff-in-diff with variable intensity of treatment
- Use variation across tax rates in countries where companies reported profits prior to reform
- Companies with profits in countries with low tax rates (tax havens) have a large change in tax incentives from HIA
- Identification assumption: changes in repatriation around 2004 would have been the same across companies with profits in different countries absent HIA

Total Repatriations by U.S. Multinational Companies



Notes: Data on aggregate repatriations are from the Bureau of Economic Analysis, U.S. International Transactions Accounts Data, Table 7b, line 3 for distributed earnings. 2006 data are preliminary.

Mean Repatriations for Different Types of U.S. Multinationals



Notes: The sample used to construct this figure includes the balanced panel of firms for which data are available from 2001-2005. The dashed line displays mean repatriations for firms that are expected to have high benefits from a tax holiday. Firms are expected to have high benefits from the tax holiday if, in 2004, they (a) face lower corporate tax rates abroad and (b) have an affiliate that is a holding company or in a tax haven. The solid line displays mean repatriations for firms that do not meet either of these criteria.

The Effects of Repatriations on U.S. Capital Expenditures, U.S. Employment Compensation, and R&D

Dependent Variable:	U.S. Capital Expenditures/ Lagged Assets		U.S. Employment Compensation/ Lagged Assets		R&D/ Lagged Assets	
	(1)	(2)	(3)	(4)	(5)	(6)
Repatriations/Lagged Assets	0.0212 (0.0268)	0.0033 (0.1272)	0.0097 (0.0538)	-0.2345 (0.5258)	0.0147 (0.0238)	-0.1454 (0.1242)
Lagged Leverage	-0.0409** (0.0070)	-0.0409** (0.0070)	-0.1639** (0.0283)	-0.16444** (0.0286)	-0.0073* (0.0039)	-0.0077* (0.0040)
Lagged Tobin's q	0.0032** (0.0009)	0.0032** (0.0009)	0.0114** (0.0023)	0.0113** (0.0023)	0.0069** (0.0012)	0.0068** (0.0013)
Lagged Cash/Lagged Assets	-0.0086 (0.0152)	-0.0084 (0.0153)	-0.1031** (0.0386)	-0.1003** (0.0368)	-0.0262** (0.0123)	-0.0243** (0.0128)
Lagged Profitability	0.0223** (0.0079)	0.0222** (0.0079)	-0.0528** (0.0234)	-0.0541** (0.0239)	0.0103* (0.0059)	0.0095 (0.0061)
Firm and year dummies?	Y	Y	Y	Y	Y	Y
Instrument with Haven or Holding Company Dummy * 2005 Dummy and High Tax Costs of Repatriation Dummy * 2005 Dummy?	N	Y	N	Y	N	Y
No. of Obs.	4,508	4,508	4,508	4,508	4,921	4,921
R-Squared	0.1645		0.1161		0.1159	

The Effects of Repatriations on Dividends and Repurchases

Dependent Variable:	Payouts/ Lagged Assets		Dividends/ Lagged Assets		Repurchases/ Lagged Assets	
	(1)	(2)	(3)	(4)	(5)	(6)
Repatriations/Lagged Assets	0.1018* (0.0589)	0.9244** (0.4192)	0.0102 (0.0127)	0.1546 (0.1082)	0.0882* (0.0511)	0.7893** (0.3606)
Lagged Leverage	-0.0389** (0.0106)	-0.0363** (0.0112)	-0.0176** (0.0033)	-0.0172** (0.0033)	-0.0144* (0.0086)	-0.0124 (0.0090)
Lagged Tobin's q	0.0038** (0.0019)	0.0042** (0.0020)	-0.0003 (0.0006)	-0.0003 (0.0006)	0.0049** (0.0015)	0.0052** (0.0015)
Lagged Cash/Lagged Assets	0.0707** (0.0181)	0.0605** (0.0182)	0.0121* (0.0068)	0.0105 (0.0067)	0.0544** (0.0138)	0.0450** (0.0139)
Lagged Profitability	0.0486** (0.0103)	0.0522** (0.0119)	0.0094** (0.0034)	0.0101** (0.0034)	0.0406** (0.0086)	0.0440** (0.0102)
Firm and Year Dummies?	Y	Y	Y	Y	Y	Y
Instrument with Haven or Holding Company Dummy*2005 Dummy and High Tax Costs of Repatriation Dummy*2005 Dummy?	N	Y	N	Y	N	Y
No. of Obs.	4,581	4,581	4,848	4,848	4,649	4,649
R-Squared	0.0796		0.0489		0.0675	

- Conclusion: tax break on repatriations did not directly lead to higher domestic investment or employment by affected firms
- Instead, \$1 increase in repatriations was associated with an increase of almost \$1 in payouts to shareholders
- But cash returned to shareholders may have been invested in the U.S. economy or consumed
- What are the implications for optimal taxation?
 - Should we tax foreign source income at all?
 - How to prevent “race to the bottom” in corporate taxation?