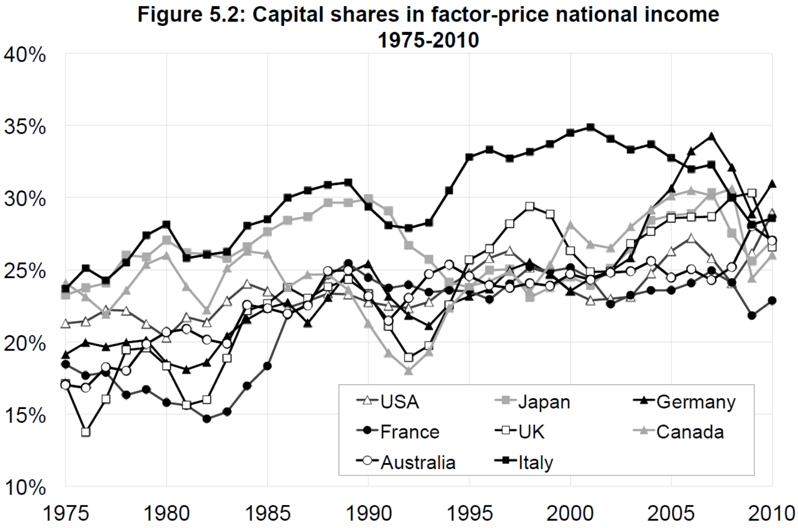
Graduate Public Economics

Lecture 1
Optimal Capital Taxation

Danny Yagan
UC Berkeley

Motivation: Large and rising capital income share

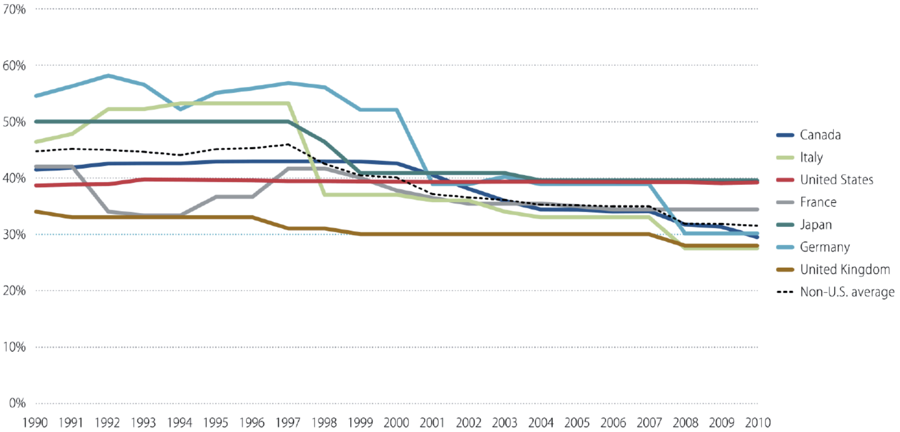


Source: Piketty-Zucman (2014)

Motivation: Declining capital income tax rates

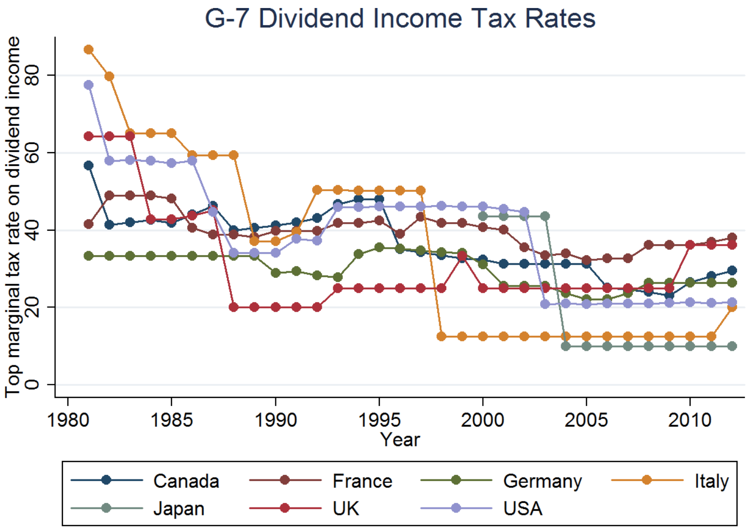
FIGURE 3

G-7 corporate tax rates, 1990-2010



Source: Auerbach (2010)

Motivation: Declining capital income tax rates



Source: Yagan (2015)

How much should we tax capital?

- Labor income taxation literature: Converged on Mirrlees (1971)
- Capital income taxation literature: All over the map
- Goals for today:
  1. Understand theoretical debate
  2. Compare theory to data
  3. Conclude with view of frontier

Arguments for zero capital income taxes

- Atikinson-Stiglitz (1976): PF pedigree
- Ramsey (1927)-Chamley (1986)-Judd (1985): macro pedigree
- Judd (1985)-Mankiw (2000): growth/history pedigree

Atkinson-Stiglitz (1976): Capital taxation is superfluous

- Consumers i earn labor in first period ("youth") and consume in both first period and second period ("retirement"):

$$\max_{c_1, c_{2,l}} U^i \left(v \left(c_{i1}, c_{i2} \right), l_{i1} \right)$$

$$\text{s.t. } c_{i1} + \frac{c_{i2}}{\left(1 + r \left(1 - \tau_k\right)\right)} = w_i l_{i1} - T_L \left(w_i l_i \right)$$

- Consumers differ by wages and weakly separable preferences over consumption and labor but share same subutility of consumption $v \left(c_{i1}, c_{i2} \right)$

- Government planner: maximize utilitarian social welfare function of individual utilities using nonlinear labor income tax $T_L \left(\cdot \right)$ and capital tax τ_k

- Optimum reached with $\tau_k = 0$

Atkinson-Stiglitz: Discussion

- Holds for arbitrarily many consumption periods and even when nonlinear labor income tax is suboptimal, as long as existing labor income tax can be perturbed (Kaplow 2006)
- Mechanism: Weak separability $\rightarrow$ capital taxation cannot relax incentive compatibility constraints or mitigate labor-leisure distortion, so all it does is distort consumption decisions
- Influential because of transparency and relationship to optimal nonlinear labor income tax problem (Mirrlees 1971)

Chamley (1986): No steady state capital tax

- Benchmark macro setup
  - Representative agent (Ramsey 1927)
  - Infinite horizon
  - Nonstochastic general equilibrium
  - Linear tax rates, no confiscation of period-zero capital
- "Primal" approach: solve for optimal allocations and back out taxes that generate those allocations (Atkinson-Stiglitz 1980; Chari-Kehoe 1999)

Chamley: Setup

1. Consumers:

$$\max_{c,l\geq 0,k,b} \sum \beta^t U(c_t, l_t) \quad \text{s.t.} \quad (1+\tau_{ct})\, c_t + k_{t+1} + b_{t+1}$$
$$\leq (1-\tau_{kt})\,(1+r_t-\delta)\, k_t + (1-\tau_{lt})\, w_t l_t + R_t b_t$$

2. Producers:

$$\max_{k_t, l_t} F(k_t, l_t) - r_t k_t - w_t l_t$$

3. Government budget constraint:

$$g_t + R_t b_t \leq \tau_{lt} w_t l_t + \tau_{kt}\,(1+r_t-\delta)\, k_t + \tau_{ct} c_t + b_{t+1}$$

4. Feasibility constraint:

$$c_t + g_t + k_{t+1} \leq F(k_t, l_t) + (1-\delta)\, k_t$$

Chamley: Government's problem

- Define "competitive equilibrium": sequence of allocations $\{c_t, l_t, k_t, b_t\}$ where (1)-(4) are satisfied

- Government's problem:

$$\max_{\tau_{lt},\tau_{kt},\tau_{ct}} \sum \beta^t U(c_t\left(\tau_{lt}, \tau_{kt}, \tau_{ct}\right), l_t\left(\tau_{lt}, \tau_{kt}, \tau_{ct}\right))$$

  where $c_t\left(\tau_{lt}, \tau_{kt}, \tau_{ct}\right)$ and $l_t\left(\tau_{lt}, \tau_{kt}, \tau_{ct}\right)$ are allocations in a competitive equilibrium and there is no lump-sum taxation: $\tau_{k0} = 0$, $\tau_{c0} = 0$.

Chamley: Focus on labor and capital taxes

- Only restrictions on tax rates come from consumer's (intratemporal and intertemporal) FOCs:

$$-\frac{U_{l,t}}{w_t U_{c,t}} = \frac{1-\tau_{lt}}{1+\tau_{ct}}$$
$$\frac{U_{c,t-1}}{\beta\left(1+r_t-\delta\right)U_{c,t}} = \frac{\left(1-\tau_{kt}\right)\left(1+\tau_{ct-1}\right)}{1+\tau_{ct}}$$

- $2*\infty$ equations $+$ $3*\infty$ unknowns $\implies$ must arbitrarily assign ∞ of them

- Here, assume $\tau_{ct}=0\ \forall t$, but note equivalence between capital taxes and increasing consumption taxes

Chamley: Solving the model

- Maximize social welfare over CE allocations (where $a_t \equiv (k_t + b_t)\, U_{c,t-1}$ for convenience):

$$\max_{c_t, k_t, a_t} \sum \beta^t U(c_t, l_t)$$

$$[\theta_t] \quad U_{c,t} c_t + U_{l,t} l_t + a_{t+1} - a_t/\beta = 0$$
$$[\lambda_t] \quad F(k_t, l_t) + (1-\delta)\, k_t - c_t - g_t - k_{t+1} = 0$$

- FOCs:

$$\{c_t\}: \ \beta^t U_{c,t} + \theta_t \left[U_{c,t} + U_{cc,} c_t + U_{cl,t} l_t\right] = \lambda_t$$

$$\{k_{t+1}\}: \ \ \lambda_{t+1}\left[1 + F_{k,t+1} - \delta\right] = \lambda_t$$

$$\{a_{t+1}\}: \ \theta_t = \beta \theta_{t-1}$$

Chamley: No steady state optimal capital tax

- Steady state: Assume $c_t \to c^{ss}, l_t \to l^{ss}, k_t \to k^{ss}, g_t \to g^{ss}$.

- Then government's FOCs reduce to:

$$\beta \left(1 + F_k^{ss} - \delta\right) = 1$$

- Recall consumer's intertemporal FOC:

$$\beta \left(1 - \tau_k^{ss}\right) \left(1 + r^{ss} - \delta\right) = 1$$

- Since $r^{ss} = F_k^{ss}$, these two conditions imply $\tau_k^{ss} = 0$.

Chamley: Discussion

- Transition: highest possible capital tax early on, declining thereafter

- Intuition: $\tau_k^{ss} > 0 \rightarrow$ ever-increasing unbounded tax on future consumption, and would rather smooth distortions (Judd 1999)

$$\frac{\beta \left(1 + r^{ss} - \delta\right) U_{c,t+1}}{U_{c,t}} = \frac{1}{1 - \tau_k^{ss}}$$

$$\lim_{T \to \infty} \frac{\beta^T \left(1 + r^{ss} - \delta\right)^T U_{c,t+T}}{U_{c,t}} = \lim_{T \to \infty} \left(\frac{1}{1 - \tau_k^{ss}} \right)^T = \infty$$

- Influential because of dynamic general equilibrium: consumers, firms, and government earn and spend in every period

Judd (1985) / Mankiw (2000): Heterogeneity

- Two types of agents: savers (as in Chamley) and spenders (hand-to-mouth: earn labor income but hold no capital)
- Striking result: Even (steady-state) spenders want no capital tax!
- Mechanism: Capital complements labor enough and is responsive enough to capital taxes that $(1 - \tau_l^{ss}) w^{ss} l^{ss}$ is maximized with $\tau_k^{ss} = 0$ and $\tau_l^{ss} > 0$
- Justification: Three hundred years of capital deepening and technology growth alongside dramatic increases in unskilled wages
- Similar force at work in Scheuer (2014) on entrepreneurship subsidies

What do these models miss?

1. Nonlinear tax instruments in Chamely-Judd
2. Steady state may not be optimal in Chamley-Judd
3. Relabeling of labor income as capital income
4. Preference heterogeneity
5. Finite tax elasticities of savings and investment
6. Divergent private and social valuations of future
7. Preferences for wealth equality
8. Future earnings uncertainty
9. Capital-labor substitutability

Saez (2013): Optimal progressive capital taxes

- Setup:
  - Infinite horizon like Chamley (1986)
  - Individuals start with different inheritances
  - Planner has access to two-bracket tax instrument: positive only above a chosen threshold
- Result: Positive rate above a wealth threshold can be optimal (also in Farhi-Werning 2010)
- Intuition: Drive large fortunes down to threshold in steady state $\rightarrow$ generate redistribution without infinitely compounding distortion
- U.S. estate tax has exactly this shape: 0% rate on estates up to $5.5m threshold, 40% thereafter

Straub-Werning (2014): Chamley fragility

- Chamley assumption: capital tax rate is bounded above in every period (in order to prevent early confiscation sunk capital)

- Key Chamley proof: "The [capital tax upper bound constraint] cannot be binding forever (the marginal utility of private consumption...would grow to infinity...which is absurd)."

- Straub-Werning: True only if interior steady state is optimal. But $c_t \to 0$ can be optimal if initial debt is large enough

- Intuition: Extraordinary distortions on consumption are bad, but extraordinary distortions on labor can be worse

- Broader point: PDV of utility can be higher with $\tau_{kt} > 0 \; \forall t$ than with $\tau_{kt} = 0 \; \forall t$, so $\tau_k^{ss} = 0$ can be poor policy guide

Relabeling labor income as capital income

- Assumption: Government observes labor income and observes capital income separately
- Practice: Can be very difficult to distinguish
  - Entrepreneurs (Bill Gates's Microsoft capital gains: labor or capital income?)
  - Hedge fund and private equity owners ("carried interest" taxed as capital gains)
  - Owner-managers of small businesses can pay themselves bonuses instead of declaring profits

Evidence on income shifting

- Gordon-Slemrod (1998, 2000): 1980s reductions in top individual income tax rates $\rightarrow$ large increase in business income being taxed as individual income (S corporation and partnership) rather than corporate income (C-corporation)

- Pirttila-Selin (2011): Finnish capital tax cut $\rightarrow$ large shift from labor income base to capital income, especially among self-employed

- Jacob-Michaely-Alstadsæter (2015): Swedish dividend tax cut $\rightarrow$ owner-managers reduced their wage compensation and increased dividends

Theory on income shifting

- Government can't distinguish capital and labor income at all and shifting elasticity is infinite $\rightarrow \tau_k = \tau_l$ (Piketty-Saez 2013; Christianson-Tuomala 2008)

- Finite but sufficiently strong shifting elasticity $\implies \tau_k^{ss} > 0$ even in Chamley (own numerical simulations)

- But this rationale for $\tau_k > 0$ requires reason for not just taxing consumption

- Income shifting seems important in real world
  - $\tau_k \approx \tau_l$ in many countries
  - U.S. S corporations allow owner-managers of closely-held businesses to have their profits taxed at individual income tax rates (obviates incentive to evade taxes by labeling profits as bonuses)

Preference heterogeneity in Atkinson-Stiglitz

- Standard Atkinson-Stiglitz: individual's allocation between c_1 and c_2 contains no information beyond income
- But if high-skilled have stronger preferences for c_2, individual's consumption allocation contains information on the person's skill
  - Taxing capital loosens the IC constraints (Akerlof 1978; Saez 2002; Diamond-Spinnewijn 2011)
- Real world: Patience is very correlated with skill (Parker-Fischhoff 2005, Bettinger-Slonim-2005, Kirby-Winston-Santiesteban 2005; see Banks-Diamond 2010)

Finite tax elasticities of savings and investment

- Why do people save?
  - Modigliani/Tobin (like Atkinson-Stiglitz application): life-cycle consumption smoothing
  - Chamley-Judd: dynastic: infinite horizon consumption smoothing

- If people save for different reasons, can get different optimal tax prescriptions

- Proceed here in three steps:
  1. Is most wealth life-cycle savings rather than inheritances?
  2. If not, is zero capital tax indeed optimal in closed economy?
  3. What about in an open economy?

Inheritance share of total wealth

- Huge debate: Kotlikoff-Summers 1981, Kotlikoff vs. Modigliani in JEP 1988
- Modigliani: does not capitalize inherited wealth $\rightarrow$ inheritance share is 20-30%
- Kotlikoff-Summers: do capitalize inherited wealth, even if heirs consume out of it $\rightarrow$ inheritance share is 80%

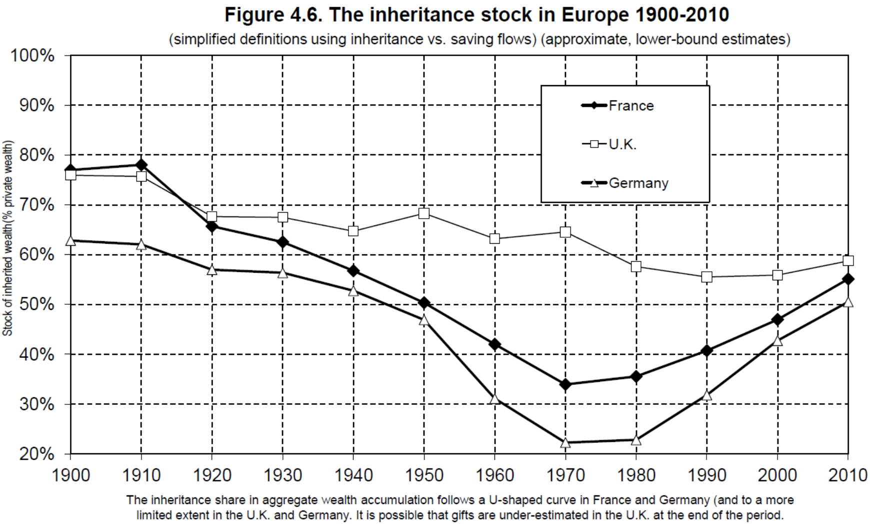
Inheritance share of total wealth

- Piketty-Postel-Vinay-Rosenthal (2013): use micro data to split population into two groups such that inheritances are capitalized but inheritance share is bounded at 100% (see Piketty-Zucman 2014)
    - "Self-made individuals": [current wealth] $>$ [capitalized value of inheritance]
    - "Rentiers": [current wealth] $<$ [capitalized value of inheritance]

- Inheritance share can grow large when $r > g$ (Piketty 2011, Piketty Zucman 2014)

- Inheritance share over 50% in Europe (likely smaller in U.S.?)

Inheritance share of total wealth in Europe



Source: Piketty-Zucman (2014)

Why do people give inheritances?

- Ramsey-Chamley-Judd: Bequest motive only
- Structural estimation: Only half of bequeathed wealth is due to bequest motive (Kopczuk Lupton 2007)
- Income shocks to parents affect parents' consumption more than kids' consumption (Altonji-Hayashi-Kotlikoff 1992, 1997)

Why do people give inheritances?

- Accidental bequests because of imperfect annuitization (Finkelstein-Poterba 2002, 2004) $\rightarrow$ potential rationale for estate taxation (bequest is not worth much to donor)

- Social status / wealth-in-the-utility-function (Carroll 2008) $\rightarrow$ potential rationale for estate (and other capital) taxation if only rank matters

- Social/family pressure $\rightarrow$ potential rationale for estate taxation since can strengthen donors' bargaining power and make them better off (Aura 2005; Wilhelm 1996; Light-McGarry 2004)

- Strategic bequests to extract labor from children (Bernheim-Shleifer-Summers 1985) $\rightarrow$ bequest is consumption for donor (Atkinson-Stiglitz no-taxation applies) but is effectively labor income for donee and thus optimally taxed

Optimal inheritance taxation (Piketty-Saez 2013)

- Atkinson-Stiglitz fails when consumers have inheritance income, not just labor income
  - Intuition: two dimensions of heterogeneity (wage and inheritance) $\rightarrow$ need two nonlinear tax instruments

- Optimal inheritance tax rate from "Meritocratic Rawlsian" perspective:

$$\tau_B = \frac{1 - \bar{b}}{1 + \varepsilon_B}$$

  where $\bar{b}$ is share of average bequest that zero-receivers leave, and ε_B is the inheritane tax elasticity of bequests

- Nests version of Ramsey-Chamley-Judd ($\varepsilon_B = \infty$ when r is exogenous)

- But value of ε_B is unresolved empirically (Kopczuk-Slemrod 2001)

Potential for small tax elasticity of savings

- Classic two-period consumption model:

$$\max_{c_1, c_2} u\left(c_1\right) + \delta\left(c_2\right)$$

$$\text{s.t. } c_1 + c_2 / \left(1 + r\right) \leq z_1 + z_2 / \left(1 + r\right)$$

- With capital tax, budget constraint becomes:

$$c_1 + c_2 / \left(1 + r\left(1 - \tau_k\right)\right) \leq z_1 + z_2 / \left(1 + r\left(1 - \tau_k\right)\right)$$

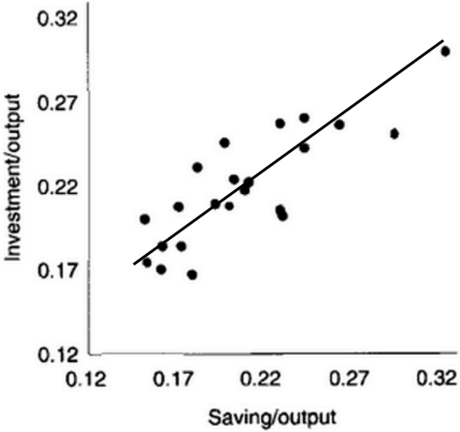
Potential for small tax elasticity of savings

- Substitution effect: $\tau_k \uparrow \Rightarrow$ price of $c_2 \uparrow$ relative to $c_1 \Rightarrow c_2 \downarrow$
- Wealth effect: $\tau_k \uparrow \Rightarrow$ poorer (if already saving) $\Rightarrow$ could increase c_2 (i.e. capital tax elasticity of savings can be zero or negative!)
- Rich donor heuristic for how much to bequeath (e.g. Bill Gates): "I want my kids to have \$10 million after taxes. What's the pre-tax amount I need to leave them?"
- Deadweight loss can be large even with zero elasticity (Feldstein 1978)
    - Policy relevance depends on marginal social welfare weight placed on savers

Tax elasticity of investment in open economy

- In small open economy, taxation of domestic residents' capital income may have no impact on domestic capital accumulation
  - Domestic investment always earns r^* no matter what $(1 - \tau_k)\, r^*$ domestic residents earn
- How internationally mobile is capital?

Challenge to international capital mobility



$\beta = .62$
$(.09)$

Figure 3.4
Industrial-country saving and investment rates, 1982–91

Source: Obstfeld-Rogoff (1996), Feldstein-Horioka (1980)

Divergent private and social valuations of future

- In overlapping generations model, steady state is dynamically inefficient if $r < g$ (Phelps 1961; Diamond 1965)
  - Economy is investing too much: can generate Pareto improvement by consuming more today and holding future consumption fixed

- Modified Golden Rule: $r = \delta + \gamma g$, with $r = F_k$, $\delta =$ social (gov.) discount rate and CRRA(γ) utility curvature $u'(c) = c^{-\gamma}$
  - Standard perturbation argument, but for planner (Piketty-Saez 2013)

$$u'(c_t) = \frac{1+r}{1+\delta} u'(c_{t+1})$$

$$\left(\frac{c_{t+1}}{c_t}\right)^{\gamma} = \frac{1+r}{1+\delta}$$

$$1 + r = (1+\delta)(1+g)^{\gamma}$$

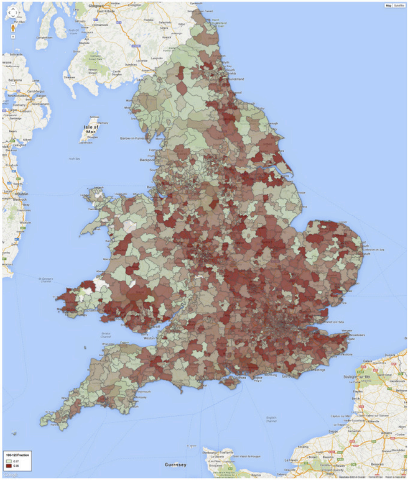
which is approximately equivalent to $r = \delta + \gamma g$ for small increments

Divergent private and social valuations of future

- What value for δ? [Nordhaus vs. Stern]
  - Reasonable upper bound: private sector rate of 1.4% (Giglio-Maggiori-Stroebel 2015)

Measuring very long discount rates

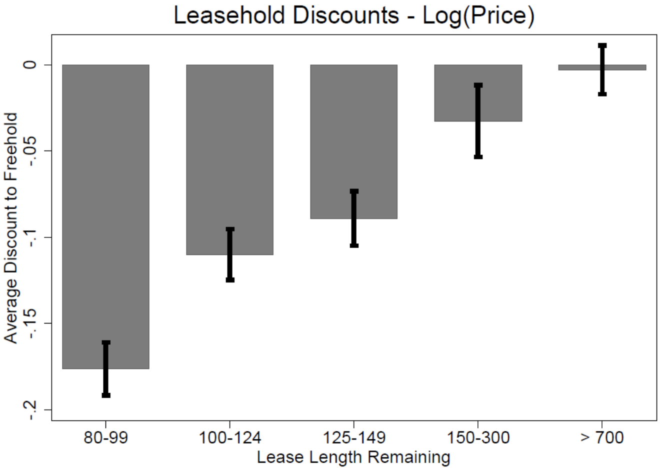
Figure A.7: U.K. Flats: Fraction of 100-124 years leaseholds



Note: Figure shows the fraction of flat transactions with 100-124 years remaining in each UK 3-digit postcode. Green and red correspond to the 10th and 90th percentile of the distribution of the fraction across postcodes.

Source: Giglio-Maggiori-Stroebel 2015

Measuring very long discount rates



(b) Price Discount by Remaining Lease Length

Source: Giglio-Maggiori-Stroebel 2015

Divergent private and social valuations of future

- What value for δ? [Nordhaus vs. Stern]
  - Reasonable upper bound: private sector rate of 1.4% (Giglio-Maggiori-Stroebel 2015)
  - Even 1.4% may be too high: extra dessert for Cleopatra $\rightarrow$ millions go without cancer treatment today (Cowen-Parfit 1991)

- What value for γ?
  - γ high $\rightarrow$ care a lot about inequality $\rightarrow$ want small capital stock and thus large r $\rightarrow$ global warming is not important
  - γ low $\rightarrow$ do not care about inequality $\rightarrow$ want large capital stock and thus small r $\rightarrow$ should care a lot about global warning

- Real world: $r > g$ $\rightarrow$ below socially optimal level of capital unless δ or γ is large $\rightarrow$ capital *subsidy* (King 1980, Atkinson-Sandmo 1980)
  - But if gov. really cares, ideally uses debt to get there, separating capital stock objectives from redistribution objectives (Piketty-Saez 2013)

Assessing golden rule / dynamic efficiency

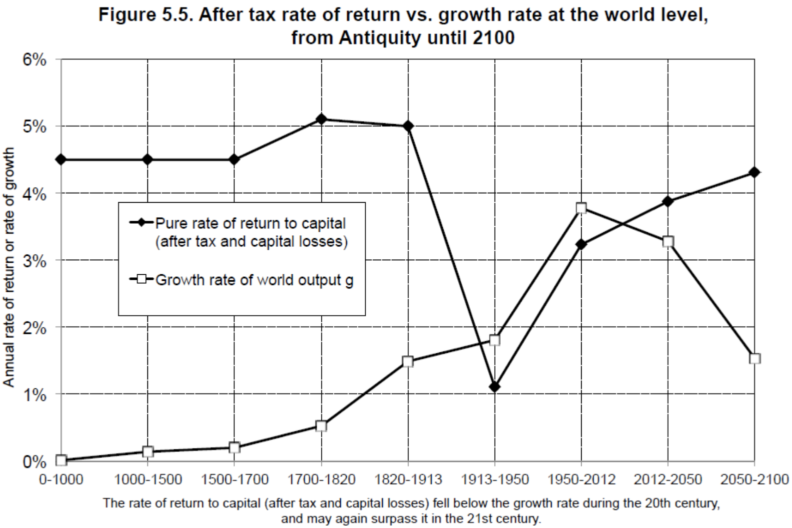
TABLE 2

Gross profit and investment: the U.S. nonfinancial corporate sector (percent)

Year	Gross profit / V		Gross investment / V	D / V
1953	29·1		14·7	13·6
1954	28·7		13·3	14·3
1955	25·1		13·3	10·2
1956	21·0		12·3	7·0
1957	19·4		10·8	7·5
1958	20·6		9·6	10·0
1959	18·2		9·8	7·6
1960	16·7	>	9·2	6·3
1961	16·9		8·9	7·5
1962	15·5		8·9	6·9
1963	18·0		9·8	8·0
1964	17·1		9·5	7·3
1965	17·5		10·5	6·9
1966	17·4		11·7	5·9
1967	18·4		11·6	6·9
1968	16·1		9·7	6·2

Source: Abel, Mankiw, Summers, Zeckhauser (1989)

Assessing golden rule / dynamic efficiency



Source: Piketty-Zucman (2014)

Preferences for wealth equality

Source: Norton-Ariely 2011

Preferences for wealth equality?

- U.S. estate tax: 40% rate above $5.5 million exemption after charitable and spousal deductions $\rightarrow$ only 0.1% of decedents liable
- Support for estate tax rises from 17% to 53% when Mechanical Turk survey respondents are (dramatically) informed that only the richest are liable (Kuziemko-Norton-Saez-Stantcheva 2013)

Preferences for wealth equality?

Besides the income tax, the government can also level the playing field with **the federal estate tax.**

The **Federal Estate Tax** (also known as the **Death Tax**) applies when a deceased person leaves **more than $5 million** in wealth to his or her heirs. Wealth left to a spouse or charitable organizations is exempt from estate tax.

Only 1 person out of 1000 is wealthy enough to face the estate tax.

Average Americans do not have anything close to $5 million in wealth, so the estate tax does not affect them and they can pass on their property to their children tax-free.

Eliminating the estate tax would allow the very richest families to pass down all of their wealth to their children tax-free. Hence, children of rich people would also start off very rich themselves.

Increasing the estate tax is a way to level the playing field between the children of wealthy parents and children of middle-class parents.

Future earnings uncertainty (New Dynamic Public Finance)

- Setup (Golosov-Kocherlakota-Tsyvinski 2003, Kocherlakota 2004):
  - Two periods of consumption: c_1 and c_2
  - Work only in second period (for simplicity)
  - Everyone is identical in period 0 but receives stochastic wage draw w in period 1
  - Utility: $u(c_1) + \beta [u(c_2) - h(l)]$ s.t. $c_1 + c_2 / (1+r) = wl / (1+r)$

Future earnings uncertainty (New Dynamic Public Finance)

- Euler with no government intervention (i.e. *private* optimum):

$$u'(c_1) = \beta(1+r)\int u'(c_2(w))\, f(w)\, dw$$

- As in Mirrlees, government wants to redistribute from high w to low w in period 1, but observes only c_1, c_2, wl

- At *government* optimum, "inverse Euler" equation holds by same type of perturbation argument for *social* welfare:

$$\frac{1}{u'(c_1)} = \frac{1}{\beta(1+r)}\int \frac{1}{u'(c_2(w))} f(w)\, dw$$

Future earnings uncertainty (New Dynamic Public Finance)

- Jensen's inequality: for $K(\cdot)$ convex

$$K\left(\int x\left(w\right)f\left(w\right)dw\right) < \int K\left(x\left(w\right)\right)f\left(w\right)dw$$

- Here, let $K\left(x\right) = 1/x$ and $x\left(w\right) = u'\left(c_2\left(w\right)\right)$:

$$\frac{1}{\int u'\left(c_2\left(w\right)\right)f\left(w\right)dw} < \int \frac{1}{u'\left(c_2\left(w\right)\right)}f\left(w\right)dw = \frac{\beta\left(1+r\right)}{u'\left(c_1\right)}$$

$$u'\left(c_1\right) < \beta\left(1+r\right)\int u'\left(c_2\left(w\right)\right)f\left(w\right)dw$$

- Result: government optimally distorts consumption to the present relative to the agent's Euler (private optimum), e.g. with a capital tax

Future earnings uncertainty (New Dynamic Public Finance)

- Mechanism: Being poorer in second period makes it costlier to pretend to be low-skilled $\rightarrow$ loosens gov.'s IC constraints

- Tangible policy implication: asset test for disability insurance (Golosov-Tsyvinski 2006)

- But overall welfare gains of optimal capital-and-labor taxation appear small (0.1% in aggregate welfare) relative to optimal labor income taxation (Farhi-Werning 2011; Golosov-Troshkin-Tsyvinski 2011)

Capital-labor substitutability

- Historically: Strong reason to think that capital has complemented labor

Big Mac Index

Table 1: McDonalds Cashier or Crew Wages and Big Mac Prices, December 1998

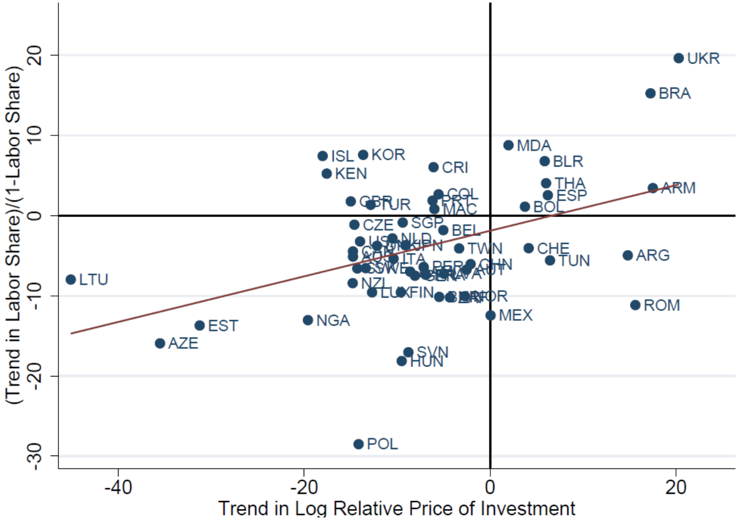
Country	Estimated hourly wage rate (1)	Reported Big Mac price (2)	Exchange Rate per $1 (3)	$ hourly wage rate (4)	$ Big Mac price (5)	Economist $ Big Mac 3/99** (6)	Big Macs per hour of work (7)
Russia	8.00	25.5	19.57	0.41	1.30		0.31
Korea	1700	3000	1210	1.41	2.48		0.57
Brazil	2.87	4.45	1.73*	1.66	2.57	1.71	0.65
Poland	4.12	5.3	3.50	1.18	1.51	1.38	0.78
Czech Rep.	45.00	53	30.30	1.49	1.75		0.85
UK	3.60		0.62*	5.80	3.07	3.07	1.89
USA	6.00		1.00	6.00	2.43	2.43	2.12
Germany	11.28	4.95	1.67	6.76	2.97	2.72	2.28
France	40.22	17.5	5.76	6.99	3.04	2.87	2.30
Italy	10417	4500	1646	6.33	2.73	2.5	2.31
Belgium	280.00	114	34.50	8.12	3.30		2.46
Sweden	64.90	25	8.03	8.09	3.11	2.88	2.60
Japan	844	280	120*	7.03	2.33	2.44	3.01

Source: Ashenfelter-Jurajda (2001)

Capital-labor substitutability

- Historically: Strong reason to think that capital has complemented labor
- Future: Unclear (Katz Murphy 1992)
  - [Elasticity of substitution between capital and labor of 1.25] + [25% global decline in relative price of investment] $\rightarrow$ explains half of 5-pp global decline in the labor share of income

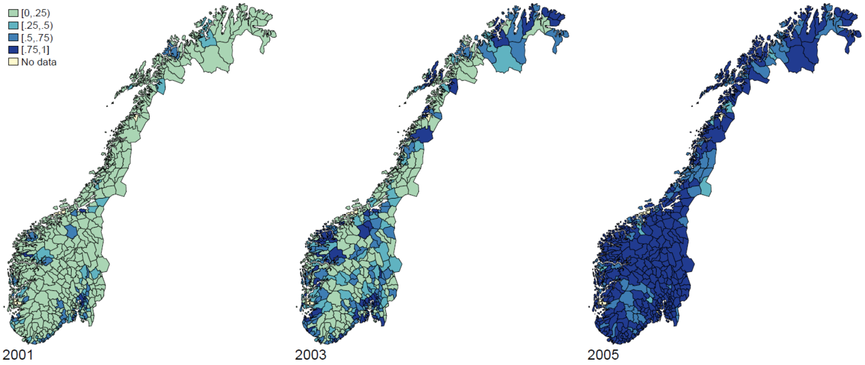
Labor share changes and the price of investment



Source: Karabarbounis-Nieman (2013)

Case study of complementarity: Rollout of broadband

Figure A4. Geographical distribution of broadband coverage rates.

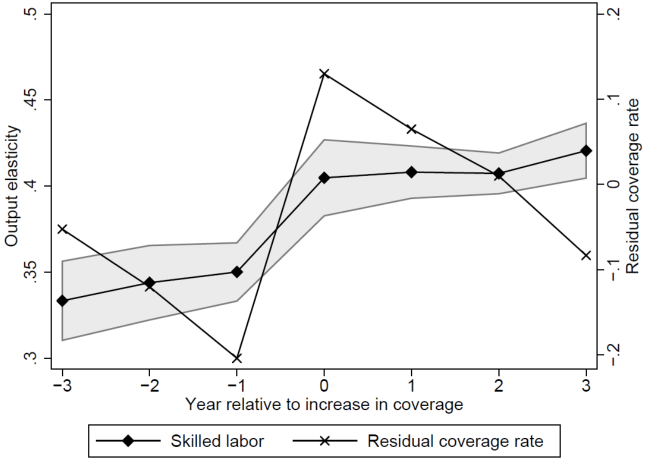


Note: The graphs show the geographical distribution of broadband coverage rates of households in 2001, 2003 and 2005.

Source: Akerman-Gaarder-Mogstad (2013)

Case study of complementarity: Rollout of broadband

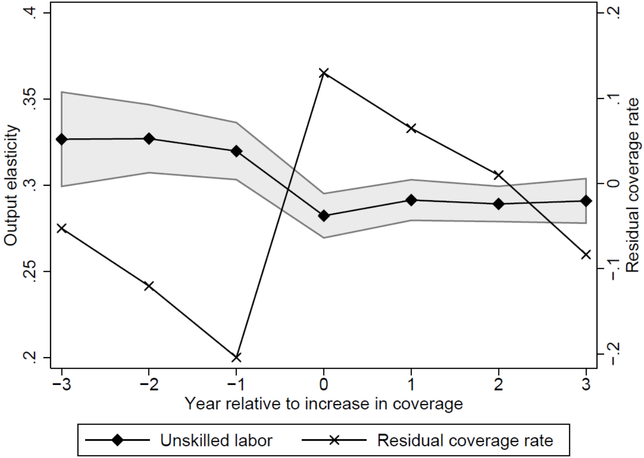
(a) Output elasticity: Skilled labor



Source: Akerman-Gaarder-Mogstad (2013)

Case study of complementarity: Rollout of broadband

(b) Output elasticity: Unskilled labor



Source: Akerman-Gaarder-Mogstad (2013)

Key empirical questions

- Tax elasticities of savings and investment
- Degree of international capital mobility
- Externalities of investment on workers
- Share of savings used for causes valued by government

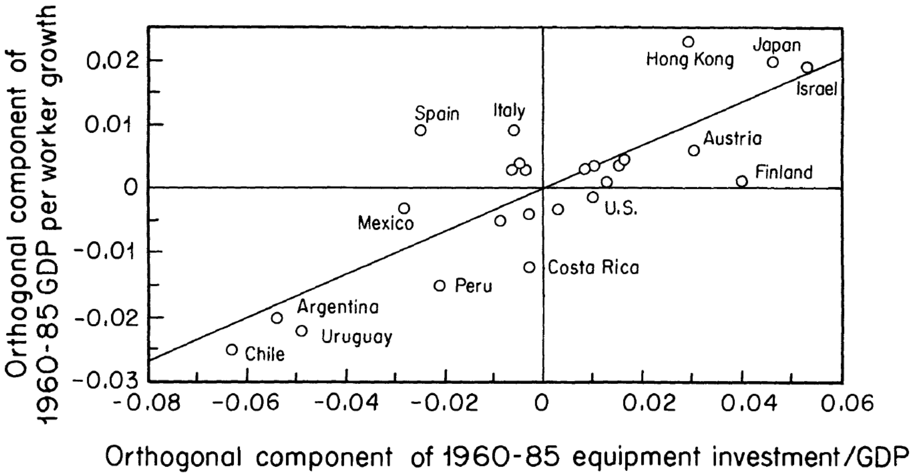
Graduate Public Economics

Lecture 2

Actual Capital Taxation – Investment

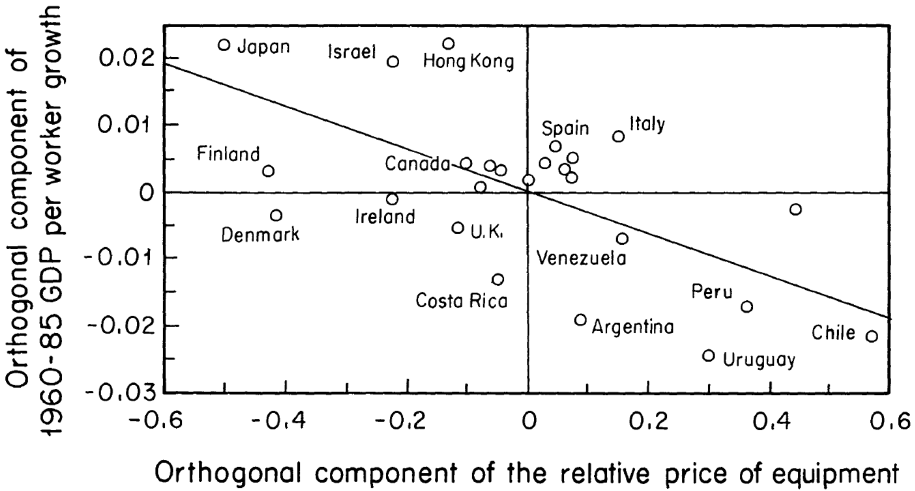
Danny Yagan
UC Berkeley

Motivation: Equipment investment and growth



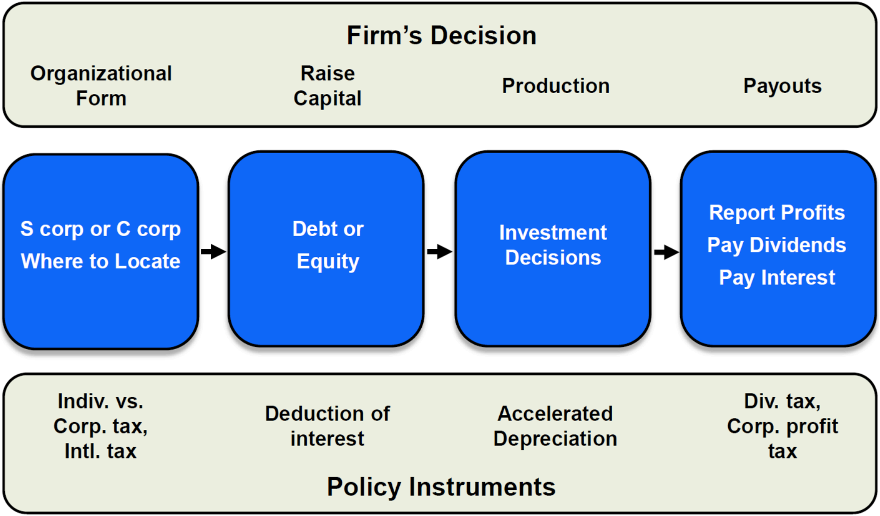
Source: De Long and Summers (1991)

Motivation: Equipment prices and growth



Source: De Long and Summers (1991)

Taxes and business investment



Source: Chetty and Bruich

How do taxes affect business investment?

- Theory: Cost of capital
- Evidence: Recent quasi-experiments
- Along the way: Departures from neoclassical considerations

Cost of capital (Hall-Jorgenson 1967)

- References: Hassett-Hubbard (2002), Auerbach (2002)
- Start with no taxes in general setup
  - Firm in period t deciding how much capital K_t to accumulate
  - Concave (gross) profit function (i.e. pre-taxes, post-deductions except depreciation deductions): $F(K_t)$
  - Price of capital goods: q_t
  - Depreciation rate (paid at purchase, before use): δ
  - Required rate of return: ρ

Cost of capital (Hall-Jorgenson 1967)

- NPV of a new machine (dK_{t+1}):

$$-q_t - \delta q_t + \frac{F'\left(K_{t+1}\right) + q_{t+1}}{1+\rho}$$

- Euler, equating marginal benefit to marginal cost at optimum:

$$F'\left(K_{t+1}\right) = q_t \left[\left(1+\delta\right)\left(1+\rho\right) - \frac{q_{t+1}}{q_t}\right]$$

$$F'\left(K_{t+1}\right) \approx q_t \left[\rho + \delta - \frac{q_{t+1} - q_t}{q_t}\right]$$

- RHS: "user cost of capital"
- With constant investment prices ($q_{t+1} = q_t$), return on marginal unit of investment $F'\left(K_{t+1}\right)/q_t$ equals required rate of return plus depreciation

Cost of capital (Hall-Jorgenson 1967)

- Add corporate income tax τ_t^{INC} (typically 35% in United States), which is assessed on gross profit (revenue minus deductions)

The U.S. corporate income tax form

Form **1120**
Department of the Treasury
Internal Revenue Service

U.S. Corporation Income Tax Return

For calendar year 2009 or tax year beginning ____________, 2009, ending ____________, 20 ____

▶ See separate instructions.

OMB No. 1545-0123

2009

A Check if:
- 1a Consolidated return (attach Form 851)
- b Life/nonlife consolidated return
- 2 Personal holding co. (attach Sch. PH)
- 3 Personal service corp. (see instructions)
- 4 Schedule M-3 attached

Use IRS label. Otherwise, print or type.

Name

Number, street, and room or suite no. If a P.O. box, see instructions.

City or town, state, and ZIP code

B Employer identification number

C Date incorporated

D Total assets (see instructions)

$

E Check if: (1) Initial return (2) Final return (3) Name change (4) Address change

Income

1a	Gross receipts or sales ____ **b** Less returns and allowances ____ **c** Bal ▶	1c	
2	Cost of goods sold (Schedule A, line 8)	2	
3	Gross profit. Subtract line 2 from line 1c	3	
4	Dividends (Schedule C, line 19)	4	
5	Interest	5	
6	Gross rents	6	
7	Gross royalties	7	
8	Capital gain net income (attach Schedule D (Form 1120))	8	
9	Net gain or (loss) from Form 4797, Part II, line 17 (attach Form 4797)	9	
10	Other income (see instructions—attach schedule)	10	
11	**Total income.** Add lines 3 through 10 ▶	11	

The U.S. corporate income tax form

Deductions (See instructions for limitations on deductions.)	12	Compensation of officers (Schedule E, line 4) . . . ▶	12	
	13	Salaries and wages (less employment credits) . . .	13	
	14	Repairs and maintenance . . .	14	
	15	Bad debts . . .	15	
	16	Rents . . .	16	
	17	Taxes and licenses . . .	17	
	18	Interest . . .	18	
	19	Charitable contributions . . .	19	
	20	Depreciation from Form 4562 not claimed on Schedule A or elsewhere on return (attach Form 4562) . . .	20	
	21	Depletion . . .	21	
	22	Advertising . . .	22	
	23	Pension, profit-sharing, etc., plans . . .	23	
	24	Employee benefit programs . . .	24	
	25	Domestic production activities deduction (attach Form 8903) . . .	25	
	26	Other deductions (attach schedule) . . .	26	
	27	**Total deductions.** Add lines 12 through 26 . . . ▶	27	
	28	Taxable income before net operating loss deduction and special deductions. Subtract line 27 from line 11 . . .	28	
	29	**Less: a** Net operating loss deduction (see instructions) . . . 29a		
		b Special deductions (Schedule C, line 20) . . . 29b	29c	
ments	30	**Taxable income.** Subtract line 29c from line 28 (see instructions) . . .	30	
	31	**Total tax** (Schedule J, line 10) . . .	31	

Cost of capital (Hall-Jorgenson 1967)

- Add corporate income tax τ_t^{INC}, which is assessed on gross profit (revenue minus deductions)

- Add NPV of depreciation deductions per dollar of investment in t:

$$\Gamma_t = \sum_{z=t}^{\infty} (1+r)^{-(z-t)} \tau_t^{INC} D_{z-t}, \text{ where } \sum_{z=t}^{\infty} D_{z-t} = 1$$

- To the extent $r > 0$ (i.e. there is discounting and inflation) and/or depreciation deductions are back-loaded (i.e. for long-lived assets), depreciation deductions are less valuable

- New Euler / cost-of-capital:

$$F'(K_{t+1}) = q_t \frac{1-\Gamma_t}{1-\tau_t^{INC}} \left[\rho + \delta - \frac{q_{t+1}(1-\Gamma_{t+1}) - q_t(1-\Gamma_t)}{q_t(1-\Gamma_t)}\right]$$

When does the corporate income tax distort capital stocks?

- Consider case of "immediate expensing" (investment cost is fully deductible immediately):

$$D_0 = 1, D_{z-t} = 0 \;\forall\, (z-t) > 0$$

$$\implies \Gamma_t = \tau_t^{INC}$$

- Then with constant taxes, the corporate income tax can raise revenue but is nondistortionary:

$$F'\left(K_{t+1}\right) = q_t \left[\rho + \delta - \frac{q_{t+1} - q_t}{q_t}\right]$$

What is going on?

- When all costs are deductible, the corporate income tax is a tax on pure profit, and the K that maximizes pure profit $\pi\left(K\right)$ also maximizes $\left(1-\tau^{INC}\right)\pi\left(K\right)$

- In real world:
  - Tax law allows only small profitable firms to immediately expense $\left(D_0 < 1\right)$
  - Tax law does not allow full deductibility of financing costs $\rightarrow$ $\rho'\left(\tau^{INC}\right) > 0$ [unless interest deduction compensates on average]

- Suggests one should "*narrow* the base and *increase* the rate," exactly the opposite of traditional logic and path of actual corp. tax policies

The U.S. corporate income tax form

Deductions (See instructions for limitations on deductions.)				
	12	Compensation of officers (Schedule E, line 4) ▶	12	
	13	Salaries and wages (less employment credits)	13	
	14	Repairs and maintenance	14	
	15	Bad debts	15	
	16	Rents	16	
	17	Taxes and licenses	17	
	18	Interest	18	
	19	Charitable contributions	19	
	20	Depreciation from Form 4562 not claimed on Schedule A or elsewhere on return (attach Form 4562)	20	
	21	Depletion	21	
	22	Advertising	22	
	23	Pension, profit-sharing, etc., plans	23	
	24	Employee benefit programs	24	
	25	Domestic production activities deduction (attach Form 8903)	25	
	26	Other deductions (attach schedule)	26	
	27	**Total deductions.** Add lines 12 through 26 ▶	27	
	28	Taxable income before net operating loss deduction and special deductions. Subtract line 27 from line 11	28	
	29	**Less:** **a** Net operating loss deduction (see instructions) 29a		
		b Special deductions (Schedule C, line 20) 29b	29c	
ments	30	**Taxable income.** Subtract line 29c from line 28 (see instructions)	30	
	31	**Total tax** (Schedule J, line 10)	31	

From optimal capital stock to optimal investment

- Hall-Jorgenson pins down the optimal capital stock
- Predicts that when τ^{INC} changes, the capital stock K adjusts immediately and permanently to a new level
- Need adjustment costs for realistic investment paths
- Hall-Jorgenson assume ad hoc adjustment path. Later research endogenized adjustment paths (Summers 1981; Abel 1982; Feldstein 1982; Auerbach-Hines 1987; Auerbach 1989; Auerbach-Hassett 1992)

Auerbach-Hassett (1992)

- Quadratic adjustment costs, Cobb-Douglas, and linearize from firm's steady state $\rightarrow$ investment is high relative to lagged capital when:
  - Near-term costs of capital are low relative to their steady-state value
  - The firm's capital stock is low relative to its steady-state value

$$\frac{I_t}{K_{t-1}} = \left[\left(\frac{1-\mu_1}{\alpha}\right)+\delta\right] - \left(\frac{1-\mu_1}{\alpha c_K^*}\right) K_{t-1}^{\alpha} E_t \sum_{s=t}^{\infty} w_{s-t} c_s$$

$$c_s = \frac{q\left(1-\Gamma_s\right)\left(\rho+\delta+\frac{\Gamma_{s+1}-\Gamma_s}{1-\Gamma_s}\right)}{1-\tau_s^{INC}}$$

(see AH appendix or Yagan 2015 appendix for full description)

- High adjustment costs $\rightarrow$ slowly declining weights w_{s-t}

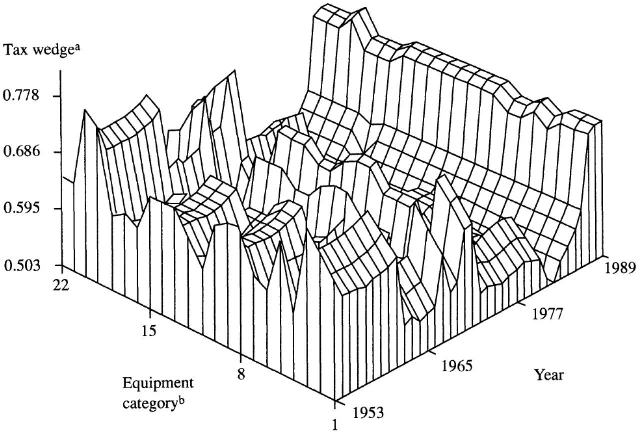
- Empirics: substantial effect of cost of capital on investment with substantial adjustment costs, but data reject model ($F(K)$ curvature α outside Cobb-Douglas feasible range $(0,1)$)

Cummins-Hassett-Hubbard (1994)

- Natural experiment idea: Estimate effect of tax reforms on investment by exploiting cross-sectional heterogeneity across firms in their cost-of-capital impacts, driven by asset length

Cummins-Hassett-Hubbard (1994)

Figure 2. After-Tax Cost of One Dollar of Equipment Investment, 1953–89



Source: Authors' calculations based upon data from the Bureau of Economic Analysis.

a. The tax wedge is calculated from Γ, which is the sum of the present value of tax savings from depreciation allowances and the investment tax credit. Higher values for $(1 - \Gamma)$ correspond to higher after-tax costs of investing.

b. See table 2 for BEA classifications.

Cummins-Hassett-Hubbard (1994)

- Natural experiment idea: Estimate effect of tax reforms on investment by exploiting cross-sectional heterogeneity across firms in their cost-of-capital impacts, driven by asset length

- Estimate year by year ("simulated instruments"):

$$\left(\frac{I_t}{K_{i,t-1}}\right) - \widehat{\left(\frac{I_t}{K_{i,t-1}}\right)} = \mu_i + \beta\left(C_{it} - \widehat{C_{it}}\right) + \varepsilon_{it}$$

- Finds large and significantly negative coefficients in tax-reform years, with implied cost-of-capital elasticity of investment equal to -0.66, $\sim$10x previous estimates

- Caveat: Method assumes no substitutability across asset types, and Caballero (1994 comment) did not replicate

Accelerated depreciation (House-Shapiro 2008)

- For long-lived capital goods, a temporary increase in the amount that can be immediately expensed $D_0 \rightarrow$ strong incentives to accelerate investment
- Major tool to stimulate investment: 30%-50% "accelerated" ("bonus") depreciation 2001-2004 for assets with recovery periods $\leq$ 20 years
- Because of discounting, this created heterogeneous subsidies (change in $1 - \Gamma$) across asset classes
- Similar DD empirical strategy to Cummins-Hassett-Hubbard (1994), except across asset classes directly rather than across firms specializing in different asset classes

Recovery periods by asset type

TABLE 2—RECOVERY PERIODS AND DEPRECIATION METHODS BY TYPE OF CAPITAL

Type of capital	Recovery period, *R* (years)	Tax depreciation rate, $\hat{\delta}$ (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 equiment	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).

Source: House-Shapiro (2008)

Subsidy from accelerated depreciation

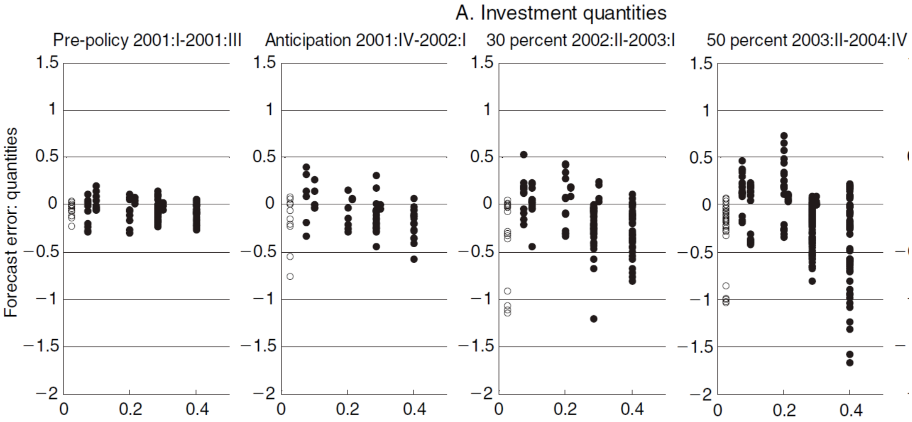
TABLE 3—QUANTIFYING DEPRECIATION ALLOWANCES

Recovery period	Nominal interest rate = 0.03			Nominal interest rate = 0.05			Nominal interest rate = 0.07		
	$\lambda^m = 0$	$\lambda^m = 0.3$	$\lambda^m = 0.5$	$\lambda^m = 0$	$\lambda^m = 0.3$	$\lambda^m = 0.5$	$\lambda^m = 0$	$\lambda^m = 0.3$	$\lambda^m = 0.5$
Panel A: Present value of depreciation allowances: $\lambda^m + (1 - \lambda^m)z^m$									
3 years	0.972	0.981	0.986	0.955	0.968	0.977	0.939	0.957	0.969
5 years	0.949	0.964	0.975	0.918	0.943	0.959	0.890	0.923	0.945
7 years	0.927	0.949	0.964	0.884	0.919	0.942	0.846	0.892	0.923
7 years (150DB)	0.914	0.939	0.957	0.863	0.904	0.932	0.818	0.872	0.909
10 years	0.896	0.927	0.948	0.837	0.886	0.919	0.786	0.850	0.893
10 years (150DB)	0.878	0.915	0.939	0.811	0.868	0.905	0.752	0.826	0.876
15 years	0.824	0.877	0.912	0.733	0.813	0.867	0.659	0.761	0.829
20 years	0.775	0.842	0.887	0.667	0.767	0.833	0.582	0.708	0.791
Panel B: Tax subsidy due to the bonus depreciation allowance, percent									
3 years	0.0	0.26	0.44	0.0	0.42	0.70	0.0	0.57	0.95
5 years	0.0	0.48	0.79	0.0	0.76	1.26	0.0	1.01	1.69
7 years	0.0	0.68	1.13	0.0	1.06	1.77	0.0	1.40	2.33
7 years (150DB)	0.0	0.80	1.33	0.0	1.25	2.08	0.0	1.64	2.73
10 years	0.0	0.96	1.60	0.0	1.47	2.45	0.0	1.91	3.18
10 years (150DB)	0.0	1.11	1.86	0.0	1.70	2.83	0.0	2.19	3.65
15 years	0.0	1.58	2.64	0.0	2.34	3.89	0.0	2.93	4.88
20 years	0.0	2.00	3.33	0.0	2.87	4.78	0.0	3.51	5.85

Source: Authors' calculations based on statutory MACRS recovery schedules, 0.3425 corporate tax rate, and 0.2975 distribution tax rate.

Source: House-Shapiro (2008)

Result: Relative increase in long-lived investment



Source: House-Shapiro (2008)

Accelerated depreciation (House-Shapiro 2008)

- Clearest finding: Large increase in investment, on average monotonically related to subsidy
- Interpretation: Very elastic investment supply (cf. Goolsbee 1998) and high internal adjustment costs
- Questions:
  - Why do investment effects persist after 2004?
  - What is the implied cost-of-capital elasticity of investment?

Accelerated depreciation (Zwick-Mahon 2014)

- House-Shapiro: Only friction to intertemporal optimization is internal adjustment cost

- Huge and contentious corporate finance literature (starting with Fazzari-Hubbard-Petersen 1988): Firms face financing constraints (a liquidity effect)

- Zwick-Mahon: Accelerated depreciation has large effect on financing constraints $\rightarrow$ perhaps explains effects on investment, rather than intertemporal substitution

Accelerated depreciation (Zwick-Mahon 2014)

- House-Shapiro's modest subsidies: accelerated depreciation increases depreciation deductions (and thus lowers tax payments) now at expense of future deductions (and thus higher tax payments) $\rightarrow$ modest subsidy (0.75-2% for five-year property) due to discounting

- Financing: Firm must pay up front for machine that pays off over time

- Financing constraint acts like high discount rate: Cash now is very valuable relative to cash later

- Accelerated depreciation generates large effective subsidy if firm is constrained

Large reduction in current taxes

Table 2: Regular and Bonus Depreciation Schedules for Five Year Items

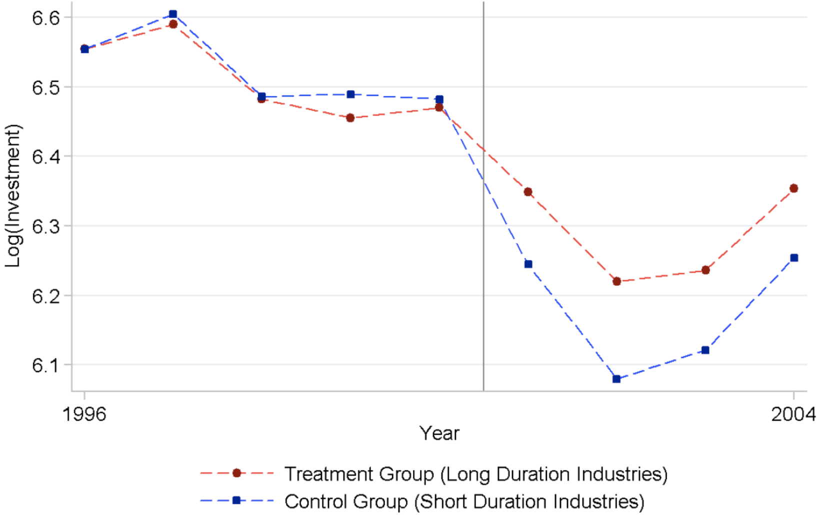
Normal Depreciation							
Year	0	1	2	3	4	5	Total
Deductions (000s)	$\boxed{200}$	320	192	115	115	58	$\boxed{1000}$
Tax Benefit ($\tau = 35\%$)	$\boxed{70}$	112	67.2	40.3	40.3	20.2	$\boxed{350}$
Bonus Depreciation (50%)							
Year	0	1	2	3	4	5	Total
Deductions (000s)	$\boxed{600}$	160	96	57.5	57.5	29	$\boxed{1000}$
Tax Benefit ($\tau = 35\%$)	$\boxed{210}$	56	33.6	20.2	20.2	10	$\boxed{350}$

Source: Zwick-Mahon (2014)

Accelerated depreciation (Zwick-Mahon 2014)

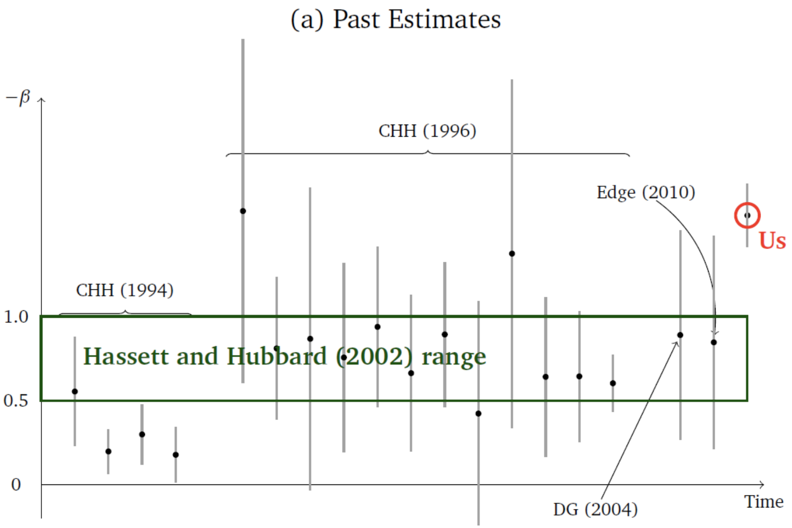
- Reduced-form effect: Compare investment across industries specializing in different asset types (computers vs. furnaces)
- Testing for financing constraints: Split firms by ex-ante markers of financing constraints (size, dividend payments, cash)
- Testing for interaction with managerial myopia: Split firms by "tax loss position,"i.e. whether they have to wait to recoup tax benefits (thus loosening constraints next year but not this year)

Large reduced-form effect



Source: Zwick-Mahon (2014)

Reconciliation with past estimates: financing constraints?

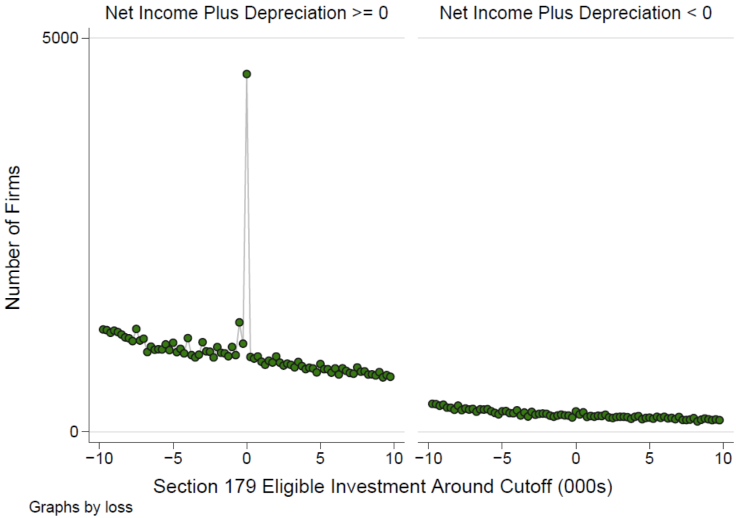


Source: Zwick-Mahon (2014)

Reconciliation with past estimates: financing constraints?

Source: Zwick-Mahon (2014)

Managerial myopia too?



Source: Zwick-Mahon (2014)

Accelerated depreciation (Zwick-Mahon 2014)

- Large absolute effect: 17-30% (again assuming no substitution across industries)
- Large cost-of-capital elasticity of investment (-1.7), using conventional cost-of-capital formulas
- Evidence of financing constraints and managerial myopia mattering for investment effects of taxes (implied discount rate of 97% for financially constrained firms!)

Payout taxes

- So far: considered only annual business income taxes
- United States has "double taxation": taxes can be assessed also when net-of-income-tax profits are distributed (paid out) to shareholders
  - Dividends: paid pro rata to all shareholders (taxed at dividend tax rate)
  - Share buyback: paid out to shareholders who sell (taxed at capital gains tax rate)
  - Retained earnings: effectively paid out when shareholder sells (taxed at accrued capital gains tax rate $<$ statutory capital gains tax rate)

Traditional view of dividend taxation

- Marginal investments are funded out of equity (Harberger 1962, 1966; Feldstein 1970; Poterba-Summers 1985) or risky debt (that can be converted to equity in bankruptcy)
    - Ex: start-ups (must issue equity in order to invest)

- In this case: τ^{DIV} is equivalent to τ^{INC}

- For compactness, ignore depreciation, uncertainty (so that $\rho = r$, fixed world interest rate), changing capital prices, and adjustment costs. Firm chooses K such that:

$$\left(1 - \tau^{DIV}\right)\left(1 - \tau^{INC}\right) F'\left(K\right) = r$$

New view of dividend taxation

- Marginal investments are funded out of retained earnings (King 1977; Auerbach 1979; Bradford 1981) or riskless debt (never converted to equity in bankruptcy)
    - Ex: Microsoft (abundant past profits from existing operations)

- In this case: permanent changes in τ^{DIV} affect value but not investment:

$$\left(1-\tau^{DIV}\right)\left(1-\tau^{INC}\right)F'\left(K\right)=\left(1-\tau^{DIV}\right)r$$

    - Firm retains cash for investment $\left(PASTPROFITS - PAYOUTS\right)$ up to point where $\left(1-\tau^{INC}\right)F'\left(PASTPROFITS - PAYOUTS\right)=r$, regardless of τ^{DIV}
    - Change in τ^{DIV} affects marginal return on investment (LHS) by the same factor that it changes the opportunity cost of investment (RHS)

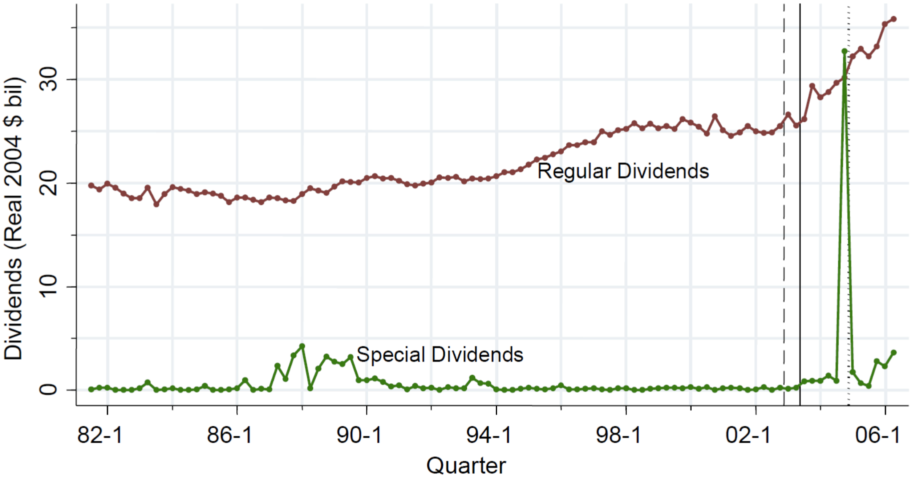
Evidence

- Original: inspect goodness of structural investment models (Poterba-Summers 1984) or cross-sectional behavior of investment and dividends (Auerbach-Hassett 2002)
- 2000s: Ignore investment and see what can be learned from payout behavior (Chetty-Saez 2005)
- 2010s: Quasi-experiments on investment (Yagan 2015)

Chetty-Saez (2005)

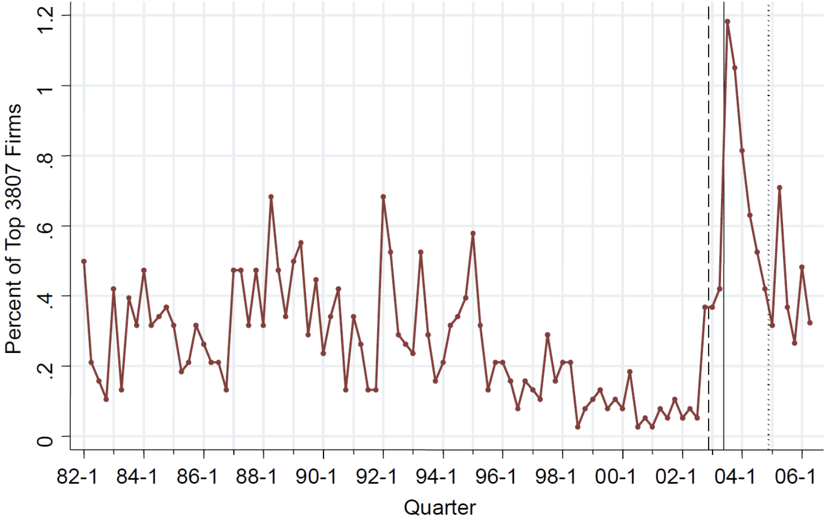
- Analyze 2003 dividend tax cut: reduced top τ^{DIV} from 38.6% to 15%
- Design:
  - Basic effect: single diff in aggregate time series (only possible because dividend initiations are high-frequency outcome, unlike investment)
  - Mechanisms: DD across firms
- Results:
  - No ringing endorsement of either traditional or new view
  - But suggests that agency considerations (imperfect monitoring of managers by owners) matter

Effect of 2003 dividend tax cut on dividend payouts



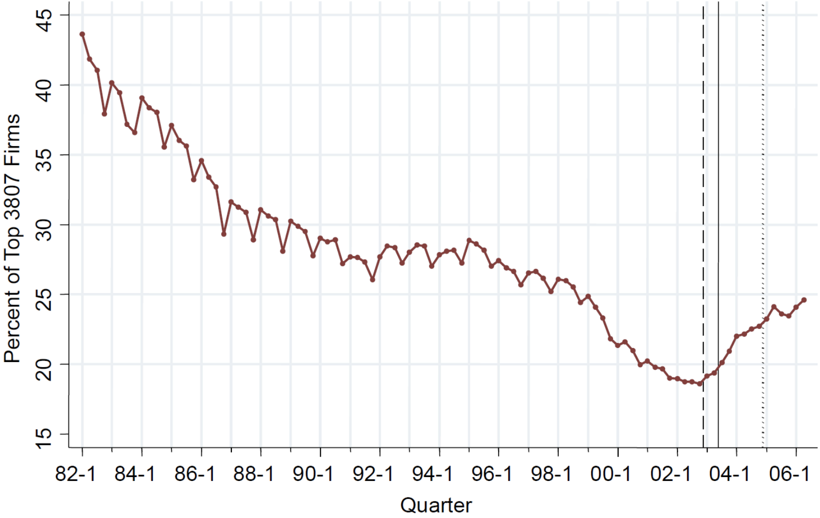
Source: Chetty-Saez (2005), updated through 2006

Effect of '03 div. tax cut on initiations of regular dividends



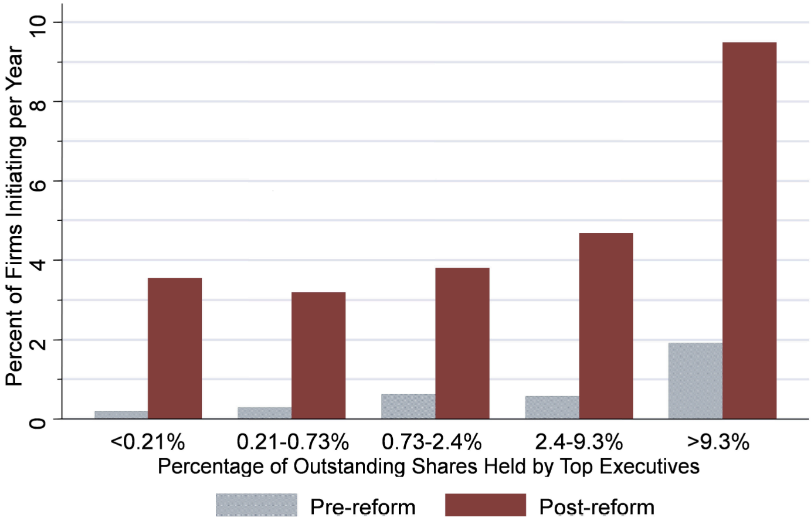
Source: Chetty-Saez (2005), updated through 2006

Effect of 2003 dividend tax cut on dividend-paying fraction



Source: Chetty-Saez (2005), updated through 2006

Heterogeneity suggestive of agency problems

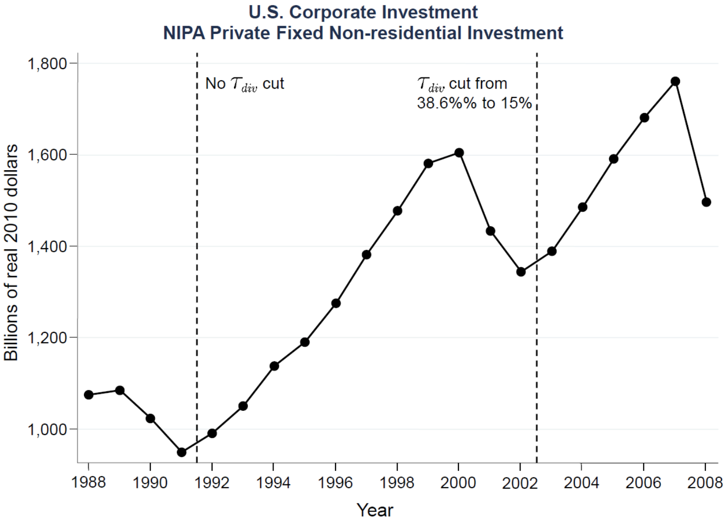


Source: Chetty-Saez (2005)

Yagan (2015)

- Chetty-Saez results consistent with positive, negative, or zero effect on investment
- Key challenge for identifying investment effects: must control for business cycle
- Design:
  - DD between C-corporations (directly affected by 2003 dividend tax cut) and S-corporations (not directly affected because never subject to dividend taxation)
- Results:
  - Zero effect that rejects basic traditional view
  - Alternative dividend tax cuts unlikely to have substantially larger effects (either new view is largely correct, or traditional view channels are inoperative in practice)

Must control for business cycle



Source: Yagan (2015)

Yagan (2015)

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

	Tax rate on annual income	Tax rate on dividends
C-corporations (treatment)	35%	15%
S-corporations (control)	35%	0%

- S-corporations: < 100 non-institutional investors, one stock class
- Operate in same narrow industries and at the same scale throughout United States $\rightarrow$ common trends

Example: Retail hardware chains

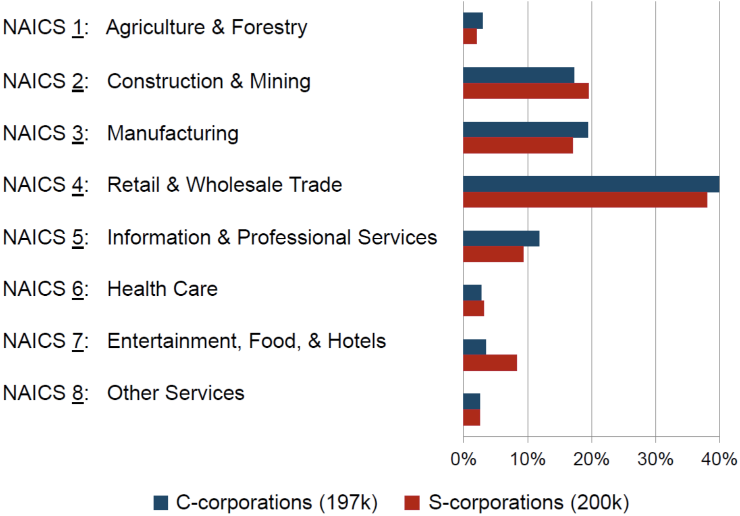
- Largest hardware chain
- C-corporation

- Third-largest hardware chain
- S-corporation

Source: Yagan (2015)

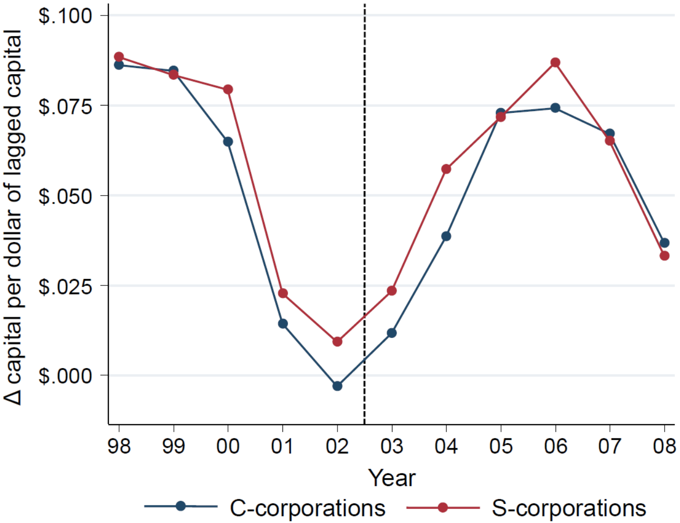
Source: Yagan (2015)

Balanced across industries and size in \$1m-\$1bn size range



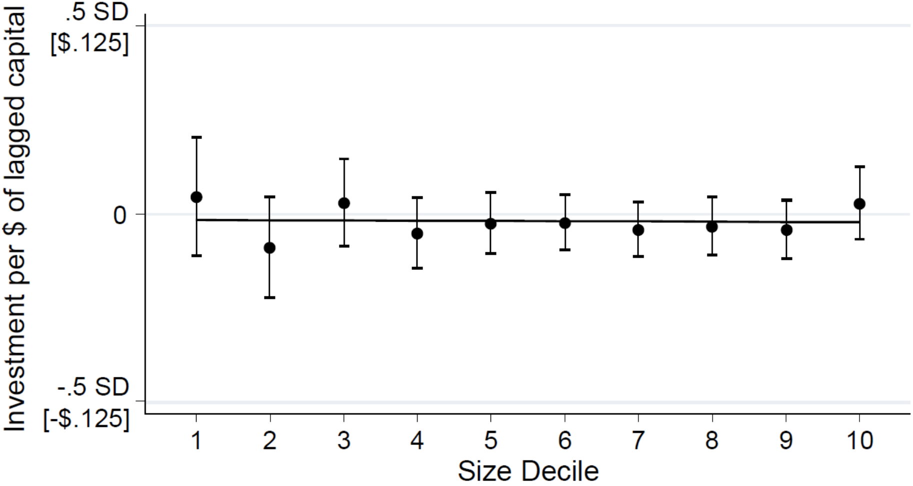
Source: Yagan (2015)

Zero effects on investment and employee compensation



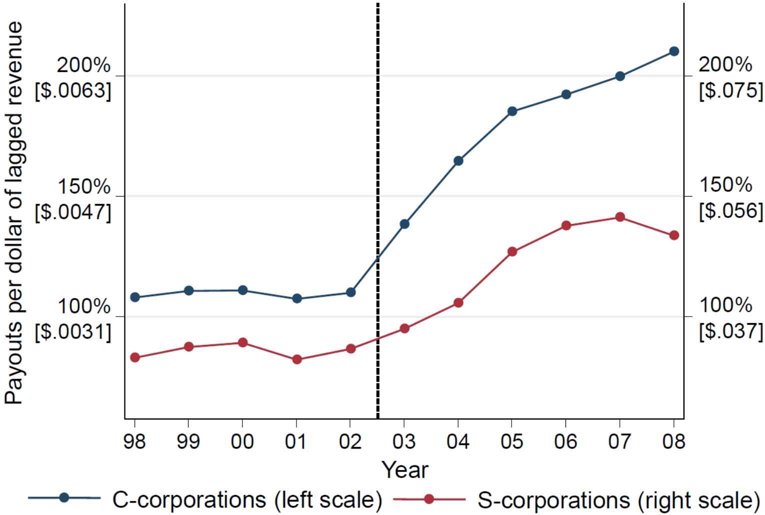
Source: Yagan (2015)

Effects constant across firm size distribution



Source: Yagan (2015)

Immediate financial response confirms relevance/salience



Source: Yagan (2015)

Yagan (2015)

- Net-of-dividend tax elasticity of investment: 0.00, with 0.08 95% confidence upper bound
- Traditional view prediction: $[0.21, 0.41]$ depending on cost-of-capital elasticity of investment (based on Hassett-Hubbard consensus range)

Yagan (2015)

- One explanation: New view is correct and most firms fund marginal investments out of retained earnings (e.g. median firm is 22 years old) $\rightarrow$ perhaps sizeable effect in very long-run when Facebook/Twitter take over U.S. production
- Alternative: Traditional view is technically correct, but tax code features blocked effects
  - Ex: Low expected permanence (originally set to expire in 2009)
  - But most investment is in short-lived assets (so six years is effectively forever)
  - And governments never commit to long-run path for tax policy: dividend tax cut has largely outlasted many "permanent" reforms, and four of the G-7 countries have substantially changed their dividend tax rates in last 10 years

Effective tax rates on business income

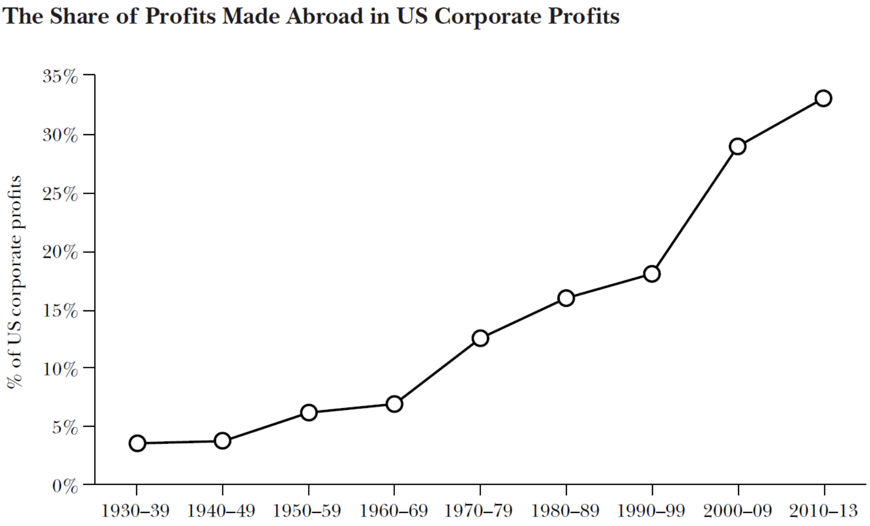
- So far: Only way to lower all-in effective tax rate on business income is to change asset mix (from long-duration to short-duration), change organizational form (e.g. from C to S), change form and timing of payouts to shareholders

- Methods available to multinationals
  - Transfer pricing: Develop property in (or sell property at low price to) foreign subsidiary, which then leases it at high price to domestic parent $\rightarrow$ domestic parent enjoys cost deductions while foreign sub pays little tax on lease earnings
  - Earnings stripping: Domestic parent borrows heavily from foreign sub in Caymans $\rightarrow$ domestic parent enjoys interest deductions while foreign sub pays little tax on interest earnings

Ex: Microsoft's Irish sub Round Island Inc.

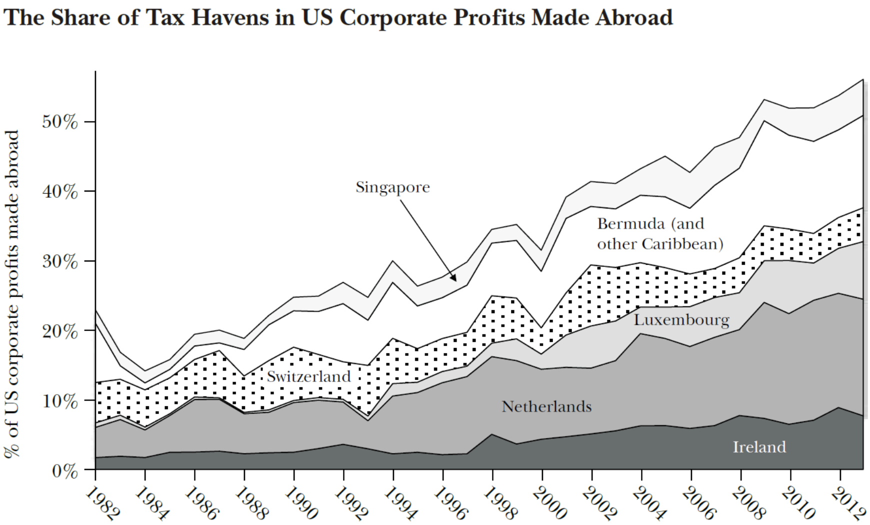
- In 2005: 1.5% of employees, 23% of assets
- Microsoft 2004 average tax rate: 33%
- Microsoft 2005 average tax rate: 26% due to "foreign earnings taxed at lower rates"

Rising importance of earnings booked abroad



Source: Zucman (2014)

Rising importance of earnings booked abroad



Source: Zucman (2014)

Worldwide taxation and repatriation

- U.S. worldwide tax: corporate earnings are taxable upon repatriation (sending profits back to U.S. parent) at 35%, less foreign taxes paid

- Hines-Rice (1994): Doesn't matter, firm avoids entire U.S. tax if it just invests abroad at r^* (fixed worldwide interest rate) and repatriates the earnings:

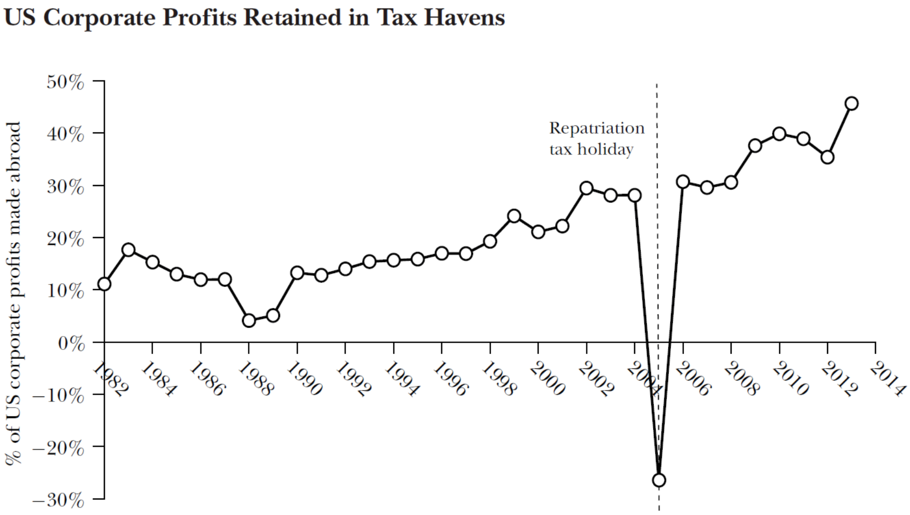
$$\sum_{T=1}^{\infty} \frac{r^* \left(1-\tau^{INC}\right)}{\left[1+r^*\left(1-\tau^{INC}\right)\right]^T} = 1$$

- Crucial assumption: firms discount future at $r^* \left(1-\tau^{INC}\right)$

- Summers (1987) survey of Fortune 200 CFOs: average discount rate of 17%

(see also Poterba-Summers 1995)

Sensitivity of repatriations to tax rate on repatriations



Source: Zucman (2014)

Desai-Foley-Forbes (2009)

- Tax holiday promoted to increase domestic investment. What did firms do with the repatriated funds?
- Design: compare investment changes across firms with different tax-haven profit concentration
- Findings: firms returned almost all money to shareholders, no direct increase in investment
- Intuition: firms can borrow against their foreign earnings, so little reason to be financially constrained in the first place

International considerations and policy

- Prescription from neoclassical cost-of-capital model: narrow base and then increase rate as much as you want
- Apparent policy consensus: leave base broad, lower the rate
- One rationalization: large perceived costs to corporations with rents moving headquarters abroad
- But are advocates trying to have it both ways?
    - "Don't tax corporations: capital is internationally mobile, so corporate taxes reduce U.S. capital accumulation, wages, and GDP!"
    - "But don't tax savings either: capital is *not* interntionally mobile, so savings taxes reduce U.S. capital accumulation, wages, and GDP!"
    - (Vice versa for Democrats)

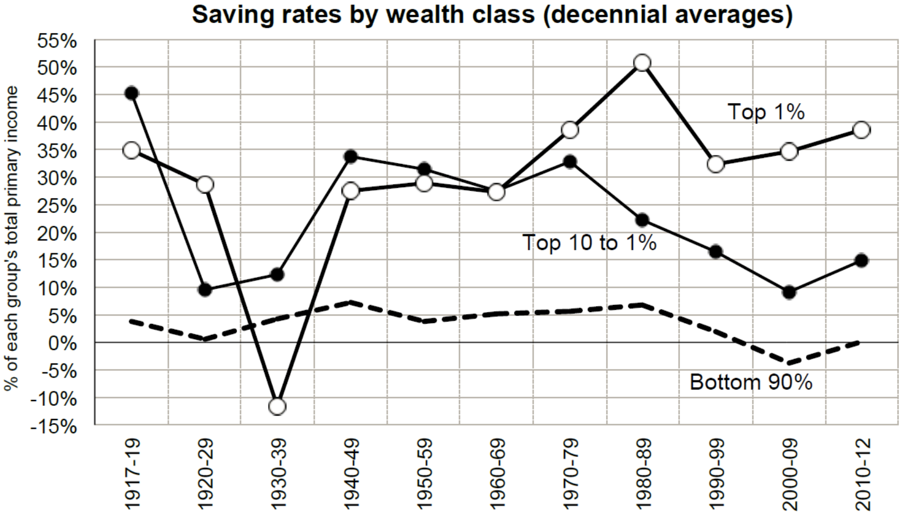
Graduate Public Economics

Lecture 3

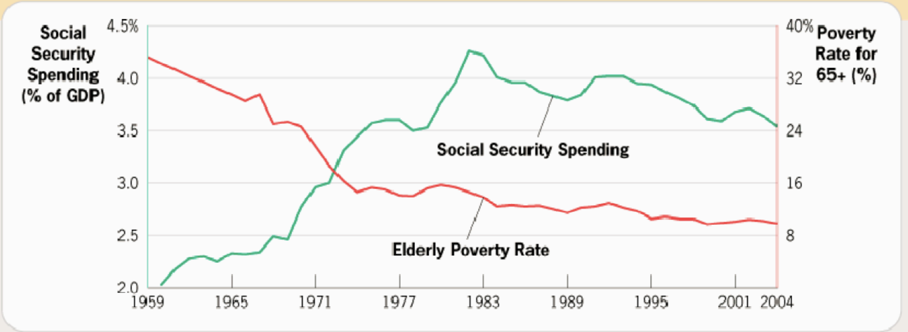
Actual Capital Taxation – Savings

Danny Yagan
UC Berkeley

Motivation: Low savings rates among non-rich



Source: Saez-Zucman (2014)



Elderly Poverty and Social Security, 1959–2004 • There is a striking negative correspondence over time between the poverty rates of the elderly (which have fallen) and the size of the Social Security program (which has risen).

Source: Gruber (2007)

Motivation: Savings matters for investment and wages?

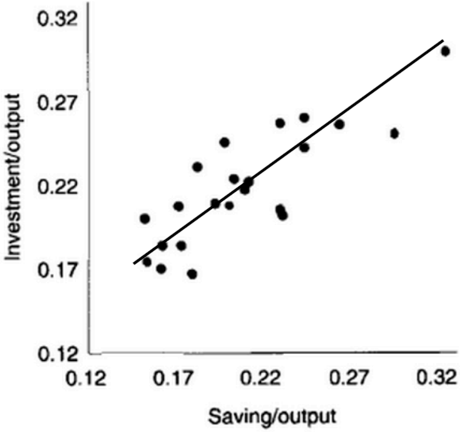


Figure 3.4
Industrial-country saving and investment rates, 1982–91

$\beta = .62$
(.09)

Source: Obstfeld-Rogoff (1996), Feldstein-Horioka (1980)

How do taxes and other policy affect savings?

- Insitutions: Taxes and tax exemptions

- Level of savings for typical households
  - Theory: Taxes and mandates
  - Evidence: Experiments and quasi-experiments

- Allocation and turnover of typical households' assets

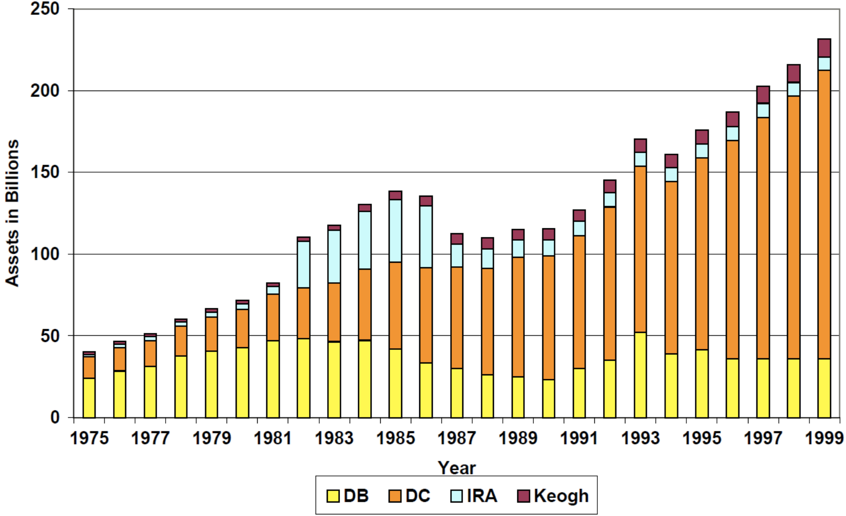
- Wealth concentration and savings of the super rich

Institutions: Taxes and tax exemptions

- Baseline taxes on capital income:
  - Annual business income tax (federal+state: 39%-46%)
  - Dividend income tax (federal+state: 30%)
  - Capital gains tax (federal+state: 29%)
  - Estate tax (federal: 40%)

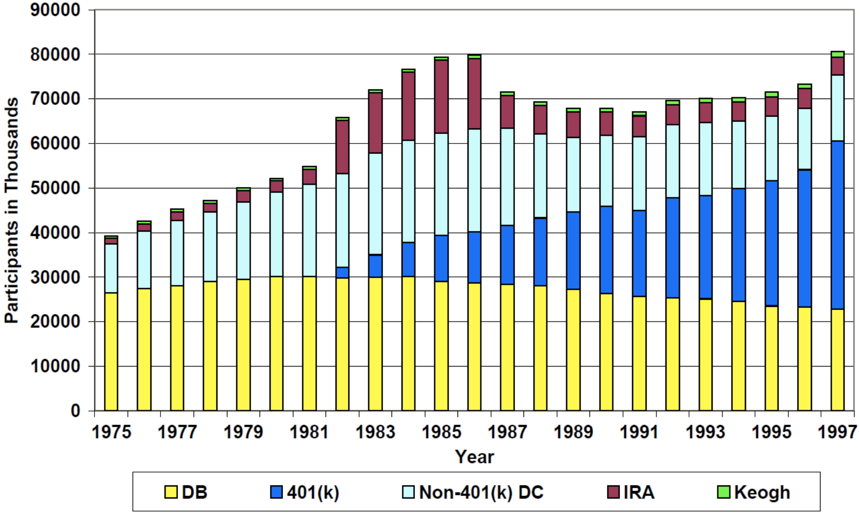
- Tax exemptions for non-rich
  - Dividend and capital gains taxes: Tax-preferred savings vehicles and Social Security
  - Estate tax: $5.5m exemption, full spouse exemption

Assets in non-Social-Security pension plans



Source: Poterba-Venti-Wise (2004)

Participation in non-Social-Security pension plans



Source: Poterba-Venti-Wise (2004)

Defined-contribution tax-preferred savings vehicles

- Defined contribution (DC) plans
  - Saver contributes and bears full risk
  - Contrasts with defined benefit (DB) plans that are in decline (and resulted in numerous employer and sovereign defaults)

- Individual Retirement Account
  - All workers eligible to contribute out of labor earnings (just call up your bank)
  - Contribution limit: $5.5k per year (indexed to inflation)
  - Taxed only on withdrawal (traditional IRA) or only on contribution (Roth IRA)

- 401(k) account
  - Workers eligible only if employer sponsors a plan (employer deducts contributions from wages)
  - Contribution limit: $17.5k per year
  - Taxed only on withdrawal (10% fee for withdrawal before age 59.5)

Comment: Back-end vs. front-end taxes

- Year-T value of \$1 of savings in year-0:
  - No taxes: $V_{NT} = (1 + r)^T$
  - Back-end taxes only (401(k) and traditional IRA): $V_{NT} = (1 + r)^T \left(1 - \tau_T^{PERSONAL}\right)$
  - Front-end taxes only (Roth IRA): $V_{NT} = \left(1 - \tau_0^{PERSONAL}\right)(1 + r)^T$
  - Regular taxable account: $V_{NT} = \left(1 - \tau_0^{PERSONAL}\right)\left(1 + r\left(1 - \tau^{DIV}\right)\right)^T$

- Choose front-end taxes over back-end taxes if you expect $\tau_T^{PERSONAL} > \tau_0^{PERSONAL}$ (e.g. contribute to Roth when you're a grad student, contribute to a 401(k) or IRA when you're a B-school professor)

Neoclassical theory

- Classic two-period consumption model:

$$\max_{c_1, c_2} u\left(c_1, c_2\right)$$

$$\text{s.t. } c_1 + c_2 / \left(1 + r\right) \leq z_1$$

- With capital tax, budget constraint becomes:

$$c_1 + c_2 / \left(1 + r\left(1 - \tau_k\right)\right) \leq z_1$$

- With 401(k)/IRA, budget constraint becomes kinked

Neoclassical theory

- Substitution effect: 401(k)/IRA $\Rightarrow$ price of $c_2 \downarrow$ relative to $c_1 \Rightarrow c_2 \uparrow$
- Wealth effect: 401(k)/IRA $\Rightarrow$ richer if already saving $\Rightarrow$ ambiguous effects on c_1 and c_2
- Subsitution effect applies only if saving below the maximum contribution threshold
- Wealth effect could easily be negative (e.g. reduce savings even if increase c_2)

Theory: Employer match is a huge subsidy

- Many employers match 401(k) contributions up to a certain level, e.g. 50% of contributions up to 6% of salary
  - Legally required to have similar participation rates across pay scale
  - May also generate advantageous selection (attracts patient workers)
  - Workers may also demand ex-ante incentives to save
- Functions as huge subsidy (e.g. $\tau_k = -.5$)

Summary of neoclassical theory

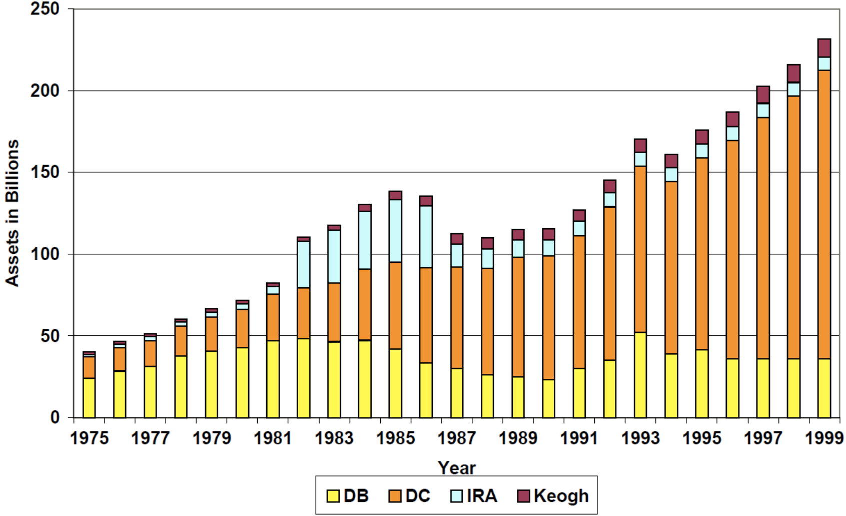
- 401(k)/IRA has ambiguous effects in general, but most likely to increase savings among poor and those with low savings rates
- 401(k) contribution rates should be high up to the match limit
- All that matters is the budget set
- Individuals differ only in preferences over (c_1, c_2)

Evidence on price effects

- Clear evidence of rise in contributions to tax-preferred accounts

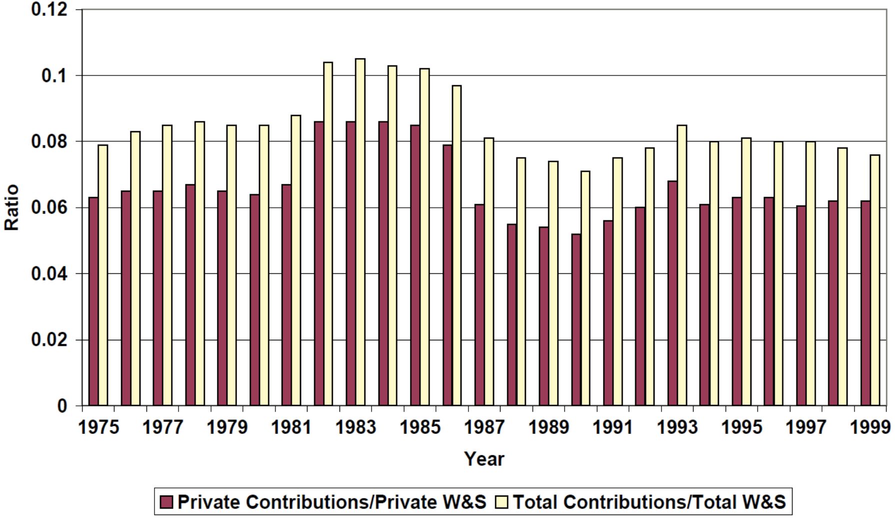
- Unclear whether this represents an increase in net savings (Poterba-Venti-Wise 1996 vs. Engen-Gale-Scholz 1996)
    - May merely substitute from taxable account to IRA or to 401(k)
    - Problem: low-quality individual-level wealth data in the U.S. and unknown counterfactual

Rising assets in private plans, especially 1982-1986



Source: Poterba-Venti-Wise (2004)

Stable ratio of pension contributions to wages



Source: Poterba-Venti-Wise (2004)

Match experiment (Duflo-Gale-Liebman-Orszag-Saez '06)

- H&R Block randomly assigned 14,000 tax filers to H&R-Block IRAs: 0% match, 20% match, or 50% match

- Compare results to Saver's Credit (non-experimental federal policy)
  - Non-refundable tax credit on first \$2,000 of contributions for people earning below AGI thresholds
  - Credit rate t is equivalent to match rate $t/\left(1-t\right)$
  - Most dramatic schedule change: effective match rate falls from 100% to 25% at \$30,000 AGI

Match experiment (Duflo-Gale-Liebman-Orszag-Saez '06)

- Experimental match has large effect:
  - 0% match: 3% contribution rate
  - 20% match: 8% contribution rate
  - 50% match: 14% contribution rate

- Saver's Credit has much smaller effect (e.g. 1.3 percentage points going form 25% match to 100% match)

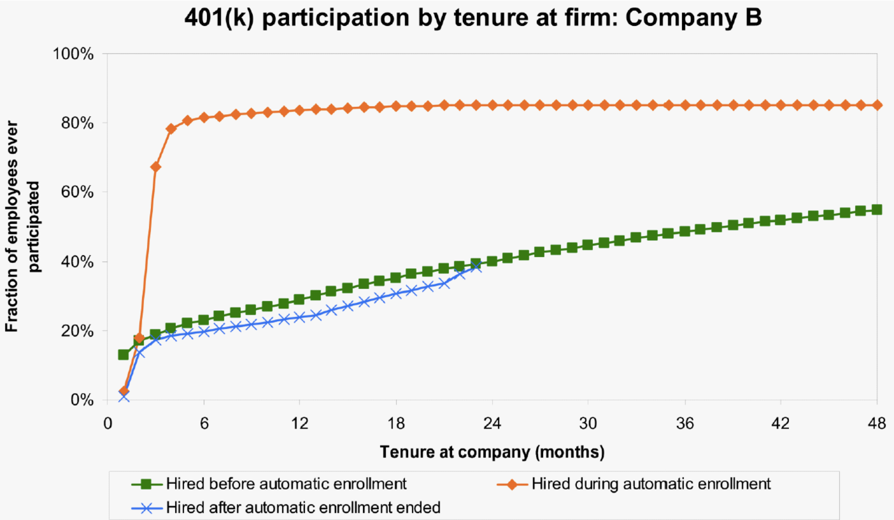
- Prices matter

- People are confused
  - Saver's Credit is confusing
  - 92% of people offered the 20% match rate declined a free lunch (contributing, getting matched, and then immediately withdrawing) (see also Choi-Laibson-Madrian 2011)

Madrian-Shea (2001)

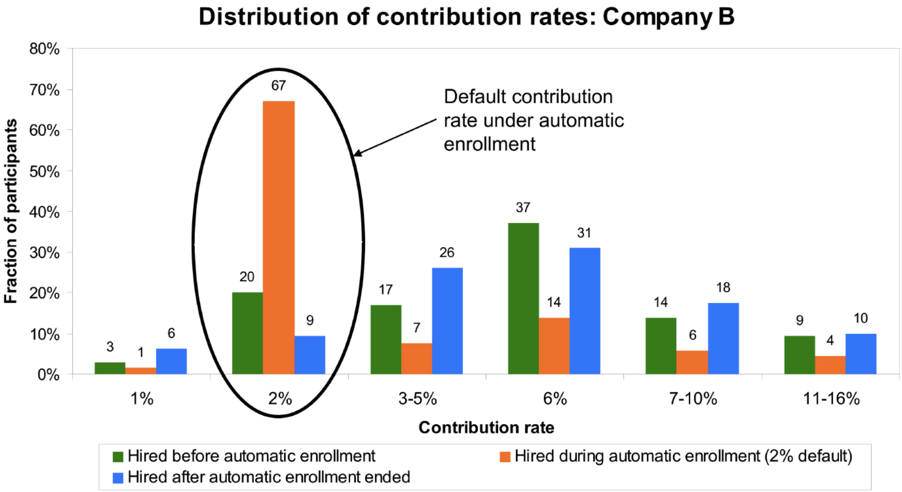
- Neoclassical theory: behavior depends only on budget set
- Analyze large firm that changed its default 401(k) contribution rate for workers
  - Original: 0% default
  - Internal policy change: New hires automatically enrolled (defaulted) to 3%
  - Internal policy reversion: Removal of automatic enrollment

Effect of automatic enrollment on 401(k) participation



Source: Madrian-Shea (2001)

Effect of automatic enrollment on contribution rate

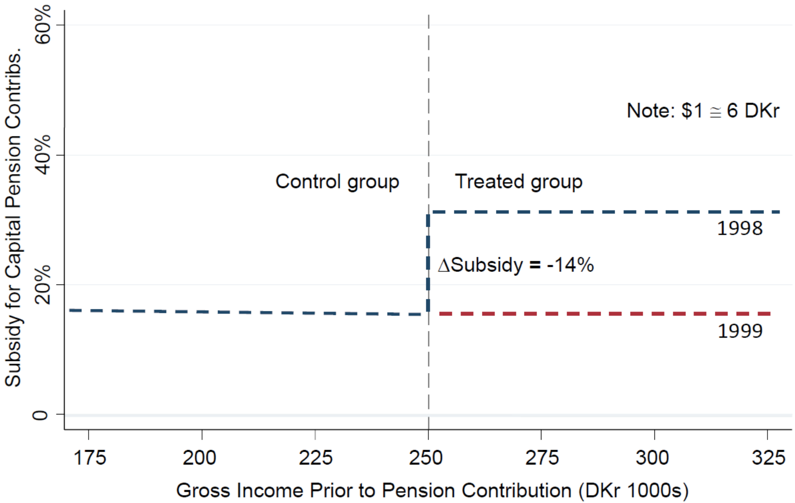


Source: Madrian-Shea (2001)

Chetty–Friedman–Leth-Petersen–Nielsen–Olsen (2014)

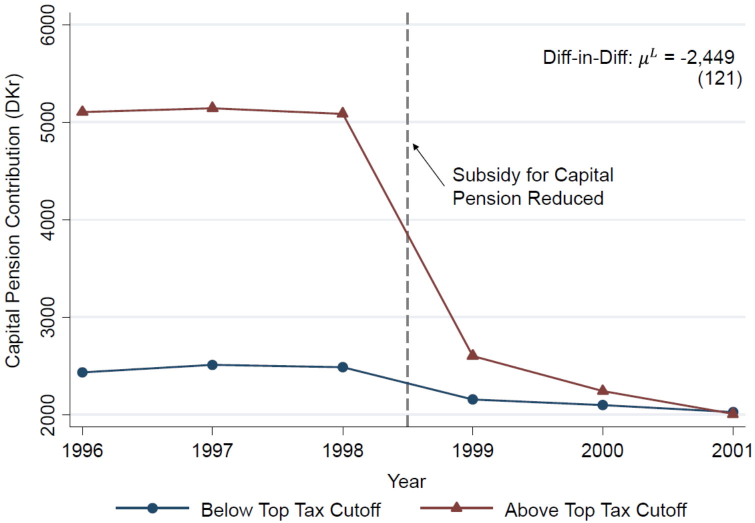
- Administrative data from Denmark
  - Third-party-reported wealth data (legacy from wealth tax)
  - Matched employer-employee data
- Two types of voluntary tax-preferred savings accounts: "capital pensions" (lump sump at retirement) and "annuity pensions" (annuitized at retirement)
- Neoclassical effects: 1999 reform reduced subsidy for saving in capital pension by 12 percentage points, only for people in top tax bracket

1999 capital pension subsidy reform



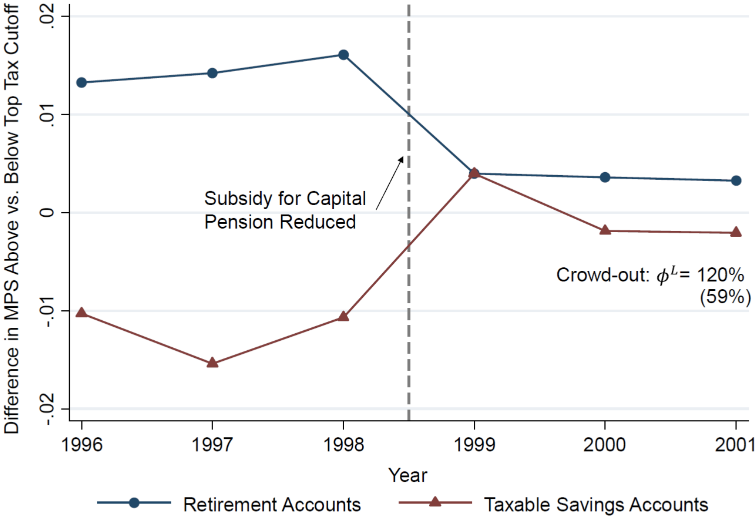
Source: Chetty–Friedman–Leth-Petersen–Nielsen–Olsen (2014)

Impact of 1999 capital pension subsidy reform



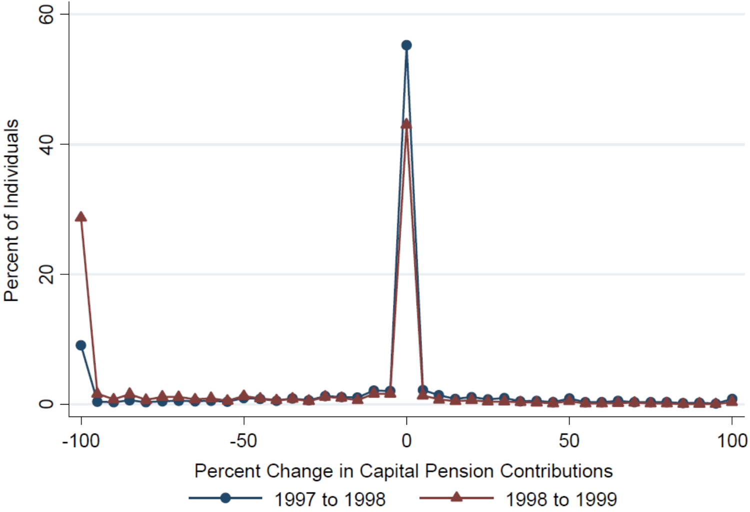
Source: Chetty–Friedman–Leth-Petersen–Nielsen–Olsen (2014)

Complete crowd-out



Source: Chetty–Friedman–Leth-Petersen–Nielsen–Olsen (2014)

Active vs. passive savers

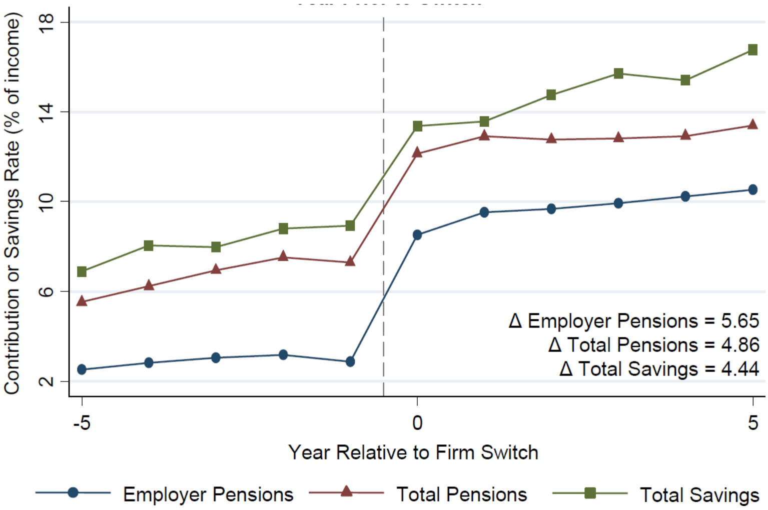


Source: Chetty–Friedman–Leth-Petersen–Nielsen–Olsen (2014)

Chetty–Friedman–Leth-Petersen–Nielsen–Olsen (2014)

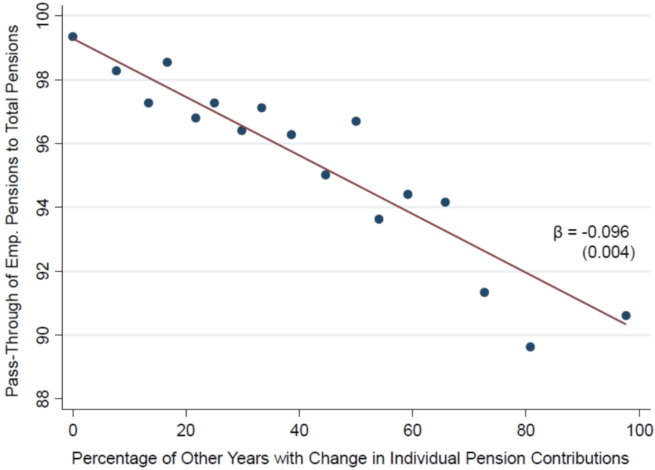
- Subsidy reduction (neoclassical mechanism) $\rightarrow$ no effect on savings, only rise in tax payments
- But retirement plans feature more than just subsidies
- Study automatic enrollment using firm switchers design
    - Neoclassical model: full offset for those not at corner

Event study when moving to firm with >3% match incr.



Source: Chetty–Friedman–Leth-Petersen–Nielsen–Olsen (2014)

Active vs. passive savers

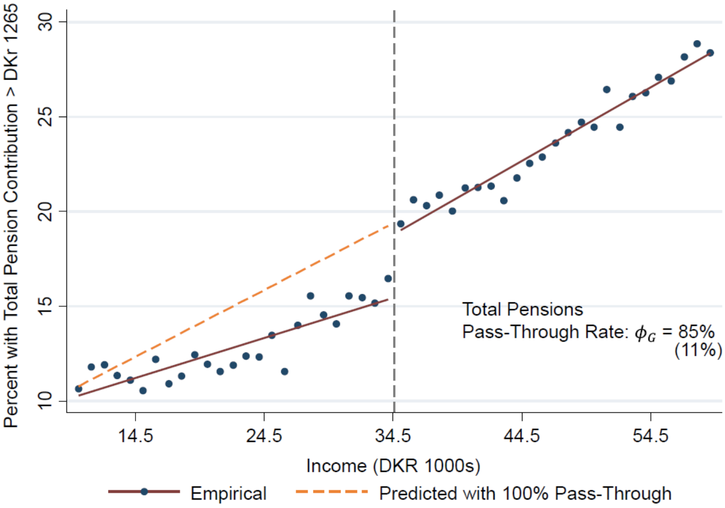


Source: Chetty–Friedman–Leth-Petersen–Nielsen–Olsen (2014)

Social Security

- Forced savings:
  - 6.2% on first \$118,500, both by employee and employer
  - Progressive schedule for Primary Insurance Amount (PIA, i.e. monthly annuity amount) based on Average Indexed Monthly Earnings (AIME, equal to average over worker's highest 35 years of earnings, including zeros)
  - But high-income live longer
- Effect in standard model with actuarily fair Social Security: full offset for high-savers, zero offset for low-savers
- Denmark: Increase of 1% in required contribution at income cutoff

Effect of 1% increase in required contribution



Source: Chetty–Friedman–Leth-Petersen–Nielsen–Olsen (2014)

Chetty–Friedman–Leth-Petersen–Nielsen–Olsen (2014)

- Roughly 85% of savers are passive savers, 15% are active savers
- Active savers tend to be wealthy and sophisticated
- Lessons
  - Standard economists' tools (prices) may be dwarfed by non-standard tools
  - High impact using foreign data: qualitative lessons more than quantitative lessons

What does "passive saving" imply about behavior?

- Individual budget constraint must hold: consumption + savings = income
- If people have fixed consumption plans, then higher default contributions would have displaced savings in other accounts (or increased debt)
- Instead, it appears that consumption is a residual for most people (at current contribution rates)
  - Intuition: At end of year, people look at their bank accounts to see if they can afford a vacation or new car

Why have Social Security

- Adverse selection (Eckstein-Eichenbaum-Peled 1985)
- Optimal disability insurance (Diamond-Sheshinski 1995)
- Myopia (Feldstein 1985;
  Beshears-Choi-Clayton-Harris-Laibson-Madrian 2015)

Graduate Public Economics

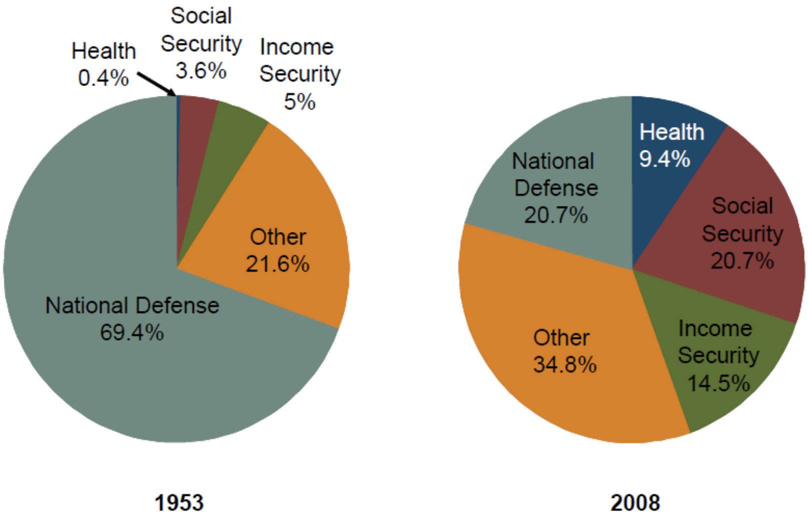
Lecture 4

Social Insurance

Danny Yagan
UC Berkeley

Note: Several slides from Chetty-Bruich

Motivation: Modern governments do social insurance



Source: Chetty; Office of Management and Budget, historical tables

Motivation: Developed governments do social insurance

Source: Chetty-Finkelstein (2013) (1996 data)

Motivation: Redistribution framed as social insurance

- Typically think of optimal tax-and-transfers as separate from social insurance
- Optimal tax: heterogeneous-wage individuals have heterogeneous marginal utilities
- Optimal social insurance: identical individuals face uncertain future marginal utilities
  - I may be laid off more than average (unemployment insurance)
  - I may get sicker than average (health insurance)
  - I may live longer than average (social security)

Motivation: Redistribution framed as social insurance

- But optimal tax is optimal social insurance from behind veil of ignorance (Rawls 1971)
  - Mirrlees (1971): Adults have heterogeneous wages. Planner uses tax system to redistribute from high-wage to low-wage:

$$\max_{\{c_w, l_w\}} \int u(c_w, l_w) f(w) dw$$

$$\text{s.t. } \{c_w, l_w\} = \arg\max u(wl_w - T(wl_w), l_w) \;\; \forall w \quad \text{(IC)}$$

$$\text{and } \int T(wl_w) f(w) dw = 0 \quad \text{(RC)}$$

  - Alternative framing: Identical unborn individuals face same risk of being born with low w. Unborn agree on optimal feasible insurance contract (tax system) that redistributes *ex post* from high-w to low-w

Social insurance definitions

- Conventional definition of “social insurance”: Transfers based on at least partially *observable events*
  - “Event”: distinguishes from redistribution across ex ante different people (useful: people vote after veil has been lifted but before other shocks)
  - “Observable”: distinguishes from dynamic optimal tax problem (fully unobservable wage shocks in new dynamic public finance)

- Key similarity: Private information forces second best

- Key differences:
  - Events generate heterogeneity
  - Universal compulsion is not taken as given

Social insurance definitions

- "Moral hazard": agent takes private action *after* contract is set in force (Mirrlees 1971; Baily 1978; Chetty 2006)
- "Adverse selection": agent takes private action *before* contract is set in force (Akerlof 1970; Rothschild-Stiglitz 1976)

Social insurance outline

- Moral hazard with application to unemployment insurance (UI)
- Advese selection with application to health insurance
- Combined application: Disability insurance (DI) and Social Security rationales
- Application of coverage externalities: Auto insurance

Rationales for government provision

- UI is valuable for consumption
- In principle, private market could provide this insurance
- In practice, adverse selection may prevent such markets (later in lecture)
- Individuals could self-insure, but they don't: median job loser has $<$ \$200 in bank

Institutional detail: U.S. UI

- When laid-off (not fired or quit), worker can claim UI benefits
- Benefits are paid weekly (26 weeks in normal times, 99 weeks in GR) and then stop
- Benefits are a function of the worker's average weekly earnings at the job, with minimum and maximum levels
- Benefits are paid to the worker by the state in which the firm establishment is located, no matter where the worker lives
- State collects UI revenue from state establishments as share of each worker's capped payroll, with share depending on firm's lagged layoffs ("experience rating")

Example: UI benefit schedule in Michigan 2009

Source: Chetty; Michigan Dept. of Energy, Labor, and Economic Growth 2009

Goal: Model of optimal UI benefit level

- Common measure of program's size is its "replacement rate"

$$r = \frac{\text{net benefit}}{\text{net wage}} \approx 0.5 \text{ currently}$$

- UI benefit $b = rw$. Net value of returning to work $= w(1-r)$

- Moral hazard in static models (Baily 1978; Chetty 2006)

- Allow for liquidity (Chetty 2009)

Baily-Chetty model

- Canonical analysis of optimal level of UI benefits: Baily (1978)
- Shows that the optimal benefit level can be expressed as a fn of a small set of parameters in a static model.
- Once viewed as being of limited practical relevance because of strong assumptions
- Chetty (2006) shows formula actually applies with arbitrary choice variables and constraints.
- Parameters identified by Baily are sufficient statistics for welfare analysis $\Rightarrow$ robust yet simple guide for optimal policy.

Baily-Chetty model: Assumptions

1. Fixed wages – no GE effects
2. No distortions to firm behavior (temporary layoffs); implicitly assume perfect experience rating
3. No externalities such as spillovers to search

Baily-Chetty model: Setup

- Static model with two states: high (employed) and low (unemployed)
- Let w_h denote the individual's income in the high state and $w_l < w_h$ income in the low state
- Let A denote wealth, c_h consumption in the high state, and c_l consumption in the low state
- Agent is initially unemployed. Controls probability of being in the bad state by exerting search effort e at a cost $\psi(e)$
- Choose units of e so that the probability of being in the high state is given by $p(e) = e$

Baily-Chetty model: Setup

- UI system that pays constant benefit b to unemployed agents
- Benefits financed by lump sum tax $t(b)$ in the high state
- Govt's balanced-budget constraint:

$$e \cdot t(b) = (1 - e) \cdot b$$

- Let $u(c)$ denote utility over consumption (strictly concave)
- Agent's expected utility is

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

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:

$$\max_{b,e} e(A + w_h - t) + (1 - e)u(A + w_l + b) - \psi(e)$$

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

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

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:

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

Second Best Problem

Problem

Optimal Social Insurance

$$\max_b V(b, t(b))$$
$$s.t. \;\; e(b)t(b) = (1 - e(b))b$$
$$e(b) = \arg\max_e e \cdot u(A + w_h - t) + (1 - e) \cdot u(A + w_l + b) - \psi(e)$$

- Formally equivalent to an optimal Ramsey tax problem with state-contingent taxes

Two Approaches to Optimal Social Insurance

1. **Structural**: specify complete models of economic behavior and estimate the primitives

    - Identify b^* as a fn. of discount rates, nature of borrowing constraints, informal ins. arrangements.

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

Kaplan 2012

- Exploiting f.o.c.'s from agent optimization particularly useful in more complex models
- Kaplan (2009): unemployed youth move back in with their parents.
    - How does this affect optimal UI?
- Kaplan takes a structural approach and estimates a dynamic model of the decision to move back home

Sufficient Statistic Approach to Kaplan 2012

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

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

- 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

$$\frac{dt}{db} = \frac{1-e}{e} - \frac{b}{e^2}\frac{de}{db} = \frac{1-e}{e}\left(1+\frac{\varepsilon_{1-e,b}}{e}\right)$$

$$\Longrightarrow \frac{dV(b)}{db} = (1-e)\left\{u'(c_l) - \left(1+\frac{\varepsilon_{1-e,b}}{e}\right)u'(c_h)\right\}$$

- 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

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

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

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

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

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

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

Chetty 2008: Baseline Assumptions

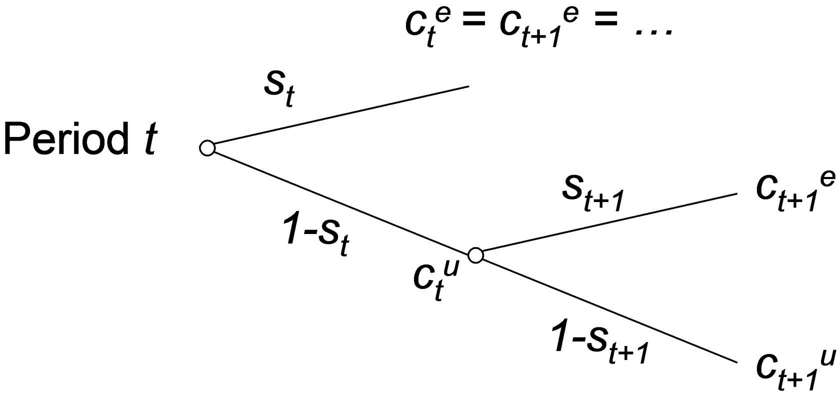
1. Assets prior to job loss exogenous
2. No heterogeneity
3. Fixed wages: choose only search intensity, not reservation wage
4. Fixed layoff probabilities (perfect experience rating)
5. No externalities (e.g. no effect of my search on your job-finding rate)

Chetty 2008: Job Search Technology

- If unemployed in period t, worker first chooses search intensity s_t
- Finds a job that begins immediately in period t with probability s_t
- If job found, consumes c_t^e. Jobs are permanent, pay wage $w_t - \tau$.

Chetty 2008: Job Search Technology

- If no job found: receives benefit b_t, consumes c_t^u, enters $t+1$ unemployed
- Cost of job search: $\psi(s_t)$



Chetty 2008: Value Functions

- Value function for agent who finds a job in period t:

$$V_t(A_t) = \max_{A_{t+1} \geq L} u(A_t - A_{t+1} + w - \tau) + V_{t+1}(A_{t+1})$$

- Value function for agent who does not find a job in period t:

$$U_t(A_t) = \max_{A_{t+1} \geq L} u(A_t - A_{t+1} + b_t) + J_{t+1}(A_{t+1})$$

where $J_{t+1}(A_{t+1})$ is value of entering next period unemployed.

- Agent chooses s_t to maximize expected utility

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

Chetty 2008: Optimal Search Behavior

- First order condition for optimal search intensity:

$$\psi'(s_t^*) = V_t(A_t) - U_t(A_t)$$

- Intuitively, s_t is chosen to equate the marginal cost of search effort with the marginal value of search effort.

- Effect of benefits on durations:

$$\partial s_t / \partial b_t = -u'(c_t^u) / \psi''(s_t)$$

Chetty 2008: Moral Hazard vs. Liquidity Decomposition

- Effect of benefits on durations can be decomposed into two terms:

$$\partial s_t / \partial b_t = \partial s_t / \partial A_t - \partial s_t / \partial w_t$$

$$\text{with } \partial s_t / \partial A_t = \frac{u'(c_t^e) - u'(c_t^u)}{\psi''(s_t)} < 0 \text{ (``liquidity effect'')}$$

$$\text{and } \partial s_t / \partial w_t = \frac{u'(c_t^e)}{\psi''(s_t)} > 0 \text{ (``moral hazard effect'')}$$

- (Technical point: need annuity version when b is extended for all periods)

- Liquidity and total benefit effects smaller for agents with better consumption smoothing capacity

Chetty 2008: Formula for Optimal UI

$$\partial s_t/\partial A_t = \{u'(c_t^e) - u'(c_t^u)\}/\psi''(s_t) \geq 0$$
$$\partial s_t/\partial w_t = u'(c_t^e)/\psi''(s_t) > 0$$
$$\Rightarrow \frac{\partial s_t/\partial A_t}{\partial s_t/\partial w_t} = \frac{\text{LIQ}}{\text{MH}} = \frac{u'(c_t^u) - u'(c_t^e)}{u'(c_t^e)}$$

- Can show that Baily-Chetty formula holds in this model:

$$\frac{u'(c_t^u) - u'(c_t^e)}{u'(c_t^e)} = \frac{\varepsilon_{1-s_t,b}}{s_t}$$

- Combining yields formula that depends solely on duration elasticities:

$$\frac{\partial s_t^*/\partial A_t}{\partial s_t^*/\partial b_t - \partial s_t^*/\partial A_t} = \frac{\varepsilon_{1-s_t,b}}{s_t}$$
$$\frac{\varepsilon_{1-e,A}}{\varepsilon_{1-e,b}\frac{A}{b} - \varepsilon_{1-e,A}} = \frac{\varepsilon_{1-s,b}}{s}$$

Intuition for Moral Hazard vs. Liquidity Formula

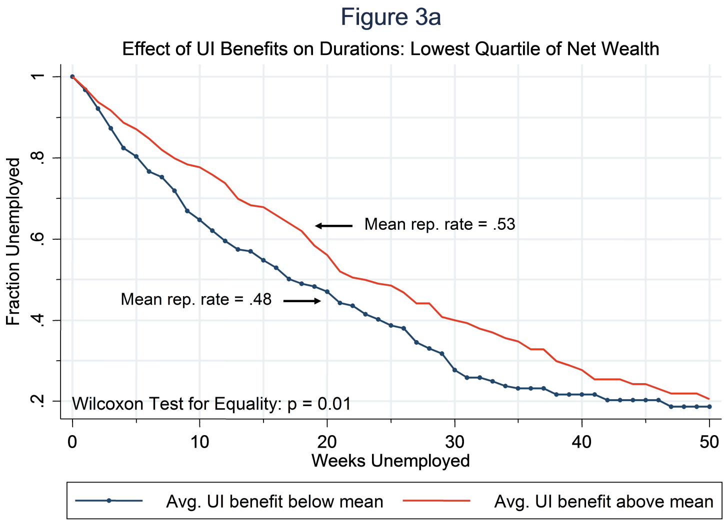
- Formula is a "revealed preference" approach to valuing insurance
  - Infer value of UI to agent by observing what he would do if money given as a cash-grant without distorted incentives
  - If agent would not use money to extend duration, infer that only takes longer because of price subsidy (moral 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

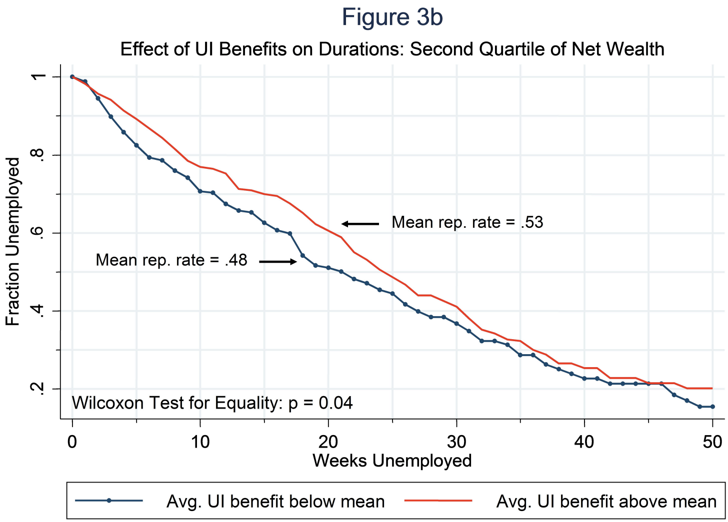
- Two empirical strategies

  1. Divide agents into liquidity constrained and unconstrained groups and estimate effect of benefits on durations using changes in UI laws.

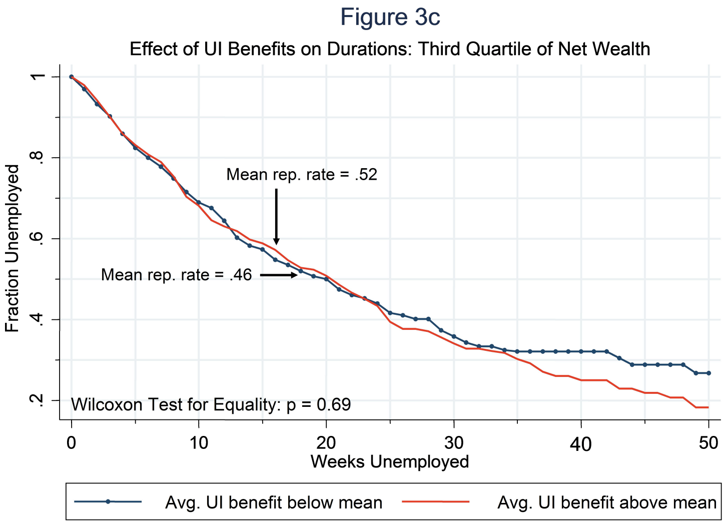
  2. Look at lump-sum severance payments to estimate liquidity effect (see also Card-Chetty-Weber 2007)



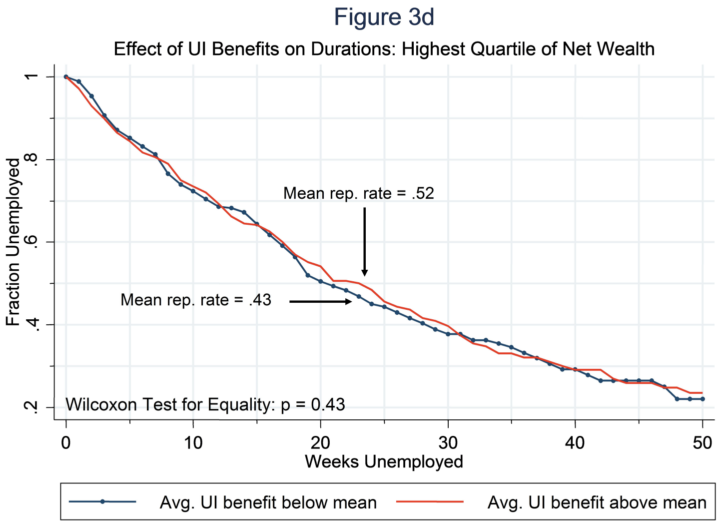
Source: Chetty 2008



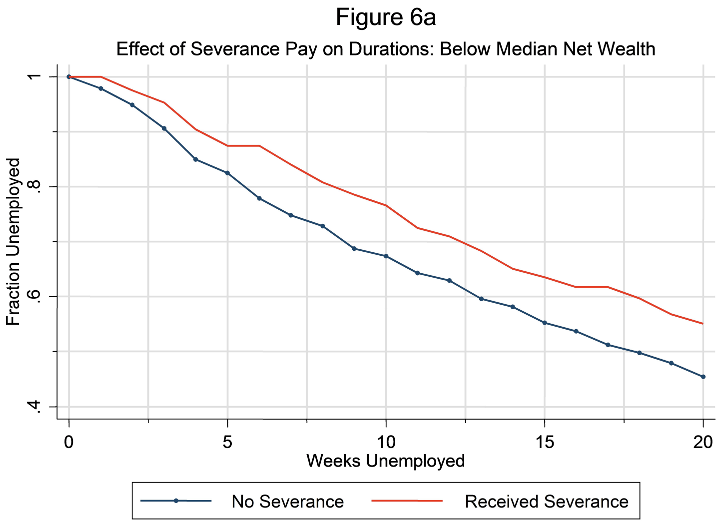
Source: Chetty 2008



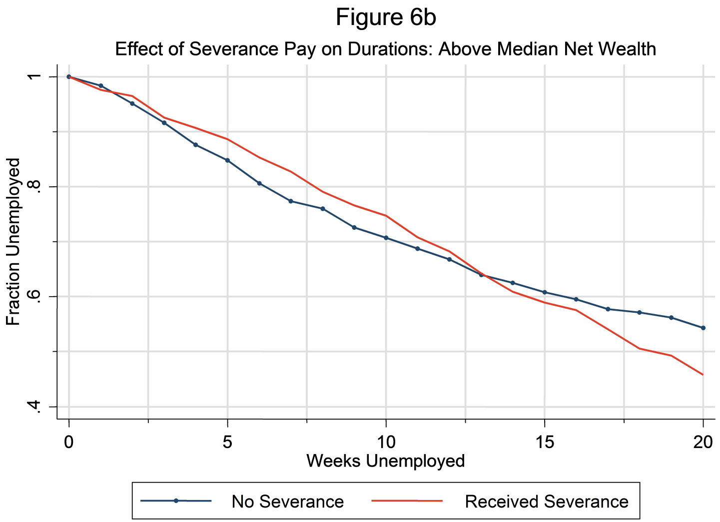
Source: Chetty 2008



Source: Chetty 2008



Source: Chetty 2008



Source: Chetty 2008

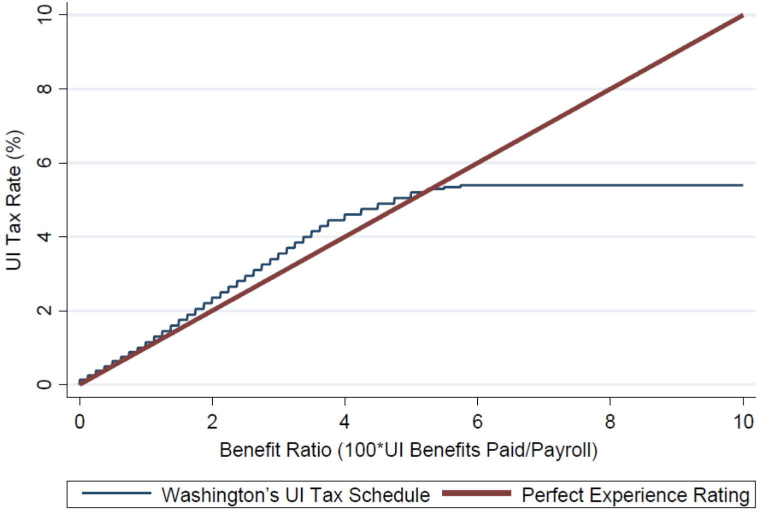
Chetty 2008: Implications for Optimal UI

- Use elasticities to compute ratio of liquidity effect to total (liquidity + moral hazard) effect (≈ 0.6) and plug in to formula for 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
- Suggests we are currently near optimal benefit level
- Ignoring liquidity effects would suggest we are way past the optimum (Baily 1978; Gruber 1997)

UI departure #1: Endogenous layoffs

- Without perfect experience rating (over half of firms according to Feldstein 1978), firms can use UI system to pay workers lower wages for same work

Imperfect experience rating in Washington State



Source: Chetty; Washington State Joint Legislated Task Force on UI Benefit Equity 2005

UI departure #1: Endogenous layoffs

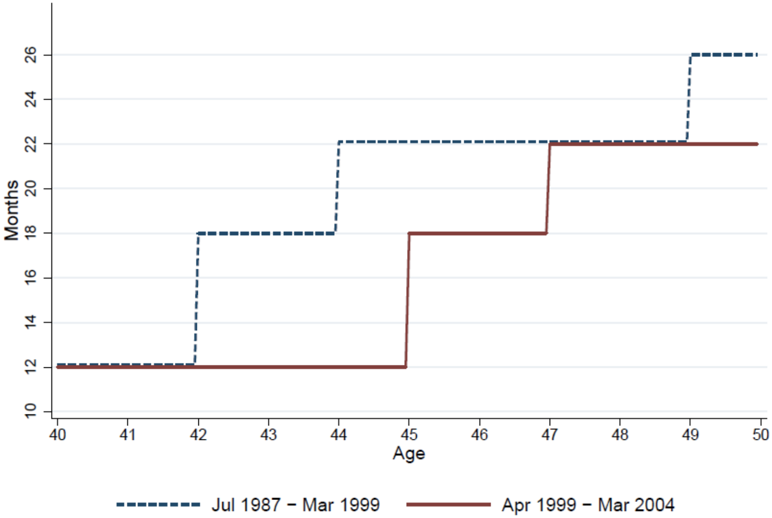
- Without perfect experience rating (over half of firms according to Feldstein 1978), firms can use UI system to pay workers lower wages for same work
  - Employ Worker Group A Jan-Mar, then lay them off
  - Employ Worker Group B Apr-June, then lay them off and hire back Worker Group A
- Feldstein (1976): models less brazen equilibrium with positive layoff rate after UI is introduced
- Topel (1983): imperfect experience rating explains 31% of temporary layoffs unemployment (see Krueger-Meyer 2002 for recent studies that find small magnitudes

UI departure #2: Cyclical considerations

- U.S. practice: extend UI in recessions (e.g. 26 weeks to 99 weeks)
    - Ljungqvist-Sargent (1998): bad if recessions involve structural reductions in wages, implying higher effective replacement rates
    - Schmieder-Wachter-Bender (2012), extension of Chetty (2008): good if provides greater liquidity because job finding rate declines
    - Landais-Michaillat-Saez (2014): even better if search externalities are negative (e.g. jobs are rationed)

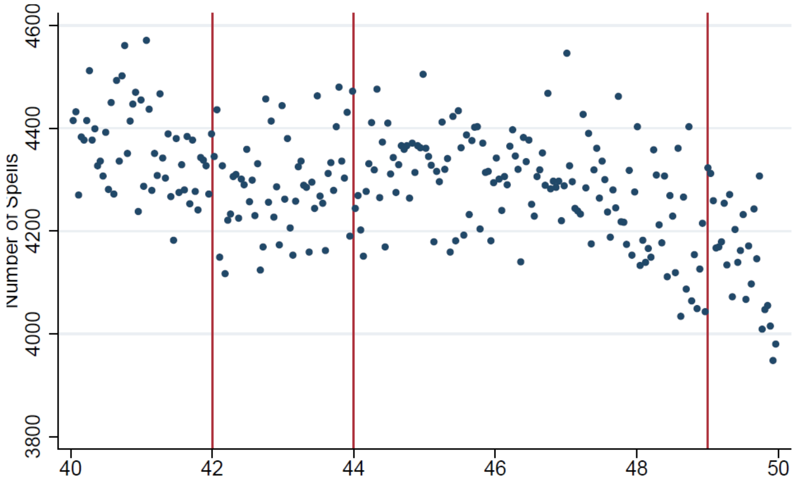
- Schmieder-Wachter-Bender use unique German context to justify U.S. practice via liquidity (see Crepon-Duflo-Gurgand-Rathelot-Zamora 2013 for evidence in favor of search externalities justification)

Age-specific potential benefit durations in Germany



Source: Schmieder-Wacther-Bender (2012)

No evidence of sorting across RD thresholds



(b) Age at Job Loss

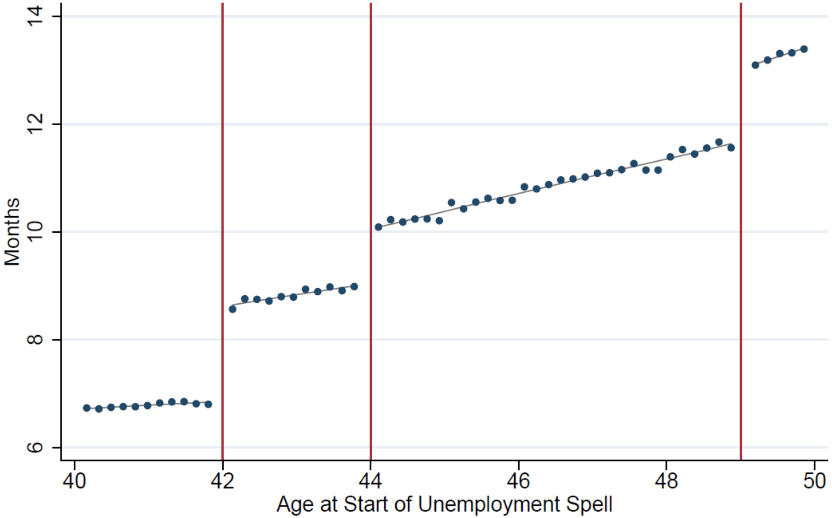
Source: Schmieder-Wachter-Bender (2012)

Schmieder-Wachter-Bender (2012)

- Welfare effect of extension in potential duration (e.g. being one day older than 42 relative to one day younger than 42):
  - Declining in effect on non-employment durations (moral hazard cost)
  - Increasing in effect on actual UI duration (since consumption is valuable non-employed state)

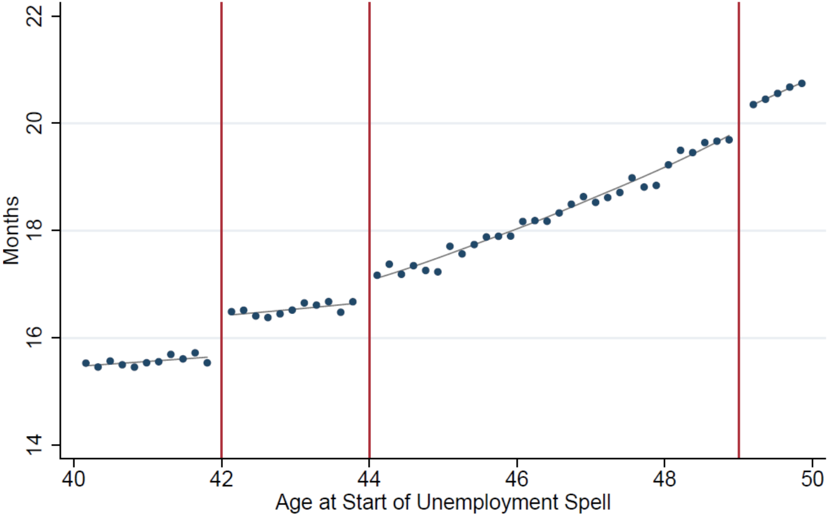
- Findings:
  - Zero or negative effect on non-employment durations
  - Large positive effect on actual UI durations

Large average effect on actual UI durations



Source: Schmieder-Wacther-Bender (2012)

Smaller average effect on non-employment durations

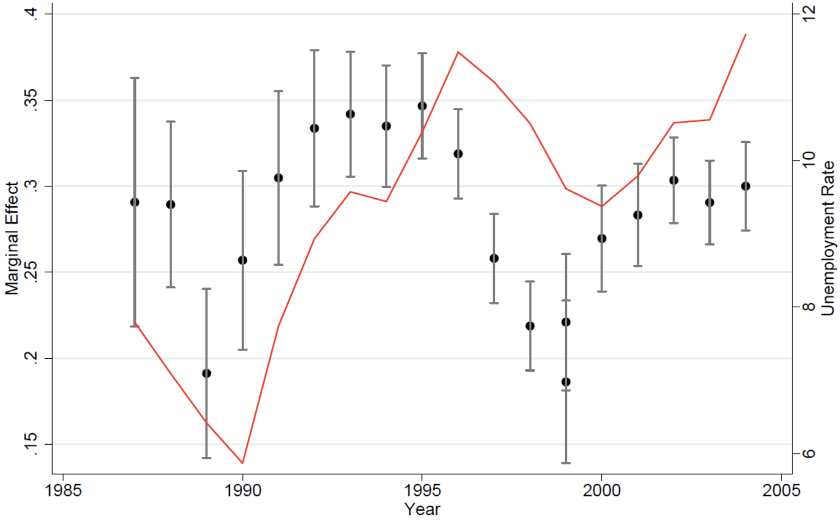


Source: Schmieder-Wachter-Bender (2012)

Non-employment duration effect correlates little with cycle

Source: Schmieder-Wacther-Bender (2012)

Actual UI duration effect correlates strongly with cycle



Source: Schmieder-Wachter-Bender (2012)

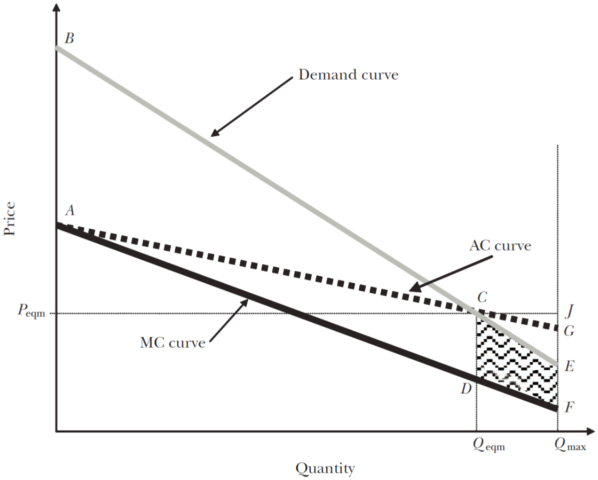
Outline: Adverse selection and health insurance

- Baseline theory: Akerlof (1970) and Rothschild-Stiglitz (1976)
- Empirical implementation: Einav-Finkelstein-Cullen (2010)

Akerlof (1970) baseline setup (Einav-Finkelstein 2011)

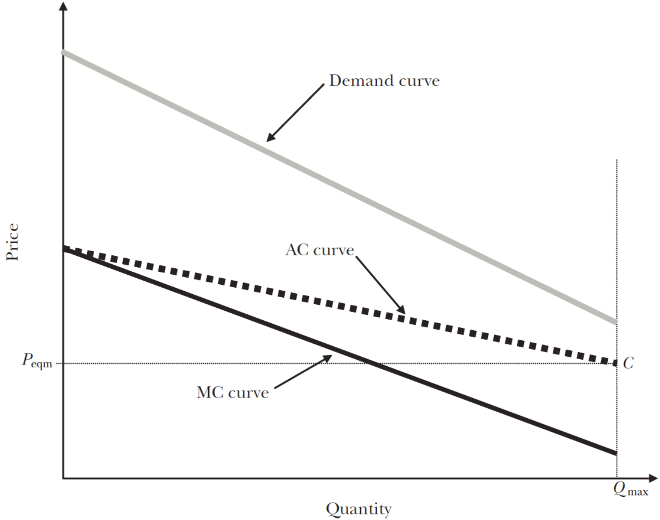
- Two states: no-loss state (endowment E_1) and loss state (endowment $E_2 < E_1$)
- Single binary insurance contract (unlike Rothschild-Stiglitz 1976) delivers $-\alpha_1$ in no-loss state and α_2 in loss state
- Individuals vary only in probability p_i of incurring loss
- Diminishing marginal utility (not assumed in used car application) $\rightarrow$ Value of contract rises with individual's private-information expected cost
- Zero overhead ("load"). Zero profits (firms break even).

Akerlof (1970): Some insured, some not (efficiency loss)



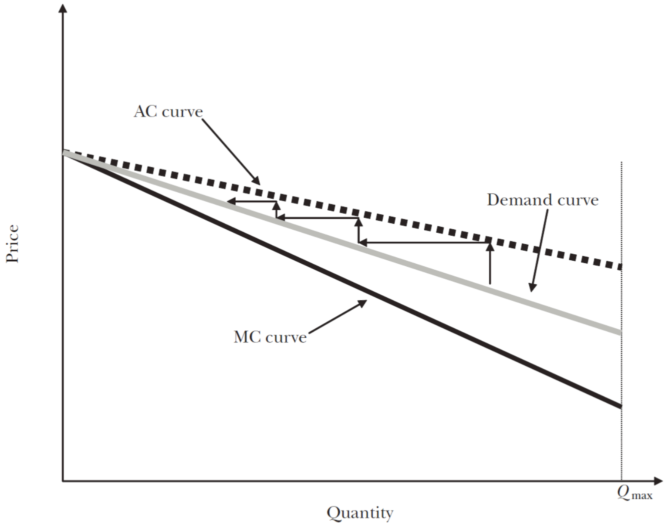
Source: Einav-Finkelstein (2011)

Akerlof (1970): Everyone insured (no efficiency loss)



Source: Einav-Finkelstein (2011)

Akerlof (1970): No one insured (large efficiency loss)

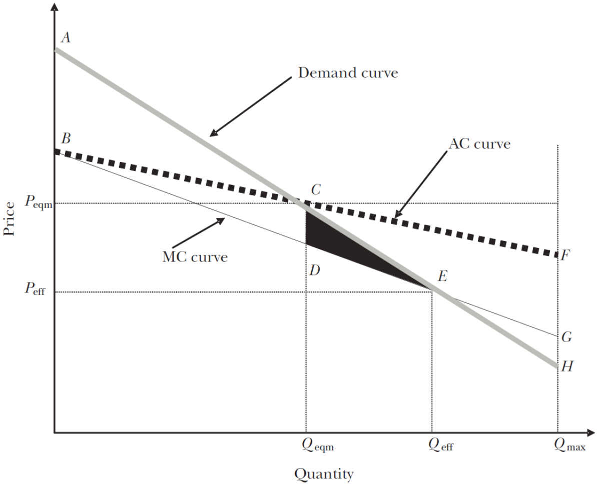


Source: Einav-Finkelstein (2011)

Comment #1: Full efficiency via mandate/subsidy

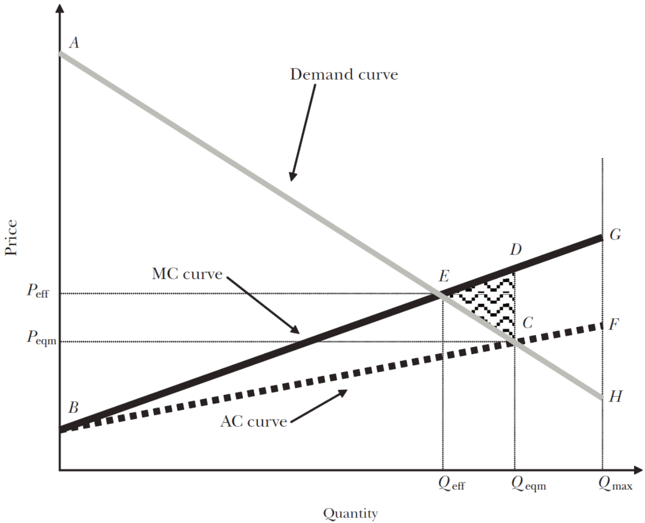
- Everyone values insurance (demand curve above marginal cost curve), so first-best involves full insurance
- Government mandate or large subsidy $\rightarrow$ eliminate efficiency loss
  - Mandate not Pareto-improving if average cost pricing
  - Subsidy Pareto-improving if high-risk types pay high enough share of subsidy

Comment #2: Full insur. may not be efficient with loads



Source: Einav-Finkelstein (2011)

Comment #3: Advantageous selection and overinsurance

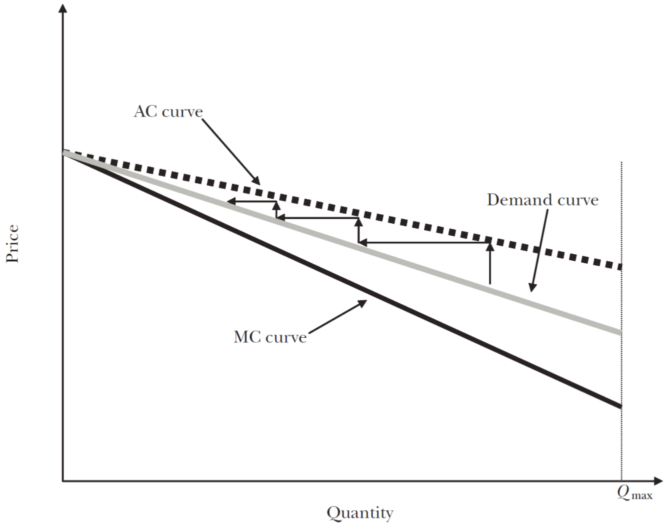


Source: Einav-Finkelstein (2011)

Comment #4: Strange pattern of who has insurance

- Policy concern: High-risk people (e.g. those with preexisting illnesses like strokes) do not have insurance
- Akerlof: Lowest-risk type is always insured!

Akerlof (1970): Full unraveling impossible for lowest type



Source: Einav-Finkelstein (2011)

Rothschild-Stiglitz model

- Economy with two types, low-risk (L) and high-risk (H)
- A fraction f of the individuals is 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

- Similarities to Akerlof
  - Static model: individuals arrive in the period either employed or unemployed; no savings/dynamics
  - No moral hazard: agents choose insurance contract but make no choices after signing a contract
  - Perfect competition: firms earn zero profits in equilibrium
- Key difference: full contract space (firms can enter and offer any zero-profit insurance contract)

Rothschild-Stiglitz: Contracts

- An insurance contract is described by a vector $\alpha = (\alpha_1, \alpha_2)$
  - Consumption in the two states: $(E_1 - \alpha_1, E_2 + \alpha_2)$
- Type i's expected utility is

$$V_i(\alpha) = (1 - p_i)u(E_1 - \alpha_1) + p_i u(E_2 + \alpha_2)$$

- Any contract that earns non-negative profits is feasible
  - Zero-profit condition $\Rightarrow$ firms price insurance s.t.

$$\alpha_2 = \frac{1-p}{p}\alpha_1$$

where p is risk rate of those who purchase contract.

Rothschild-Stiglitz: Equilibrium

Definition

An **equilibrium** is defined by a set of insurance contracts such that
(1) individuals optimize: both types cannot find a better contract than the ones they chose
(2) firms optimize: all firms earn zero profits

- Two types of equilibrium:

1. **Pooling**: both types are offered the same contract α.
2. **Separating**: high-risk types choose a contract α_H while low-risk types choose a different contract α_L.

Rothschild-Stiglitz: First Best Solution

- In first best, insurer can distinguish types (perfect information)
  - In this case, equilibrium is separating
- Plugging in $\alpha_2 = \frac{1-p_i}{p_i}\alpha_1$, each type solves

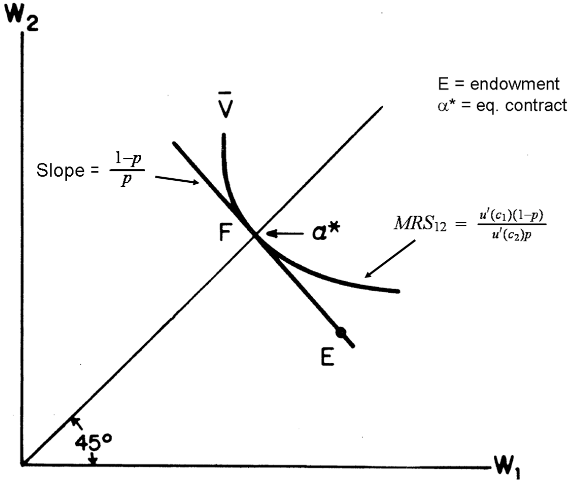
$$\max_{\alpha_1} (1-p_i)u(E_1-\alpha_1) + p_i u(E_2 + \frac{1-p_i}{p_i}\alpha_1).$$

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: consume their expected income $(1-p_i)E_1 + p_i E_2$ regardless of the state.

Equilibrium with Perfect Information



Source: Rothschild and Stiglitz 1976

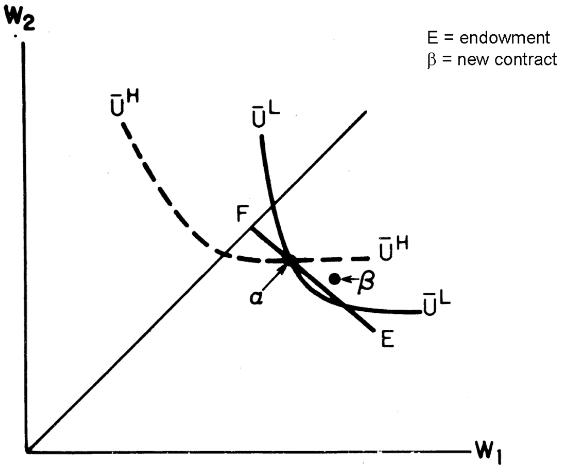
Rothschild-Stiglitz: Second-Best Problem

- 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 contracts and insurer goes out of business
- Hence optimal contracts differ when information is asymmetric

Rothschild-Stiglitz: Second Best Solution

- **Result #1:** no pooling equilibrium exists
- If H and L types are pooled in a contract α, low-risk types lose money in expectation.
- Zero-profit condition requires $\alpha_2 = \frac{1-\overline{p}}{\overline{p}}\alpha_1$ but $\overline{p} > p_L$.
  - Low-risk type gets fewer dollars in state 2 than he should if the insurance were fair for him.
- Creates an opportunity for a new insurer to enter and "pick off" low-risk types by offering slightly less insurance at a better price: higher c_1, lower c_2.
  - Only low-risk types switch, because they value c_1 more.

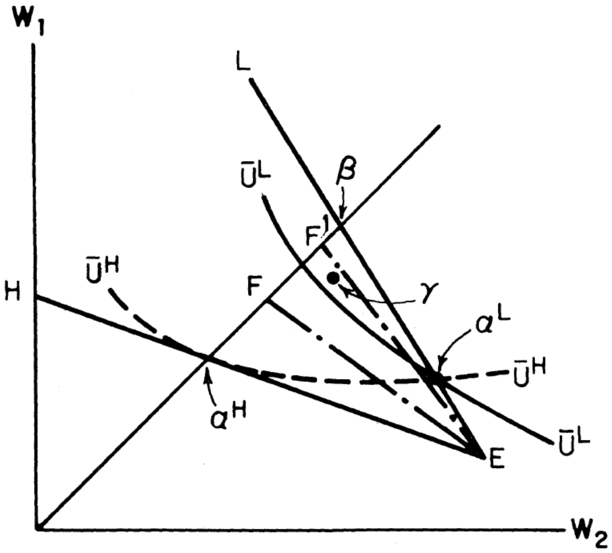
No Pooling Equilibrium with Asymmetric Information



Source: Rothschild and Stiglitz 1976

Rothschild-Stiglitz: Second Best Solution

- **Result #2:** in a separating eq, Type H obtains full insurance and Type L is under-insured
- **Intuition:** in any sep. eq., both types are getting actuarially fair insurance because of the zero-profits condition
  - For H, no cost to firm in providing full ins (worst that can happen is that L will join the pool, raising profits)
  - But for L, full ins. would create an incentive for H to buy this (cheaper) policy, forcing firm into negative profits
- Incentive constraints always bind downward—"no distortion at the top" result in standard asymmetric info. models
- In eq., L gets as much insurance as possible without inducing H to deviate and pretend to be low-risk



Source: Rothschild and Stiglitz (1976)

Comment: Still strange pattern of who has insurance

- Both Akerlof and Rothschild-Stiglitz: highest-risk type (e.g. person with history of strokes) has full insurance

- Real world: Highest-risks sometimes cannot buy insurance at any price (fully unraveled market for them)

- Hendren (2013): Akerlof market fully unravels for a market segment when types are continuous if

$$\frac{p}{1-p}\frac{u'\left(E_2+\alpha_2\right)}{u'\left(E_1-\alpha_1\right)} \leq \frac{E\left[P|P \geq p\right]}{1-E\left[P|P \geq p\right]}\ \forall p$$

  - For every type p: "people riskier than me are sufficiently riskier relative to my value of insurance"
  - Evidence: rejected insurance applicants have more private information than accepted insurance applicants, and markup larger than insurance value based on other work

Comment: Does adverse selection explain ObamaCare?

- Key parts of Affordable Care Act ("ObamaCare"): insurance mandate (actually a tax on no-insurance, similar to a subsidy) and "community rating": premiums determined by average costs of locally insured and cannot condition on preexisting conditions
- Community rating in Akerlof: can either increase or decrease efficiency
- Possible rationale: redistribution, not social insurance
  - Likely want to redistribute from low-types to high-types since high-types are poorer (also because they have to pay high premiums)
  - Justified as social insurance only from behind veil of ignorance (same as optimal tax)

Application: Why do we have Social Security

1. Adverse selection in annuities markets: Finkelstein-Poterba (2004)

2. Optimal disability insurance with screening costs (Diamond-Sheshinski 1995)

3. Myopia (“internalities”) (Beshears-Choi-Clayton-Harris-Laibson-Madrian 2015)

4. Moral hazard with lack of political commitment

Adverse selection as rationale for Social Security

- Lifecycle solution to risk of outliving one's savings: annuitize assets at retirement into lifetime income stream
- If mortality risk is private information, annuity market can unravel (Eckstein-Eichenbaum-Peled 1985)
- U.S. annuity market is tiny, but Americans may already be over-annuitized by Social Security (want to hang on to assets to cover other risks)

Finkelstein and Poterba 2004

- Study two types of annuity markets: compulsory vs. voluntary.
- Examine two features of annuity contracts
  - degree of **backloading** (inflation indexing and escalation of payments over time)
  - **payments to estate** in event of death (guarantees and capital protection).
- Positive correlation (Chiappori-Salanie 2000) predictions
  1. In eq., those who purchase backloaded annuities have lower mortality rates
  2. In eq., those who purchase annuities with payment to estate have higher mortality rates
  - Both effects should be stronger in voluntary markets

TABLE 2
SELECTION EFFECTS AND ANNUITY PRODUCT CHARACTERISTICS

Explanatory Variable	Estimates from Hazard Model of Mortality after Purchasing an Annuity: Compulsory Market (1)	Estimates from Hazard Model of Mortality after Purchasing an Annuity: Voluntary Market (2)	Estimates from Linear Probability Model of Probability of Dying within Five Years: Compulsory Market (3)	Estimates from Linear Probability Model of Probability of Dying within Five Years: Voluntary Market (4)
Index-linked	−.839***	−.894**	−.053***	−.185***
	(.217)	(.358)	(.019)	(.050)
Escalating	−1.085***	−1.497***	−.072***	−.152***
	(.113)	(.253)	(.010)	(.030)
Guaranteed	.019	.216***	.007*	.046***
	(.029)	(.060)	(.004)	(.016)
Capital-protected	...	.056	...	.064***
		(.051)		(.016)
Payment (£100s)	−.003***	.001**	−.0003***	.0003***
	(.0006)	(.0004)	(.0001)	(.0001)
Male Annuitant	.640***	.252***	.044***	.044***
	(.039)	(.051)	(.005)	(.014)
Observations	38,362	3,692	24,481	3,575
Number of deaths in sample	6,311	1,944	2,693	822

Source: Finkelstein and Poterba 2004

Limitations of Positive Correlation Test

1. Does not account for other dimensions of heterogeneity that may confound the correlation
   - Literature on "advantageous selection" (e.g., Finkelstein and McGarry 2006)
2. 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
3. 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

DI screening costs as rationale for Social Security

- Diamond-Sheshinski (1995): Individuals have different disutilities of working ψ_i
- To max social welfare, not desirable for those with high ψ_i to work.
- First best: Individual i works iff

$$\text{Marginal product} > \psi_i$$

- But govt observes only an imperfect signal of $\psi_i \rightarrow$ sets a higher threshold for disability
- Result: lower benefit rate if screening mechanism has higher noise-to-signal ratio

DI screening costs as rationale for Social Security

- Suppose there exists a marginal cost (e.g. paying the DI judge) to screening each DI applicant
- And suppose actual disability rates rise with age toward 100%
- At some age and before actual disability rate reaches 10%, it will be optimal to give everyone DI without screening them (call it "Social Security")
    - Though need reason for providing a transfer to people if disability rate actually reaches 100% (no insurance value)

Other rationales for Social Security

- Myopia (Beshears-Choi-Clayton-Harris-Laibson-Madrian 2015): if small share of population fails to save for retirement but policy cannot be targeted to them, policy caters to them because welfare benefits are huge for myopic
  - Feldstein (1985) argues that myopia is typically not a sufficient rationale but requires strong assumptions on capital retardation from pay-go (current Social Security payments fund current retirees' rather than business investment)

- Moral hazard with lack of political commitment: Youth foresee that politicians/public will never allow elderly to live in poverty, so youth do not save

Areas for future work

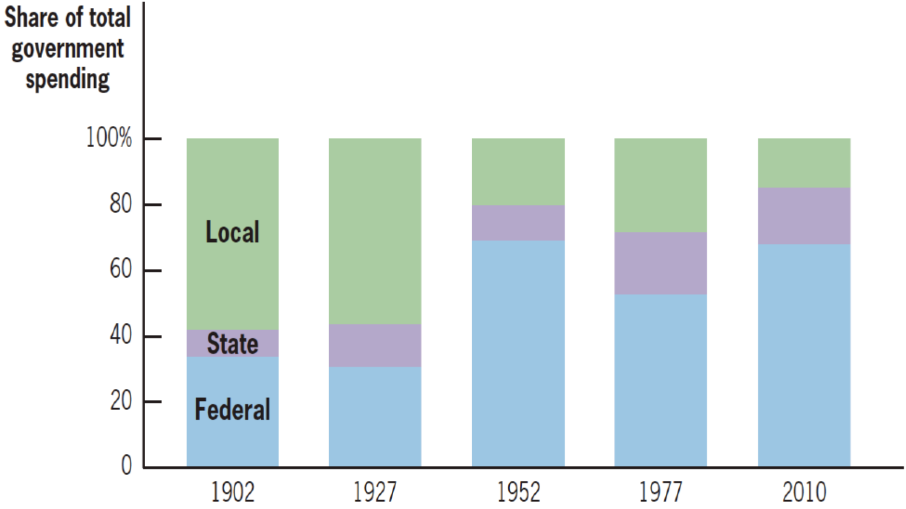
- Connecting optimal DI to empirical DI
- Revealed preference estimates for all types of insurance
- Optimal public-vs-private provision (e.g. Cabral-Geruso-Mahoney 2015)
  - Social Security and Medicare: gov. just writes checks
  - UK National Health Service: gov. provides the care too

Graduate Public Economics

Lecture 5
Local Public Finance

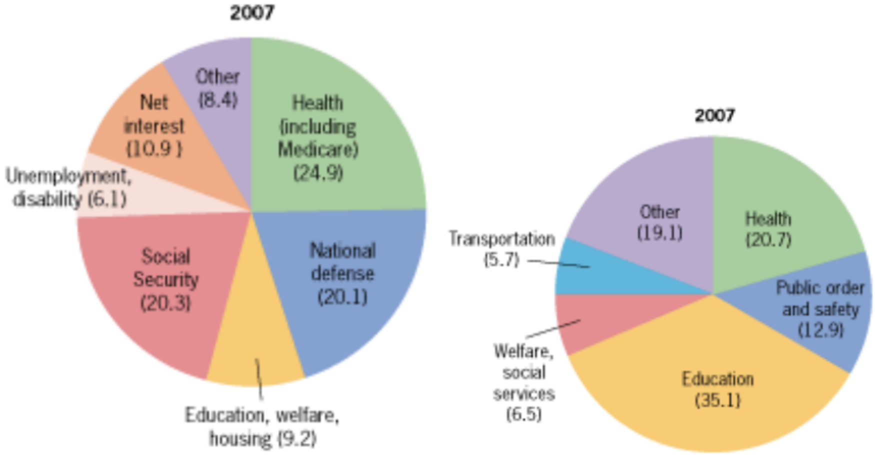
Danny Yagan
UC Berkeley

Motivation: One-third of U.S. gov. spending is subnational

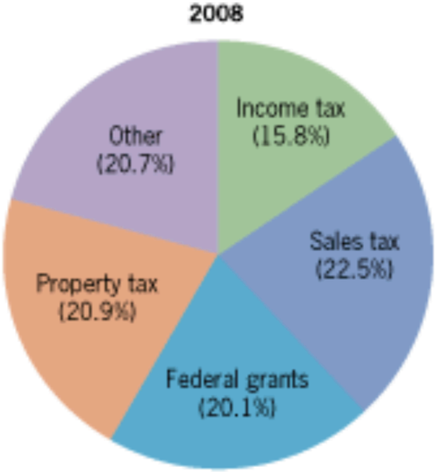


Source: Gruber (2013)

Motivation: Subnational govs do different things



Source: Gruber (2013)



Source: Gruber (2013)

Motivation: Fertile ground for research

- Excellent policy variation
- Excellent data
- Economic richness b/c of general equilibrium
- Diverse constituency (PE, labor, urban, trade, development)

Local public finance: Outline

- Tiebout (1956): Efficient provision of local public goods
- Empirical tests of Tiebout
- Fiscal federalism and education finance

Intellectual progression

- Arrow-Debreu (1954): Existence and Pareto efficiency of competitive equlibrium under no externalities
- Samuelson (1954): Massive underprovision of publicly provided goods in decentralized market
- Tiebout (1956): Pareto-efficient provision of publicly provided goods in decentralized market

Tiebout (1956): Efficient local provision

- References since Tiebout is just a sketch: Bewley (1981) and Glaeser (2008)
- Consumers:
  - Each consumer i is one of finite types $\alpha_i \in A$, measure one in total
  - Possesses endowment z, must be living in a city $c \in C$, must pay city's "head" tax t_c
  - Costlessly chooses city to maximize quasi-linear utility in private consumption and publicly provided good g_c:

$$\max_{c \in C} z - t_c + \alpha_i v\left(\frac{g_c}{n_c}\right)$$

  - Idea: city produces g_c worth of teachers which are split across n_c residents

Tiebout (1956): Efficient local provision

- City governments:
  - At least as many city governments as types of people: $|C| \geq |A|$
  - Produces good g_c with some share $\mu_c \in [0, 1]$ of tax revenue. Leaders pocket rest (profit).
  - Technology: $g_c = \mu_c t_c n_c$ (e.g. one dollar allocated to paying teachers generates one dollar's worth of teaching)
  - Chooses (t_c, g_c) s.t. $g_c \leq t_c n_c$ to maximize profit (zero in equilibrium):

$$\max_{t_c, g_c} t_c n_c - g_c$$

- Competitive equilibrium: Allocation of consumers to cities and vector of city taxes and public good levels such that consumers and governments are optimizing

Tiebout (1956): Efficient local provision

- Result: Competitive equilibria exist and are Pareto optimal
  - Mathematically identical to Arrow-Debreu
- Mechanism: Consumers "vote with their feet" and sort into homogenous communities
  - At least one city specializes in attracting highest-α type (high taxes with great schools) while others do opposite
  - Consumers of each type move to city satisfying MRS $=$ MRT: $\alpha v' \left(g_c / n_c \right) = 1$
  - Allocation is same as if consumers bought publicly provided good on private market!

What just happened here?

- The publicly provided good is a "public service," not a canonical "public good" (Bewley 1981)
- Public good (e.g. national defense, public radio): non-excludable (everyone enjoys it) and non-rival (my enjoyment does not affect your enjoyment)
- Public service: somewhat excludable (only city residents can enjoy city's teachers) and rival (costs *and* benefits of teachers are evenly divided among residents)
- Tiebout: $|C| \geq |A|$ provides sufficient excludability for self-interested providers (profit-maximizing cities) to provide efficient amount

Tiebout (1956): Inefficient local provision of non-rival good

- Suppose the publicly provided good is non-rival, so consumers maximize:

$$\max_{c \in C} z - t_c + \alpha_i v\left(g_c\right)$$

- First-best: everyone lives in same city (exploiting economies of scale) and each type pays an α-specific tax

- Free-rider and coordination problem without differentiated taxes
  - A single large city can be second-best
  - But for sufficiently wide variation in α, another "entrepreneurial" (Bewley 1981) city can pick off the lowest-α type (a la Rothschild-Stiglitz 1976)

- Lesson: Tiebout requires no scale economies, or else back to Samuelson (1954) problem of externalities are not being internalized

Tiebout (1956): Inefficient local provision of non-rival good

- Every model is wrong. The question is "how wrong?" Is decentralization better than centralization, even if not first-best?
- Scale economies: Huge in national defense, perhaps little in schooling
    - Can argue Constitution delegates powers optimally: national defense to central gov, schooling to local gov (de Tocqueville 1835)
    - Road building? Road cleaning? Law enforcement?

Tiebout (1956): Head tax vs. property tax

- Tiebout: Every resident pays same dollar amount in taxes ("per head")
- Margaret Thatcher (UK Prime Minister 1979-1990) tried a head tax ("community charge"/"poll tax") . . .

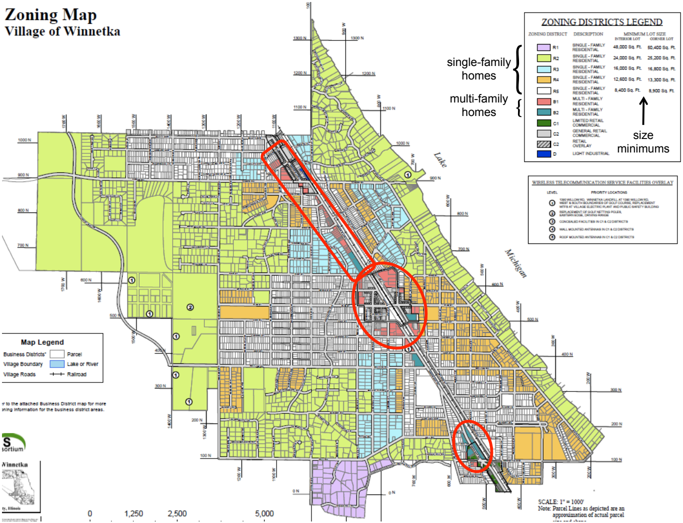
1990 Poll Tax Riots: 200,000 people in Central London

1990 Poll Tax Riots: 200,000 people in Central London

Tiebout (1956): Head tax vs. property tax

- In reality, cities finance themselves by a property tax (and intergovernmental transfers)
- Property tax for resident $i = \tau_c^{PROP} * ASSESSEDHOMEVALUE_i$
- Creates incentive for poor to move into rich cities (enjoy great schools at low cost)
- Hamilton (1975): Can restore efficiency if allow zoning
    - Ex. no apartments, minimum plot size of houses
    - Efficiency now requires homogeneity in housing preferences too

Zoning ex.: Winnetka IL (per-pupil spending 2x state avg.)



Source: http://www.villageofwinnetka.org/ (multi-family homes allowed only in red and teal areas)

Is Tiebout a good model?

- Capitalization: Property tax changes are benefits taxes (fully valued) and have no impact on house price
- Knowledge: People know what their property taxes buy
- Sorting: People sort along observable dimensions

Capitalization as a way to measure incidence

- With efficient markets, asset price immediately reflects effect of an event on PDV of asset's lifetime return flow (Summers 1985; Cutler 1988)
    - Works for both immediate and future tax changes announced today
    - Implies incidence is fully borne by today's owners (or else could arbitrage)
- Difficulty: Expectations matter
    - If tax changes today but was fully expected to, no asset price change
    - If tax changes unexpectedly today but is expected to revert soon, small asset price change
- Difficulty: With elastic supply, quantities can matter too

Rosen (1982): CA Prop 13 effect on house prices

- In 1978, CA became first of nearly forty states to limit property taxes
  - Rate: no more than 1% (except to cover pre-existing bond payments)
  - Base: market price at purchase plus no more than 2% annual appreciation
- Some jursidictions had higher pre-existing tax rates $\rightarrow$ more affected than others
- Regresses 1976-1979 house price change on 1976-1979 annual property tax savings

Rosen (1982): CA Prop 13 effect on house prices

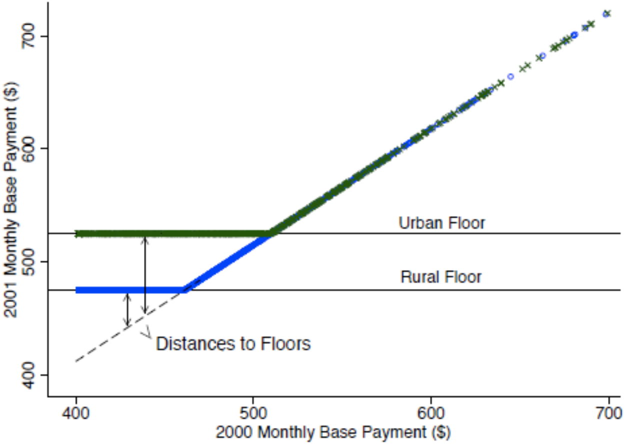
- Result: Coefficient of 7 (\$1 of annual property tax reduction $\rightarrow$ \$7 increase in house price)
- At interest rate of 12-15%, which implies full capitalization $(1/.135 \approx 7)$
- Tiebout: Should have no capitalization
- Explanation: "these tax cuts occurred without any substantial corresponding reduction in services as the state of California's surplus was used to bail out local communities"?

Sidebar: Always make a graph

- Yagan (and others') Doctrine: If you can't show your quasi-experimental result in a graph, you don't have a quasi-experimental result.
- Rosen shows no graph, so reader can't evaluate "common trends":
  - DD identifying assumption: in absense of Prop 13, property values would have trended similarly across highly affected and lightly affected jurisdictions
- Ex. of similar design: Cabral-Geruso-Mahoney (2015)

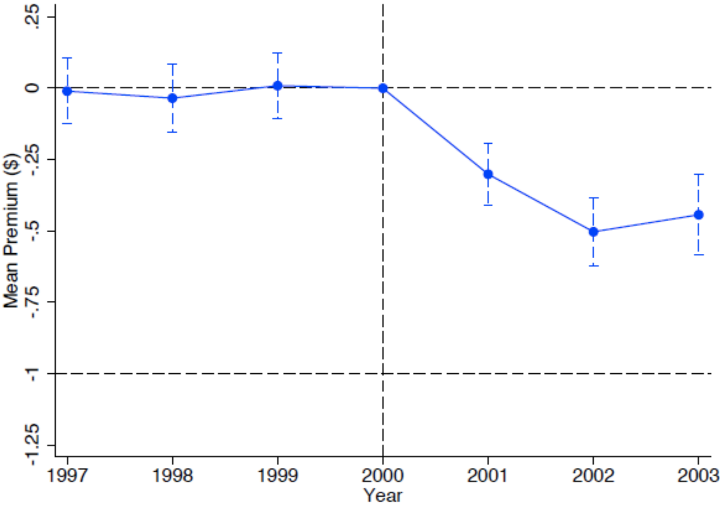
Treatment heterogeneity across space, just like Rosen

Figure 1: Payment Floors: Pre- and Post-BIPA Monthly Base Payments



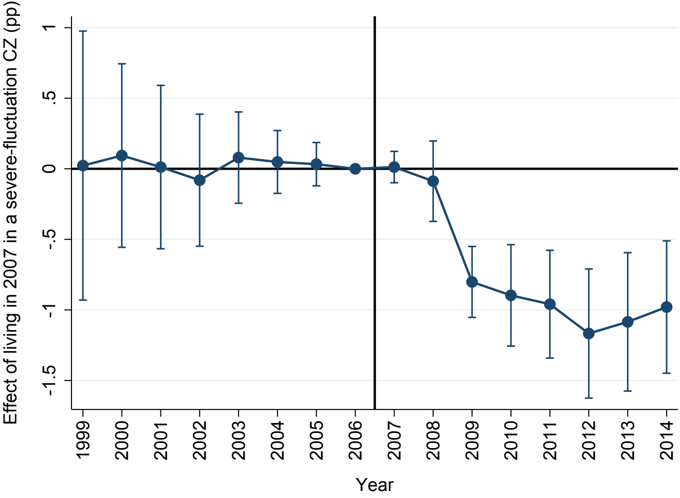
Source: Cabral-Geruso-Mahoney (2015)

Eval. pre-trends with time series of within-year estimates



Source: Cabral-Geruso-Mahoney (2015)

Less parametric: Conditional means across quantiles

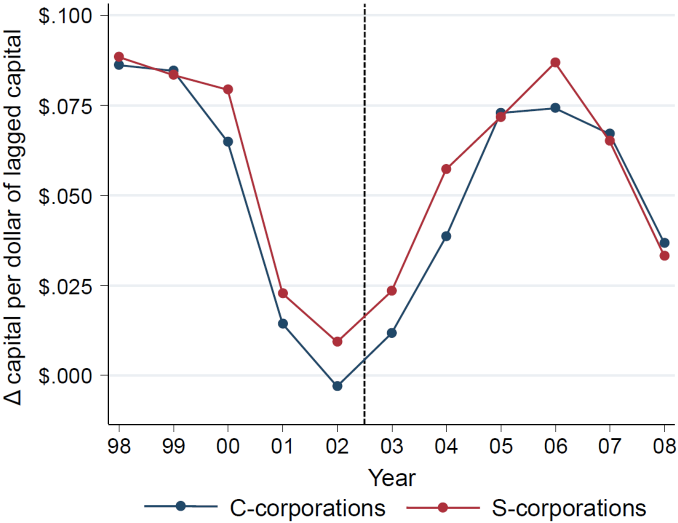


Source: Yagan (2016)

What about when doing a pooled DD?

- Cabral-Geruso-Mahoney and Yagan (2016) do DDs with only one pre-treatment year and one post-treatment year for main estimates
- Often lack power $\rightarrow$ need to pool pre-treatment years and post-treatment years in main DD specification
- Should still show year-by-year version $\rightarrow$ separate identification demonstration (common trends) from inference (main specification)

Year-by-year estimates in pooled DD setting



Source: Yagan (2015)

Cellini-Ferreira-Rothstein (2010): Marginal spending

- Locales issue bonds to build new schools, paid for by future taxes
- In CA, residents must approve proposed bond issues
- CFR: Value marginal bond issue by comparing close house prices in locales that barely approved bond issue to those in locales that barely failed

CFR: Sample bond issue

Shall Alhambra Unified School District repair, upgrade and equip all local schools, improve student safety conditions, upgrade electrical wiring for technology, install fire safety, energy efficient heating/cooling systems, emergency lighting, fire doors, replace outdated plumbing/sewer systems, repair leaky rundown roofs/bathrooms, decaying walls, drainage systems, repair, construct, acquire, equip classrooms, libraries, science labs, sites and facilities, by issuing $85,000,000 of bonds at legal rates, requiring annual audits, citizen oversight, and no money for administrators' salaries? (Institute for Social Research 2006)

Source: Cellini-Ferreira-Rothstein (2010)

Cellini-Ferreira-Rothstein (2010): Dynamic RD

- Locales can keep trying $\rightarrow$ "dynamic RD"
  - Passed and failed bonds equally likely to pass ex ante
  - But ex post, failed bond issue may soon succeed (expectations matter!)

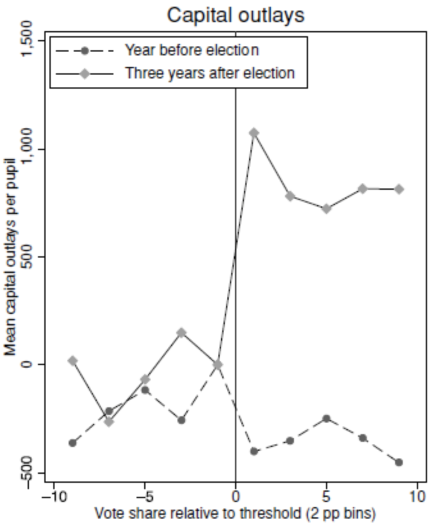
- Easiest to estimate intent-to-treat (ITT): effect of initial treatment status (pass/fail), regardless of compliance (whether failures later succeed)

Cellini-Ferreira-Rothstein (2010): Dynamic RD

- Key input for Tiebout is treatment-on-the-treated (TOT): effect of initial treatment status under full compliance (failures stay failed)
  - Most useful Tiebout parameter: TOT (equals willingness-to-pay for marginal dollar of spending)

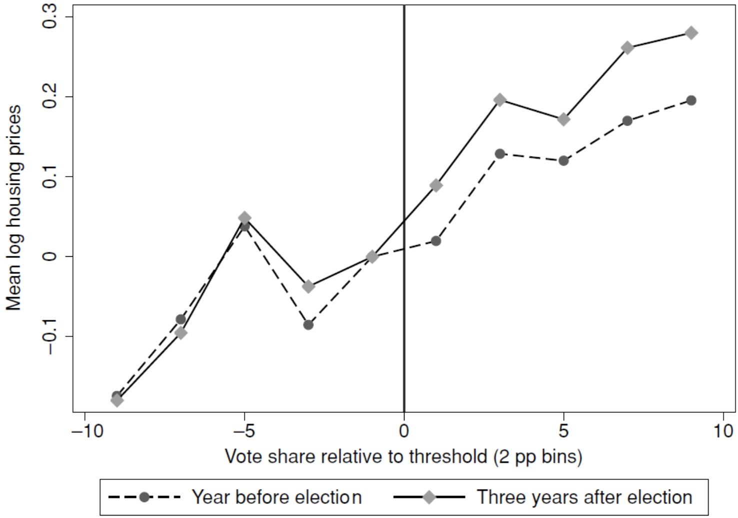
- CFR implement "dynamic RD" (similar to fuzzy RD) that recursively nets out effect of bond failure on subsequent approval, assuming time-invariant effects
  - TOT = ITT − [cumulative effect of bond passage on fewer subsequent approvals]

CFR: First-stage



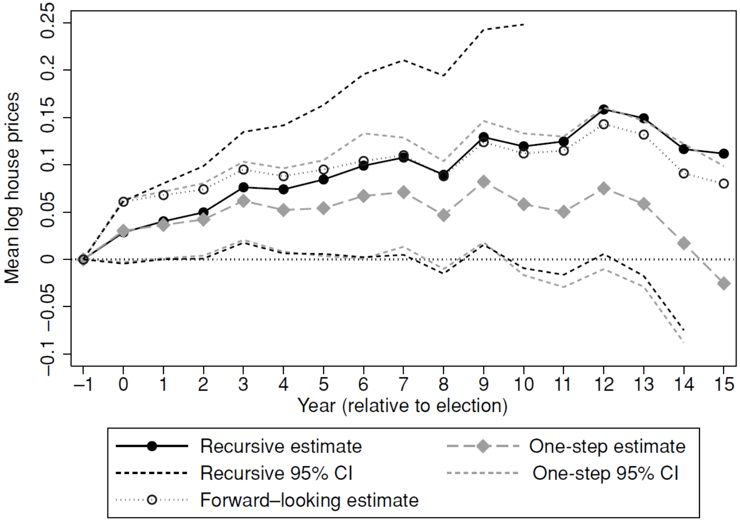
Source: Cellini-Ferreira-Rothstein (2010)

CFR: ITT



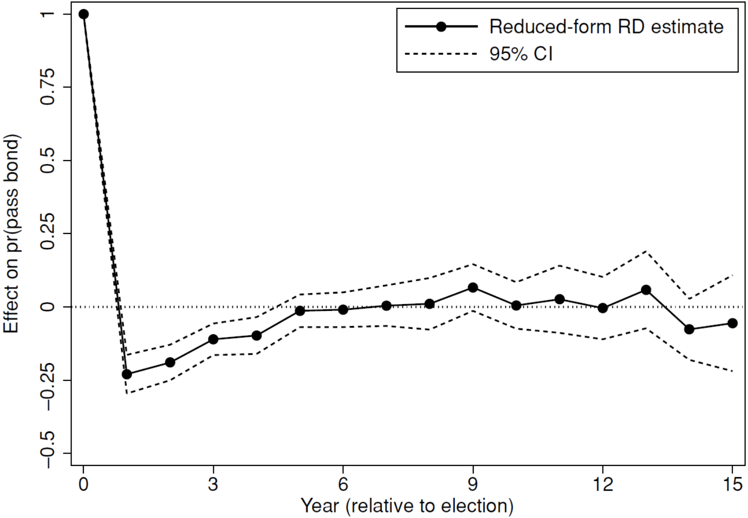
Source: Cellini-Ferreira-Rothstein (2010)

CFR: TOT



Source: Cellini-Ferreira-Rothstein (2010)

CFR: TOT only somewhat larger than ITT



Source: Cellini-Ferreira-Rothstein (2010)

CFR vs. Rosen

- Rosen: Property taxes $\downarrow \longrightarrow$ House prices $\uparrow\uparrow \implies$ Had been way over-spending
  - Looks like Tiebout failure: \$1 in property taxes value near \$0 (relative to \$1 Tiebout benchmark)
- CFR: Property taxes $\uparrow \longrightarrow$ House prices $\uparrow \implies$ Had been under-spending a bit
  - Tiebout looks better: \$1 in property taxes valued at \$1.14-\$1.44

Average vs. marginal valuations

- Rosen and CFR: Capitalization of marginal dollar
- Black (1999): Capitalization of average dollar
  - If have same expenditure but different quality (e.g. school technology variation) $\longrightarrow$ different house prices
  - Calculate house price difference within same school districts (holding property tax rate constant) but across school attendance boundaries
  - Finding: 5% (1 s.d.) higher test scores $\rightarrow$ 2.5% higher house price

Black (1999): School attendance border design

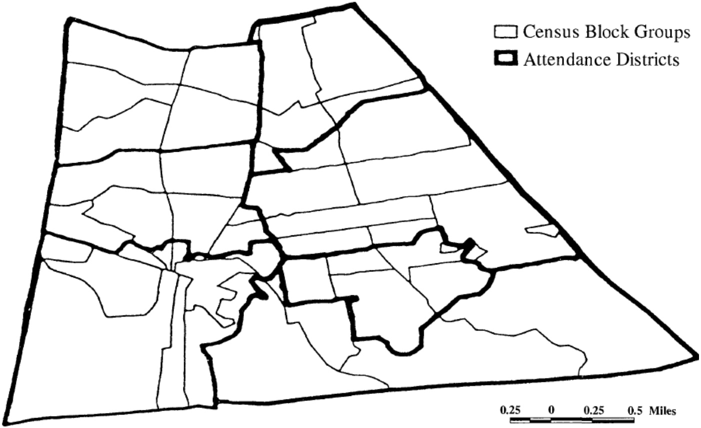


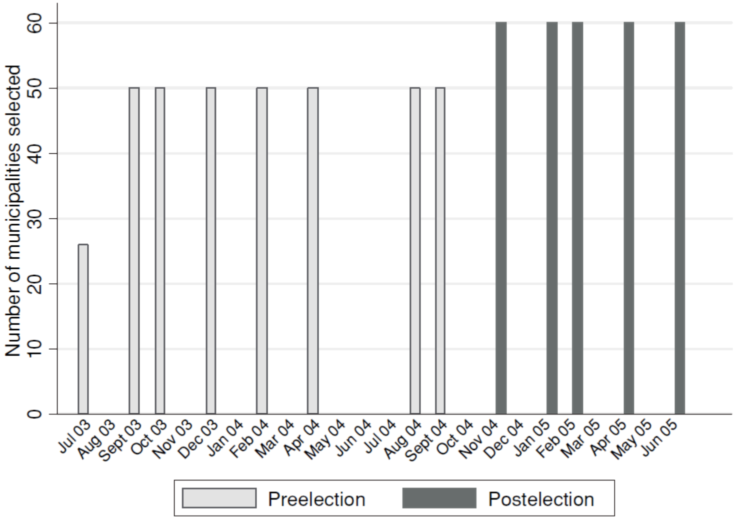
FIGURE II
Example of Data Collection for One City: Melrose
Census Block Groups and Attendance District Boundaries

Source: Black (1999)

Ferraz-Finan (2008): Knowledge frictions

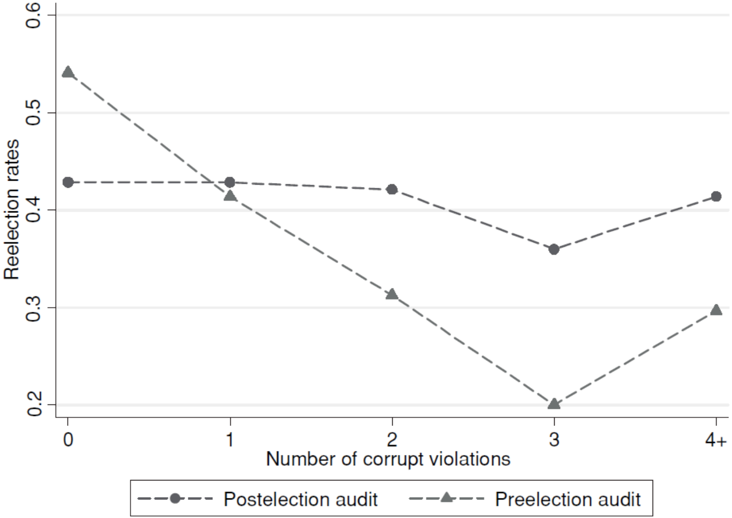
- Tiebout: Every property tax dollar efficiently goes to public service
- Major potential friction: Consumers may not know how well their property tax dollars are being spent (á la No Child Left Behind)
- Ferraz-Finan (2008): Effect of Brazilian local audits on incumbent mayors

FF: Audit release timing (treatment heterogeneity)



Source: Ferraz-Finan (2008)

Ferraz-Finan: Effect on mayors' reelection rates

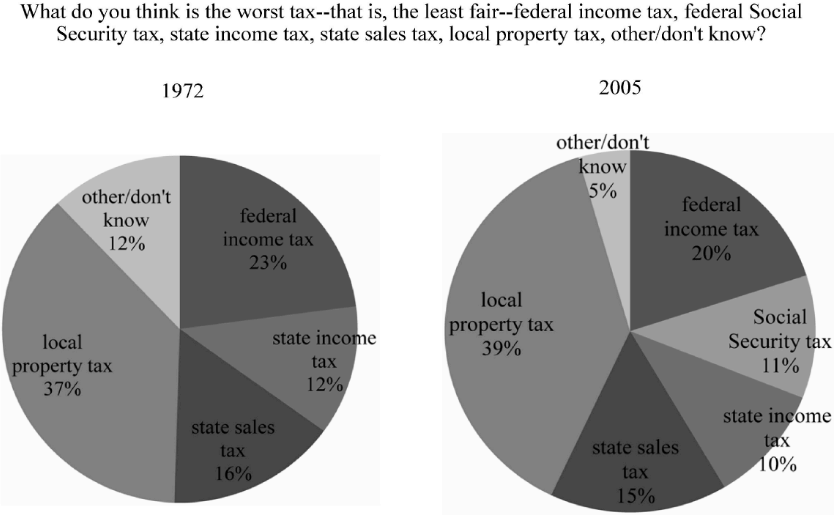


Source: Ferraz-Finan (2008)

Knowledge: Do people treat it like a benefits tax?

- Tiebout: Property tax is a "benefits tax": it pays for benefits that accrue fully to the payer, valued at cost
- Almost no other taxes are a theoretically pure benefits tax
- Should see that people think the property tax is great and other taxes are pernicious

People hate the property tax



Source: Cabral and Hoxby (2013), based on polls by Advisory Commission on Intergovernmental Relations and Gallup.

Rhode-Strumpf (2003): 140 years of Tiebout sorting?

- Tiebout: Frictionless mobility
- As mobility costs decline, should see greater heterogeneity in local taxes and other measures of residential sorting (segregation)
- Rhode-Strumpf: Exactly the opposite 1850-1990
  - 2/3 reduction in variation in per-capita local taxes
  - Similar reductions in income/racial/political segregation

Rhode-Strumpf (2003): Possible explanations

- Dispersion in employment opportunities (e.g. professor jobs everywhere, not just Cambridge, MA)
- Idiosyncratic tastes for places (e.g. surfers can finally all move to SF) that effectively reduces $|C| - |A|$
- Decline in local funding/provision relative to higher-level funding/provision of publicly provided goods (true)

Fiscal federalism interactions

	Source of Funds in 2004		
Unit	Central	State	Localities
State government	24%	75%	1%
Local government	4%	30%	66%

Source: Census of Governments: State & Local Finances (2004)

Why transfer from central to local?

- Preference heterogeneity
- On-the-ground implementation/optimization
- Competition across local areas
- Redistribution
- Paternalism (Farhi-Werning 2007)
- Internalize fiscal externalities (Gordon 1983; Oates 1999; Bovenberg Jacobs 2005)

Bovenberg-Jacobs (2005): Internalizing fiscal externality

- Subsidizing education (e.g. tuition at UC Berkeley) seems regressive: students will soon be high-wage people
- But future income taxes can reduce human capital accumulation
- Government has equity stake in you $\longrightarrow$ education subsidy $=$ labor income tax rate under specific conditions

Hoxby (2001): School finance equalization

- Every state has deviated from Tiebout: local school spending deviates from local property tax revenue
  - Redistribution
  - Paternalistism: value children more than parents do

- "Categorical aid": Locales get flat per-capita amount based on income
  - Tax price (amount of revenue locale must raise in order to spend an extra dollar) = \$1

Hoxby (2001): School finance equalization

- "Foundation aid": assesses statewide property tax and rebates lump sum, but locales can top up with their own property tax
  - Tax price = \$1
- "Guaranteed Tax Revenue" (e.g. CA/TX): In extreme, locales cannot top up with own property tax
  - Tax price $= (1, \infty)$

Hoxby (2001): Estimation

- Goal: estimate effect of FA tax rate, FA rebate, and (inverse of) tax price on school district spending using 1972, 1982, 1992 Census of Governments

- RHS vars reflect direct tax effects and endogenous reponses

- Uses simulated instruments to exclude variation from endogenous responses
  - Predict changes in RHS variables using 1970 characteristics and subsequent tax changes (Gruber-Saez 2002; Weber 2014)

Baicker-Staiger (2005): Matching categorial grants

- Categorical grant: Fixed amount
- Matching categorical grant: Subsidize local expenditures
- Medicaid Disproportionate Share Hospital Program: Federal gov. in 1989 started giving states 50%-82% match rate on poor hospital expenditures
- Enforcement: Many states wrote checks to hospitals, which then wrote checks right back to the state

Baicker-Staiger (2005): Findings

- Federal government audit: States captured 19% of DSH transfers through circular payments, with large heterogeneity across states
- Findings: Diverting states experience no decline in mortality while non-diverting states experience large declines
- Non-diverting states: Possibly flypaper effect (Hines-Thaler 1995)

Albouy (2009): Federal taxes and local cost of living

- Cities differ in attractiveness: productivity (higher wages) and amenities
- These differences can be baked into house/rental prices
    - High nominal incomes in high-productivity/low-amentity places, but equal real incomes across space because housing prices adjust to make people indifferent across space (Rosen 1979; Roback 1982)

Albouy (2009): Federal taxes and local cost of living

- Federal taxes are assessed on nominal incomes
  - Subsidizes places with low nominal incomes (low-productivity/high-amenity) relative to others
- Albouy: Workers in cities with above-average nominal wages pay 27% more in federal income taxes than workers in other cities
- Some tax deductions help (mortgage interest deductions and state/local taxes)
- Large or small welfare loss?

Graduate Public Economics

Lecture 6

Place-Based Policies

Danny Yagan
UC Berkeley

Motivation: Geographically concentrated output

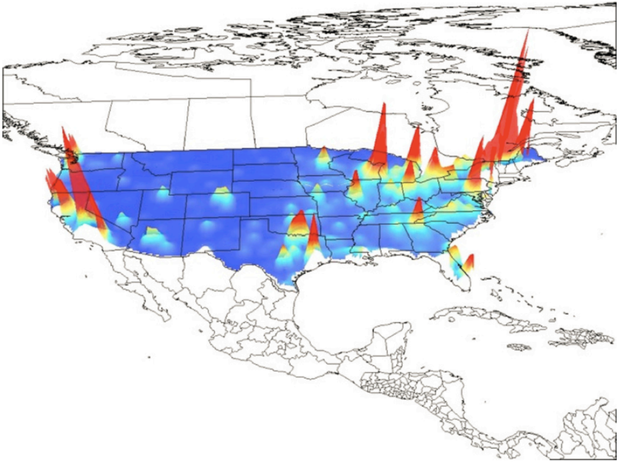
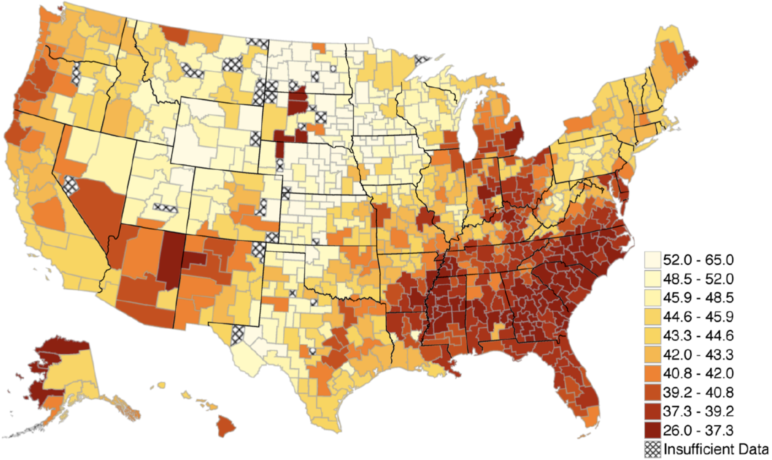


Figure 1 ***Spatial distribution of economic output in the US, by square mile.*** Notes: This figure reports the value of output produced in the US by square mile.

Source: Moretti (2011)

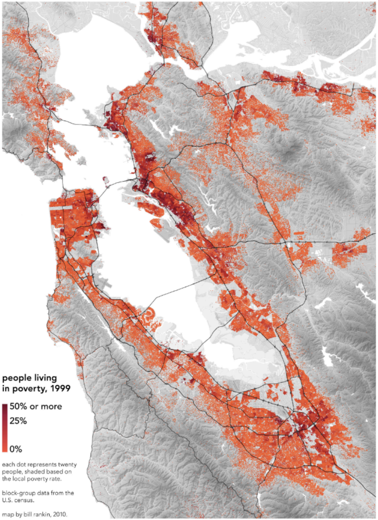
Motivation: Geographically concentrated upward mobility

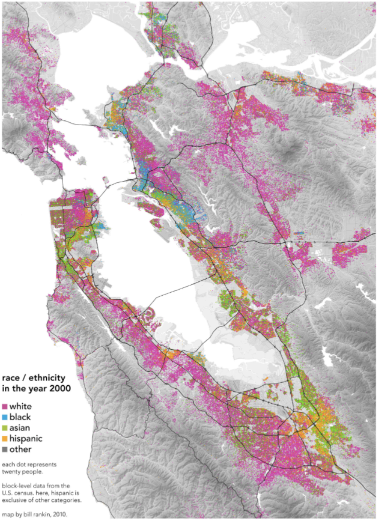
A. Absolute Upward Mobility: Average Child Rank for Below-Median Parents $(\bar{y}_{25})$ by CZ



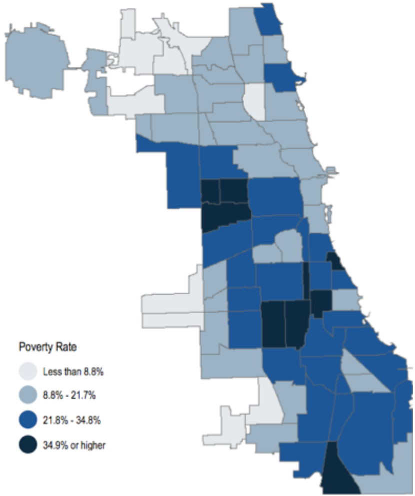
Source: Chetty-Hendren-Kline-Saez (2014)

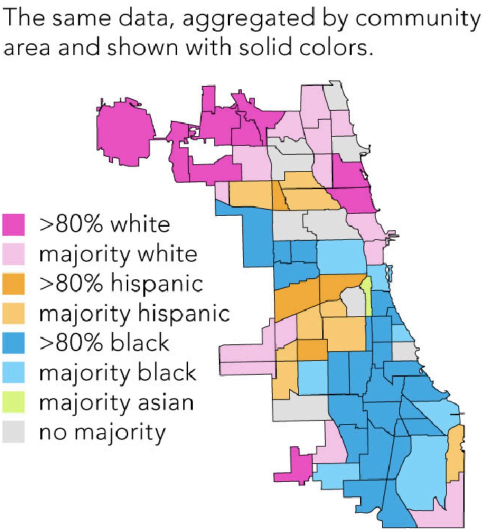
Motivation: Geographically concentrated poverty/race





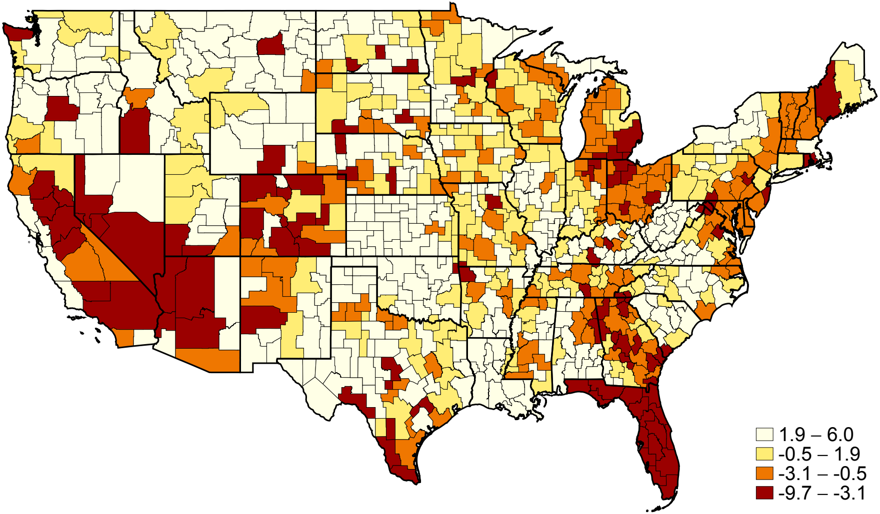
Source: Rankin (2010) using 2000 Census (http://www.radicalcartography.net/)



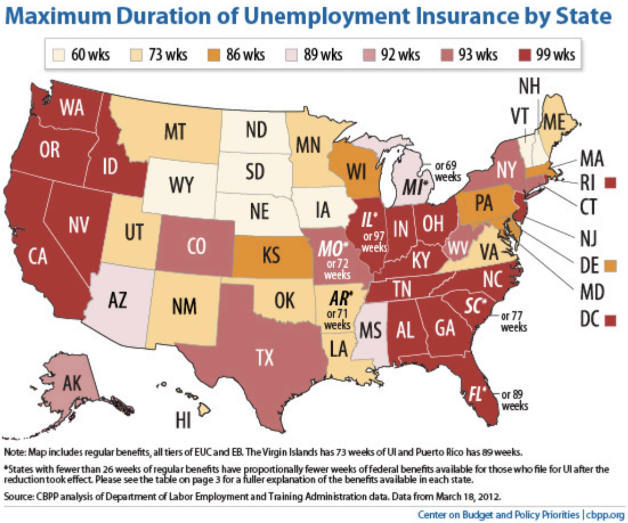


Source: Rankin (2010) and http://capitolfax.com/2013/01/17/todays-maps-illinois-poverty/ using 2010 Census

Motivation: Geographically concentrated recessions



Source: Yagan (2016)



Source: CBPP (2012)

Outline: Place-Based Policies

- Key questions:
  1. Equality: Is place useful for directing redistribution / social insurance?
  2. Efficiency: Can place-based policies increase output?

- Baseline spatial equilibrium theory/evidence in Rosen (1979)–Roback (1982) tradition

- New wave of spatial equilibrium theory/evidence

- Place-based policies over the business cycle

Reminder on tagging

- Mirrlees (1971): Full redistribution if planner had full information
- Akerlof (1978): Ease equity-efficiency tradeoff by conditioning transfers/taxes on relatively *immutable* personal characteristics
- Policymakers constrained in choice of tags (Mankiw-Weinzierl-Yagan 2009, Weinzierl 2014)
- Rich history of place-based policies (state/local governments, inner cities, stimulus, disaster relief) but place is not immutable

Rosen (1979)–Roback (1982): Basic spatial equilibrium

- Original goal: Value nonmarketed amenities (city "quality of life")
- Subsequently used to explain economic geography (prices, quantities, growth)
- Here: Simplified Roback, following notation in Kline (2010)

Rosen (1979)-Roback (1982): Setup

- Measure-one continuum of workers choosing one of two cities $j \in \{1, 2\}$
    - Inelastically supply one unit of labor locally, earning city-specific wage w_j (independent of population / no downward-sloping demand)
    - Inelastically demand one unit of housing
    - Identical quasi-linear preferences over consumption and city-specific amenity:

$$u_{ij} = w_j - r_j + A_j$$

- Landlords supply housing according to weakly increasing housing supply function:

$$r_j = g_j \left(L_j \right)$$

Rosen (1979)-Roback (1982): Equilibrium

- Spatial equilibrium:

$$g_1(L_1) - g_2(1 - L_1) =$$

$$(w_1 + A_1) - (w_2 + A_2)$$

- Difference in real wages pins down difference in amenity values:

$$A_1 - A_2 = (w_2 - r_2) - (w_1 - r_1)$$

Rosen (1979)-Roback (1982): Robustness

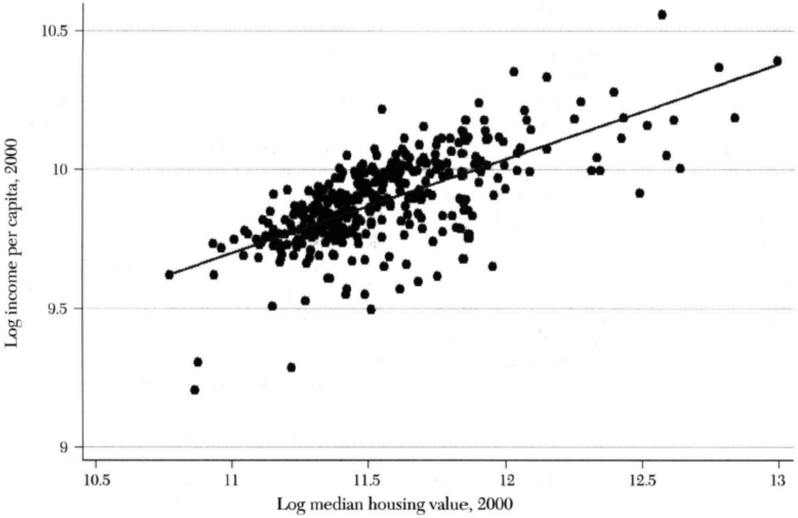
- Criticism 1: "Frictionless mobility is a crazy assumption"
- Response: All results hold even when there's only epsilon frictionless movers

- Criticism 2: "Flexible prices is a crazy assumption"
- Response: Americans are mobile and will vote with their feet (1/3 of adults do not live in birth state, Molloy-Smith-Wozniak 2011)

Rosen (1979)-Roback (1982): Predictions

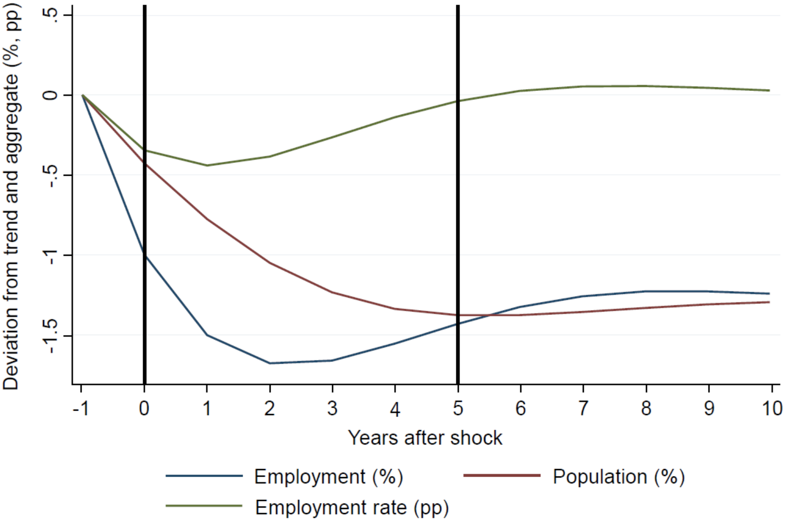
- Expensive cities pay commensurately higher nominal wages (nominal wage variance $>$ real wage variance)
- Contraction in labor demand $\rightarrow$ out-migration until parity in worker outcomes is restored across space

Glaeser-Gottlieb (2009): Income and home prices



Source: Glaeser-Gottlieb (2009) (Slope = 0.34, close to housing expenditure share of 0.41)

Blanchard-Katz (1992): Impulse-response to demand shock



Source: Yagan (2016), replicating Blanchard-Katz

Rosen-Roback: Implications for place-based policies

- RR implication: Place is a terrible tag
  - Even though poor people live in Detroit, a $1 subsidy to anyone living in Detroit raises Detroit rent by $1
- Worker utility is always equal across space: $u_i = \bar{u} \ \forall i$
- See extensions in Glaeser (2008) with similar punchlines ("help poor people, not poor places")

Kline (2010): Consumer surplus in spatial equilibrium

- Rosen-Roback: Everyone has identical preferences for places (amenities not person-specific)
- Kline (2010) and Moretti (2011) (building on Glaeser 1998): Heterogeneous preferences for places and elastic housing supply
  - Workers earn surplus/rents in equilibrium ($u_i \neq \bar{u}$)
  - A place-specific policy benefits incumbents

Kline (2010): Consumer surplus in spatial equilibrium

- Workers now have mean-zero idiosyncratic taste for local amenities ε_{ij}:

$$u_{ij} = w_j - r_j + A_j + \varepsilon_{ij}$$

- Denote relative preference for city 2 $\xi_i \equiv \varepsilon_{i2} - \varepsilon_{i1}$, $\xi_i \sim F(\cdot)$

- Number of workers living in city 1 is now:

$$L_1 = F\left[(w_1 - r_1 + A_1) - (w_2 - r_2 + A_2)\right]$$

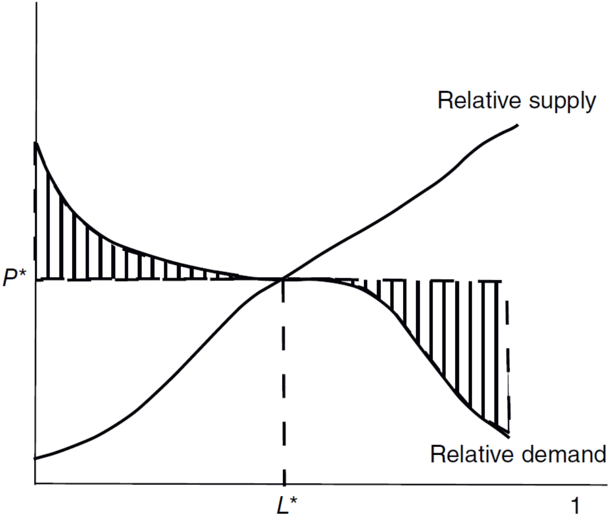
Kline (2010): Consumer surplus in spatial equilibrium

- Spatial equilibrium:

$$g_1(L_1) - g_2(1 - L_1) = (w_1 + A_1) - (w_2 + A_2) - F^{-1}(L_1)$$

- RHS = Relative supply = difference in cost of new housing in 1 relative to 2
- LHS = Relative demand = difference in value of living in 1 relative to 2 *for marginal mover*

Kline (2010): Consumer surplus in spatial equilibrium



Source: Kline (2010)

Kline (2010): Consumer surplus in spatial equilibrium

- Let $\bar{\xi} \equiv F^{-1}\left(L_1^*\right)$ denote marginal mover's relative preference for 2

- Consumer surplus:

$$\int_{\bar{\xi}}^{\infty} \left(\xi_i - \bar{\xi}\right) f\left(\xi_i\right) d\xi_i + \int_{-\infty}^{\bar{\xi}} \left(\bar{\xi} - \xi_i\right) f\left(\xi_i\right) d\xi_i$$
$$= \int_{-\infty}^{\infty} |\xi_i - \bar{\xi}| f\left(\xi_i\right) d\xi_i$$

- Zero surplus if no taste heterogeneity (i.e. $\xi_i = \bar{\xi} \ \forall i$) and thus relative demand is completely elastic and $u_i = \bar{u}$
- Very large surplus if very heterogeneous tastes (value of choice always increasing in taste heterogeneity)

Kline (2010): Subsidy to living in city 1

- Foreign government starts giving \$1 to anyone living in city 1. Are city 1's original residents made better off? [DRAW GRAPHS]
- Extreme case A (Rosen-Roback): No taste heterogeneity $\rightarrow$ rent in city 1 rises by \$1 $\rightarrow$ only city 1 landlords benefit (city 2 landlords lose)
- Extreme case B: Inelastic housing supply $\rightarrow$ rent in city 1 rises by \$1 $\rightarrow$ only city 1 landlords benefit (city 2 landlords unaffected)
- General case: Original residents benefit (i.e. place is effective tag) to extent that tastes are heterogeneous (inelastic relative demand) and housing supply is elastic
  - Testable in cross section: Small rent increase, small population increase

Kline (2010): General comments

- Rosen-Roback: Unexplained cross-city differences in real wages pin down unobserved difference in amenity value (“Name that residual”)

- Kline: Both amenities and preferences

- Other forms of heterogeneity? Do we care?
  - Wages (Moretti 2011)
  - Moving costs (Topel 1986; Bound-Holzer 2000)
  - Segmented housing markets

Busso-Gregory-Kline (2013): Welfare analysis

- Goal: Welfare analysis of a place-based policy
- Context: 1990s Empowerment Zones (poor census tract groups)
  - 20% subsidy (tax credit) to wages of residents who lived and worked in EZ
  - Large block grant

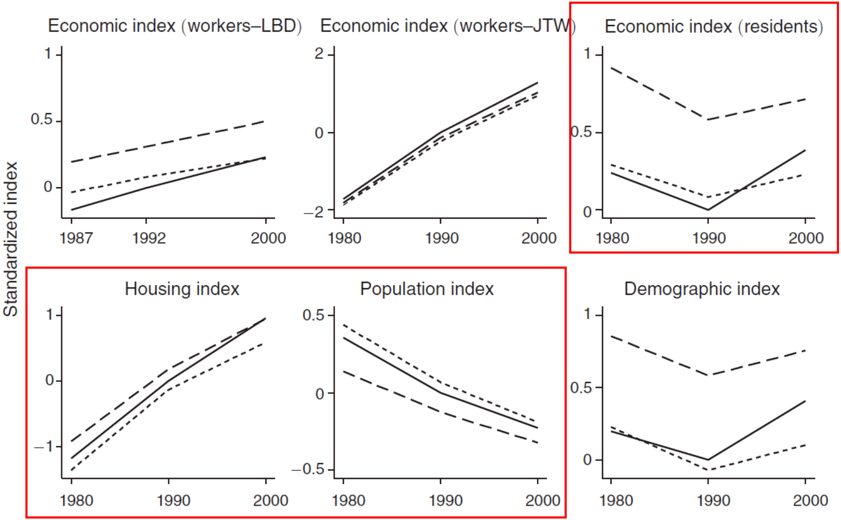
Empowerment Zone example

Source: Busso-Gregory-Kline (2013)

Busso-Gregory-Kline (2013): Strategy

- Challenges
  - Challenging to handle income heterogeneity in spatial equilibrium
  - Income heterogeneity is important (presumably why EZ program exists)
  - Cross-sectional data
- Strategy: Follow older PE literature (e.g. Feldstein 1999)
  - Estimate tract-level regressions and estimate DWL valuing dollars equally across agents (i.e. mostly ignore heterogeneity)
  - Use model to infer incidence based on equilibrium outcomes (prices, quantities)
  - Leave it to reader to weigh incidence vs. DWL

Design: EZ-accepted vs. EZ-denied



Source: Busso-Gregory-Kline (2013) (Similar to Greenstone-Hornbeck-Moretti 2010 and Abadie-Gardeazabal 2003.)

Busso-Gregory-Kline (2013): Results

- Reduced-form
  - Large increase in employment and wages
  - No increase in population, some increase in rent levels
- Implications
  - Small DWL (no population change)
  - Benefits to workers likely accrued by original residents

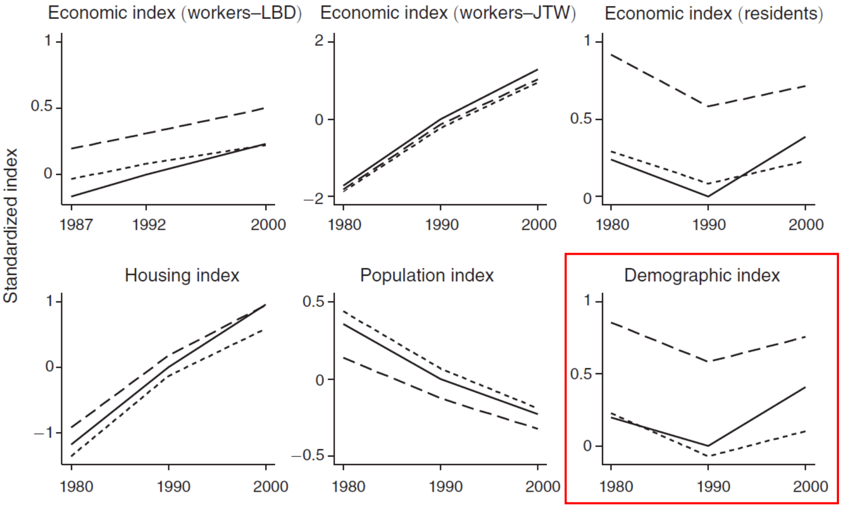
Sidebar: You don't know why your paper is interesting

- Ex. Busso-Gregory-Kline title progression:
  - Original: "Do Local Economic Development Programs Work? Evidence from the Federal Empowerment Zone Program"
  - Published: "Assessing the Incidence and Efficiency of a Prominent Place-Based Policy"
- Ex. Yagan (2015) title progression
  - Original: "Quasi-Experimental Evidence on Dividend Taxes, Payouts, Investment, and Employment"
  - Later: "Did the 2003 Dividend Tax Cut Stimulate Investment? Quasi-Experimental Evidence"
  - Published: "Capital Tax Reform and the Real Economy: The Effects of the 2003 Dividend Tax Cut"
- Good tactics: Start with solid core. Engage non-PE person in 15 seconds. Write/rewrite introduction. Write one-paragraph referee report (why will someone cite/teach it?)

Thought experiment: What if BGK had panel data?

- Broadly two types of incidence analysis
  - Cross-section: Model-based inference from equilibrium prices and quantites (e.g. BGK)
  - Panel: Reduced-form inference from treatment-vs-control DD (ITT/TOT) (e.g. recent papers Reed Walker 2012, Autor-Dorn-Hanson-Song 2014, Yagan 2016)
- What would BGK have done with panel data? (see Tong-Zhou in-progress)

Design: EZ-accepted vs. EZ-denied

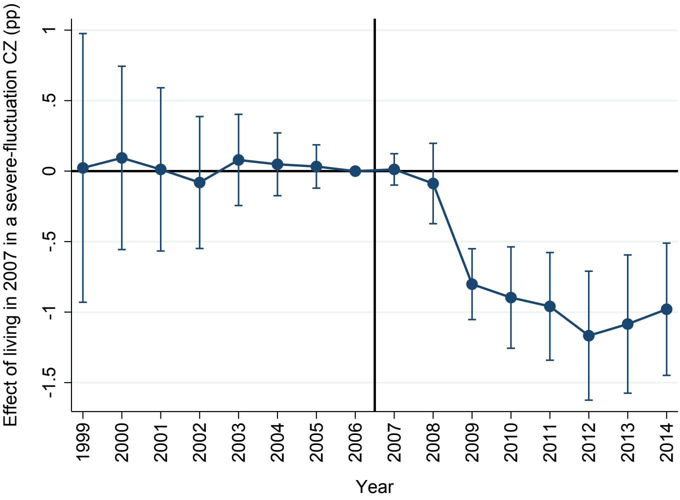


Source: Busso-Gregory-Kline (2013) (Similar to Greenstone-Hornbeck-Moretti 2010 and Abadie-Gardeazabal 2003.)

Ex. Place-based policies after economic shocks

- U.S. extends state unemployment insurance durations when state unemployment rate reaches thresholds (26 weeks $\rightarrow$ 99 weeks)
- Rationale: helps affected workers without getting baked into rents
  - UI receipt is not place-based: can move and continue receipt (good tag since based on past location)
  - Blanchard-Katz suggests there is no long-run need
  - UI *eligibility* is place-based so RR suggests would be costly
- But does spatial equilibrium actually insure workers in long run?
  - Blanchard-Katz: Could miss composition effects, recessions could be different

Enduring employment impact of great recession location



Source: Yagan (2016)

Place-based policy responses to economic shocks?

- Large “location lottery” in spite of spatial equilibrium (Yagan 2016)
  - Uniform city wage and frictionless mobility are both dangerous assumptions even in longer term
  - Past location may be useful tag for DI, not just UI (Black-Daniel-Sanders 2002)
  - Rationale for place-based stimulus spending (Mundell 1961)

- Reminder: NOT spatial DISequilibrium. Just not Rosen-Roback.

Kline-Moretti (2014): Efficiency gains?

- BGK: Is it costly to redistribute to geographically concentrated poor?
- Kline-Moretti: Are there efficiency gains/losses to place-based policies?
    - Intuition: Externalities or "big push" to better equilibrium (Krugman 1991; Kline 2010)

Kline-Moretti (2014): Efficiency gains?

- Context: Tennessee Valley Authority during 1930s
- Design: Similar to BGK (TVA vs. six nominated but unapproved authorities)
- Find that TVA region has:
    - Higher long-run manufacturing employment (agglomeration economies)
    - No "big push" result on net: no aggregate efficiency gain from reallocating manufacturing activity because of constant agglomeration elasticity

Incidence of state taxes

- Relatively underexplored (Moretti-Wilson 2015)
- Of major relevance to governors/mayors
- Potential for race to bottom (Gordon 1983; Ossa 2015)
  - Ex. Goolsbee-Maydew (2000) "Coveting Thy Neighbor's Manufacturing"

Incidence of state business taxes

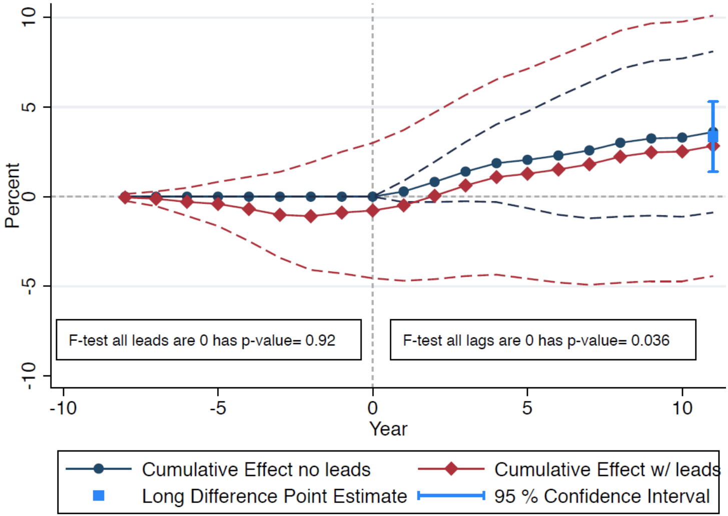
- Baseline model: State business tax is not borne by firm owners
- Huge variation in state corporate (and personal and capital gains) taxes
  - e.g. 10% in CA, 0% in WA/NV
  - Yet investment (and savers) have not all fled to WA/NV $\rightarrow$ rents to tax
- Suarez-Serrato-Zidar (2016): 40% borne by firm owners, 25-30% by landowners, 30-35% by workers

SS-Z: Stratagy

- Challenges: Similar to BGK but more complicated
  - Want to allow firms to bear incidence (idiosyncratic preferences of workers *and* firms)
  - Allow for downward-sloping local labor demand
  - Complicated "apportionment" system

- Strategy: Similar to BGK
  - Use apportionment system to advantage (other states' tax changes affect firms/workers in my state)

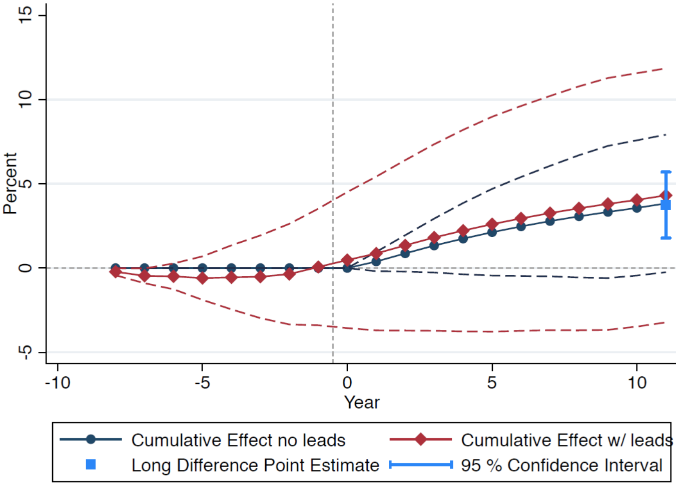
SS-Z: Firms are similarly mobile...



Source: Suarez-Serrato-Zidar (2016)

SS-Z: ...as workers

Figure A7: Cumulative Effects of Business Tax Cuts on Population Growth



Source: Suarez-Serrato-Zidar (2016)

Spatial equilibrium in PE

- PE:
  - Policy incidence
  - Optimal policy
  - Sometimes efficiency
  - Reduced-form empirics
- Not PE:
  - Pure economic geography
  - Non-policy incidence