Motivation: Large and rising capital income share

Figure 5.2: Capital shares in factor-price national income 1975-2010

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

G-7 Dividend Income Tax Rates

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

- Atkinson-Stiglitz (1976): PF 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”):

$$\max_{c_1, c_2, l} U^i(v(c_{i1}, c_{i2}), l_{i1})$$

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

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

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

- Optimum reached with $\tau_k = 0$
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 (Merrlees 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)
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:

\[
c_t + g_t + k_{t+1} \leq F(k_t, l_t) + (1 - \delta) k_t
\]
Define “competitive equilibrium”: sequence of allocations \( \{c_t, l_t, k_t, b_t\} \) where (1)-(4) are satisfied

Government’s problem:

\[
\max_{\tau_{lt}, \tau_{kt}, \tau_{ct}} \sum \beta^t U(c_t(\tau_{lt}, \tau_{kt}, \tau_{ct}), l_t(\tau_{lt}, \tau_{kt}, \tau_{ct}))
\]

where \( c_t(\tau_{lt}, \tau_{kt}, \tau_{ct}) \) and \( l_t(\tau_{lt}, \tau_{kt}, \tau_{ct}) \) 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:

\[ \frac{-U_{l,t}}{w_t U_{c,t}} = \frac{1 - \tau_{lt}}{1 + \tau_{ct}} \]

\[ \frac{U_{c,t-1}}{\beta (1 + r_t - \delta) U_{c,t}} = \frac{(1 - \tau_{kt}) (1 + \tau_{ct-1})}{1 + \tau_{ct}} \]

- 2 * ∞ equations + 3 * ∞ unknowns \( \Rightarrow \) must arbitrarily assign ∞ of them

- Here, assume \( \tau_{ct} = 0 \) \( \forall t \), but note equivalence between capital taxes and increasing consumption taxes
Maximize social welfare over CE allocations (where $a_t \equiv (k_t + b_t) U_{c,t-1}$ for convenience):

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

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

FOCs:

\[
\begin{align*}
\{c_t\} & : \quad \beta^t U_{c,t} + \theta_t [U_{c,t} + U_{cc}, c_t + U_{cl}, t l_t] = \lambda_t \\
\{k_{t+1}\} & : \quad \lambda_{t+1} [1 + F_{k,t+1} - \delta] = \lambda_t \\
\{a_{t+1}\} & : \quad \theta_t = \beta \theta_{t-1}
\end{align*}
\]
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:

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

Recall consumer’s intertemporal FOC:

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

Since $r^{ss} = F_k^{ss}$, these two conditions imply $\tau_k^{ss} = 0$. 
• Transition: highest possible capital tax early on, declining thereafter

• Intuition: $\tau^ss_k > 0 \rightarrow$ ever-increasing unbounded tax on future consumption, and would rather smooth distortions (Judd 1999)

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

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

• Influential because of dynamic general equilibrium: consumers, firms, and government earn and spend in every period
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 \((1 - \tau_{ss}^l) w_{ss}^l \) is maximized with \( \tau_k^{ss} = 0 \) and \( \tau_i^{ss} > 0 \)

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 → 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
• 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 $\tau_{kss} = 0$ 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 → $\tau_k = \tau_l$ (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


- Modigliani: does not capitalize inherited wealth → inheritance share is 20-30%

- Kotlikoff-Summers: do capitalize inherited wealth, even if heirs consume out of it → 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.?)
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) → potential rationale for estate taxation (bequest is not worth much to donor)

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

• Social/family pressure → 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) → 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) → 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 \( \varepsilon_B \) is the inheritance tax elasticity of bequests

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

- But value of \( \varepsilon_B \) is unresolved empirically (Kopczuk-Slemrod 2001)
Potential for small tax elasticity of savings

- Classic two-period consumption model:

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

  s.t.  $$c_1 + c_2 / (1 + r) \leq z_1 + z_2 / (1 + r)$$

- With capital tax, budget constraint becomes:

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

\[ \beta = 0.62 \pm 0.09 \]

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

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 $r = F_k$, $\delta = \text{social (gov.) discount rate and CRRA}^{(\gamma)} \text{ utility curvature } u'(c) = c^{-\gamma}$
  - 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 $\delta$? [Nordhaus vs. Stern]
  - Reasonable upper bound: private sector rate of 1.4%
    (Giglio-Maggiori-Stroebel 2015)
Measuring very long discount rates

Source: Giglio-Maggiori-Stroebel 2015

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.
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 $\delta$? [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 $\gamma$?
  - $\gamma$ high $\rightarrow$ care a lot about inequality $\rightarrow$ want small capital stock and thus large $r$ $\rightarrow$ global warming is not important
  - $\gamma$ low $\rightarrow$ do not care about inequality $\rightarrow$ want large capital stock and thus small $r$ $\rightarrow$ should care a lot about global warming

- Real world: $r > g$ $\rightarrow$ below socially optimal level of capital unless $\delta$ or $\gamma$ 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)
### TABLE 2

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

<table>
<thead>
<tr>
<th>Year</th>
<th>Gross profit</th>
<th>Gross investment</th>
<th>$D/V$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$V$</td>
<td>$V$</td>
<td></td>
</tr>
<tr>
<td>1953</td>
<td>29.1</td>
<td>14.7</td>
<td>13.6</td>
</tr>
<tr>
<td>1954</td>
<td>28.7</td>
<td>13.3</td>
<td>14.3</td>
</tr>
<tr>
<td>1955</td>
<td>25.1</td>
<td>13.3</td>
<td>10.2</td>
</tr>
<tr>
<td>1956</td>
<td>21.0</td>
<td>12.3</td>
<td>7.0</td>
</tr>
<tr>
<td>1957</td>
<td>19.4</td>
<td>10.8</td>
<td>7.5</td>
</tr>
<tr>
<td>1958</td>
<td>20.6</td>
<td>9.6</td>
<td>10.0</td>
</tr>
<tr>
<td>1959</td>
<td>18.2</td>
<td>9.8</td>
<td>7.6</td>
</tr>
<tr>
<td>1960</td>
<td>16.7</td>
<td>9.2</td>
<td>6.3</td>
</tr>
<tr>
<td>1961</td>
<td>16.9</td>
<td>8.9</td>
<td>7.5</td>
</tr>
<tr>
<td>1962</td>
<td>15.5</td>
<td>8.9</td>
<td>6.9</td>
</tr>
<tr>
<td>1963</td>
<td>18.0</td>
<td>9.8</td>
<td>8.0</td>
</tr>
<tr>
<td>1964</td>
<td>17.1</td>
<td>9.5</td>
<td>7.3</td>
</tr>
<tr>
<td>1965</td>
<td>17.5</td>
<td>10.5</td>
<td>6.9</td>
</tr>
<tr>
<td>1966</td>
<td>17.4</td>
<td>11.7</td>
<td>5.9</td>
</tr>
<tr>
<td>1967</td>
<td>18.4</td>
<td>11.6</td>
<td>6.9</td>
</tr>
<tr>
<td>1968</td>
<td>16.1</td>
<td>9.7</td>
<td>6.2</td>
</tr>
</tbody>
</table>

Source: Abel, Mankiw, Summers, Zeckhauser (1989)
Figure 5.5. After tax rate of return vs. growth rate at the world level, from Antiquity until 2100

The rate of return to capital (after tax and capital losses) fell below the growth rate during the 20th century, and may again surpass it in the 21st century.

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 → 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)
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 \(c_1, c_2, w_l\)

- 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

\[
K \left( \int x(w) f(w) \, dw \right) < \int K(x(w)) f(w) \, dw
\]

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

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

\[
u'(c_1) < \beta(1 + r) \int u'(c_2(w)) f(w) \, dw
\]

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

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

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] → 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
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

Organizational Form

Raise Capital

Production

Payouts

Firm’s Decision

S corp or C corp
Where to Locate

Debt or
Equity

Investment
Decisions

Report Profits Pay Dividends Pay Interest

Policy Instruments

Indiv. vs.
Corp. tax,
Intl. tax

Deduction of
interest

Accelerated
Depreciation

Div. tax,
Corp. profit
tax

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)


- 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): $\delta$
  - Required rate of return: $\rho$
Cost of capital (Hall-Jorgenson 1967)

- NPV of a new machine \(dK_{t+1}\):
  \[-q_t - \delta q_t + \frac{F'(K_{t+1}) + q_{t+1}}{1 + \rho}\]

- Euler, equating marginal benefit to marginal cost at optimum:
  \[F'(K_{t+1}) = q_t \left[(1 + \delta)(1 + \rho) - \frac{q_{t+1}}{q_t}\right]\]
  \[F'(K_{t+1}) \approx q_t \left[\rho + \delta - \frac{q_{t+1} - q_t}{q_t}\right]\]

- RHS: “user cost of capital”

- With constant investment prices \(q_{t+1} = q_t\), return on marginal unit of investment \(F'(K_{t+1}) / q_t\) equals required rate of return plus depreciation
Cost of capital (Hall-Jorgenson 1967)

- Add corporate income tax $\tau_{t}^{INC}$ (typically 35% in United States), which is assessed on gross profit (revenue minus deductions)
### The U.S. corporate income tax form

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Compensation of officers (Schedule E, line 4)</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Salaries and wages (less employment credits)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Repairs and maintenance</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Bad debts</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Rents</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Taxes and licenses</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Interest</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Charitable contributions</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Depreciation from Form 4562 not claimed on Schedule A or elsewhere on return (attach Form 4562)</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Depletion</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Advertising</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Pension, profit-sharing, etc., plans</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Employee benefit programs</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Domestic production activities deduction (attach Form 8903)</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Other deductions (attach schedule)</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>27</td>
<td>Total deductions. Add lines 12 through 26</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Taxable income before net operating loss deduction and special deductions. Subtract line 27 from line 11</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Less: a Net operating loss deduction (see instructions)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b Special deductions (Schedule C, line 20)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Taxable income. Subtract line 29c from line 28 (see instructions)</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Total tax (Schedule J, line 10)</td>
<td></td>
</tr>
</tbody>
</table>
Cost of capital (Hall-Jorgenson 1967)

- Add corporate income tax $\tau_t^{INC}$, 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):

\[
D_0 = 1, \quad D_{z-t} = 0 \quad \forall \ (z-t) > 0
\]

\[
\Rightarrow \quad \Gamma_t = \tau_t^{INC}
\]

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

\[
F' (K_{t+1}) = q_t \left[ \rho + \delta - \frac{q_{t+1} - q_t}{q_t} \right]
\]
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 $\pi(K)$ also maximizes $\left(1 - \tau^{INC}\right) \pi(K)$.

- In real world:
  - Tax law allows only small profitable firms to immediately expense ($D_0 < 1$).
  - Tax law does not allow full deductibility of financing costs $\rightarrow \rho'(\tau^{INC}) > 0$ [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.
<table>
<thead>
<tr>
<th>Deductions (See instructions for limitations on deductions)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12 Compensation of officers (Schedule E, line 4)</td>
<td>12</td>
</tr>
<tr>
<td>13 Salaries and wages (less employment credits)</td>
<td>13</td>
</tr>
<tr>
<td>14 Repairs and maintenance</td>
<td>14</td>
</tr>
<tr>
<td>15 Bad debts</td>
<td>15</td>
</tr>
<tr>
<td>16 Rents</td>
<td>16</td>
</tr>
<tr>
<td>17 Taxes and licenses</td>
<td>17</td>
</tr>
<tr>
<td>18 Interest</td>
<td>18</td>
</tr>
<tr>
<td>19 Charitable contributions</td>
<td>19</td>
</tr>
<tr>
<td><strong>20 Depreciation from Form 4562 not claimed on Schedule A or elsewhere on return (attach Form 4562)</strong></td>
<td><strong>20</strong></td>
</tr>
<tr>
<td>21 Depletion</td>
<td>21</td>
</tr>
<tr>
<td>22 Advertising</td>
<td>22</td>
</tr>
<tr>
<td>23 Pension, profit-sharing, etc., plans</td>
<td>23</td>
</tr>
<tr>
<td>24 Employee benefit programs</td>
<td>24</td>
</tr>
<tr>
<td>25 Domestic production activities deduction (attach Form 8903)</td>
<td>25</td>
</tr>
<tr>
<td>26 Other deductions (attach schedule)</td>
<td>26</td>
</tr>
<tr>
<td><strong>27 Total deductions. Add lines 12 through 26</strong></td>
<td><strong>27</strong></td>
</tr>
<tr>
<td><strong>28 Taxable income before net operating loss deduction and special deductions. Subtract line 27 from line 11</strong></td>
<td><strong>28</strong></td>
</tr>
<tr>
<td>29 Less: a Net operating loss deduction (see instructions)</td>
<td><strong>29a</strong></td>
</tr>
<tr>
<td>b Special deductions (Schedule C, line 20)</td>
<td><strong>29b</strong></td>
</tr>
<tr>
<td><strong>29c</strong></td>
<td><strong>29c</strong></td>
</tr>
<tr>
<td><strong>30 Taxable income. Subtract line 29c from line 28 (see instructions)</strong></td>
<td><strong>30</strong></td>
</tr>
<tr>
<td><strong>31 Total tax (Schedule J, line 10)</strong></td>
<td><strong>31</strong></td>
</tr>
</tbody>
</table>
Hall-Jorgenson pins down the optimal capital stock

Predicts that when $\tau^{INC}$ 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)
Quadratic adjustment costs, Cobb-Douglas, and linearize from firm’s steady state → 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

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

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

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

- High adjustment costs → 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 \(\alpha\) outside Cobb-Douglas feasible range \((0, 1)\))
• 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
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 $\Gamma$, 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”):
\[
\left( \frac{I_t}{K_{i,t-1}} \right) - \left( \frac{\hat{I}_t}{K_{i,t-1}} \right) = \mu_i + \beta \left( C_{it} - \hat{C}_{it} \right) + \varepsilon_{it}
\]

- Finds large and significantly negative coefficients in tax-reform years, with implied cost-of-capital elasticity of investment equal to \(-0.66\), \(\sim10\)x 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
## Table 2—Recovery Periods and Depreciation Methods by Type of Capital

<table>
<thead>
<tr>
<th>Type of capital</th>
<th>Recovery period, ( R ) (years)</th>
<th>Tax depreciation rate, ( \delta ) (percent)</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tractor units for over-the-road use, horses over 12 years of age or racehorses with over 2 years in service</td>
<td>3</td>
<td>66.7</td>
<td>200 DB</td>
</tr>
<tr>
<td>Computers and office equipment; light vehicles, buses and trucks</td>
<td>5</td>
<td>40.0</td>
<td>200 DB</td>
</tr>
<tr>
<td>Miscellaneous equipment, office furniture, agricultural equipment</td>
<td>7</td>
<td>28.6 or 21.4</td>
<td>200 DB or 150 DB</td>
</tr>
<tr>
<td>Water transportation equipment (vessels and barges); single-purpose agricultural structures</td>
<td>10</td>
<td>20.0 or 15.0</td>
<td>200 DB or 150 DB</td>
</tr>
<tr>
<td>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</td>
<td>15</td>
<td>10.0</td>
<td>150 DB</td>
</tr>
<tr>
<td>Farm buildings (other than single purpose structures), railroad structures, telephone communications, electric utilities, water utilities structures including dams, and canals</td>
<td>20</td>
<td>7.5</td>
<td>150 DB</td>
</tr>
<tr>
<td>Nonresidential real property (office buildings, storehouses, warehouses, etc.)</td>
<td>39</td>
<td>2.6</td>
<td>SL</td>
</tr>
</tbody>
</table>

*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)
## Table 3—Quantifying Depreciation Allowances

<table>
<thead>
<tr>
<th>Recovery period</th>
<th>Nominal interest rate = 0.03</th>
<th></th>
<th>Nominal interest rate = 0.05</th>
<th></th>
<th>Nominal interest rate = 0.07</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda^m = 0$</td>
<td>$\lambda^m = 0.3$</td>
<td>$\lambda^m = 0.5$</td>
<td></td>
<td>$\lambda^m = 0$</td>
<td>$\lambda^m = 0.3$</td>
</tr>
<tr>
<td>3 years</td>
<td>0.972</td>
<td>0.981</td>
<td>0.986</td>
<td></td>
<td>0.955</td>
<td>0.968</td>
</tr>
<tr>
<td>5 years</td>
<td>0.949</td>
<td>0.964</td>
<td>0.975</td>
<td></td>
<td>0.918</td>
<td>0.943</td>
</tr>
<tr>
<td>7 years</td>
<td>0.927</td>
<td>0.949</td>
<td>0.964</td>
<td></td>
<td>0.884</td>
<td>0.919</td>
</tr>
<tr>
<td>7 years (150DB)</td>
<td>0.914</td>
<td>0.939</td>
<td>0.957</td>
<td></td>
<td>0.863</td>
<td>0.904</td>
</tr>
<tr>
<td>10 years</td>
<td>0.896</td>
<td>0.927</td>
<td>0.948</td>
<td></td>
<td>0.837</td>
<td>0.886</td>
</tr>
<tr>
<td>10 years (150DB)</td>
<td>0.878</td>
<td>0.915</td>
<td>0.939</td>
<td></td>
<td>0.811</td>
<td>0.868</td>
</tr>
<tr>
<td>15 years</td>
<td>0.824</td>
<td>0.877</td>
<td>0.912</td>
<td></td>
<td>0.733</td>
<td>0.813</td>
</tr>
<tr>
<td>20 years</td>
<td>0.775</td>
<td>0.842</td>
<td>0.887</td>
<td></td>
<td>0.667</td>
<td>0.767</td>
</tr>
</tbody>
</table>

**Panel A:** Present value of depreciation allowances: $\lambda^m + (1 - \lambda^m) z^m$

**Panel B:** Tax subsidy due to the bonus depreciation allowance, percent

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?
• 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 → 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) → 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
Table 2: Regular and Bonus Depreciation Schedules for Five Year Items

<table>
<thead>
<tr>
<th>Normal Depreciation</th>
<th>Year</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deductions (000s)</td>
<td></td>
<td>200</td>
<td>320</td>
<td>192</td>
<td>115</td>
<td>115</td>
<td>58</td>
<td>1000</td>
</tr>
<tr>
<td>Tax Benefit ($\tau = 35%$)</td>
<td></td>
<td>70</td>
<td>112</td>
<td>67.2</td>
<td>40.3</td>
<td>40.3</td>
<td>20.2</td>
<td>350</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bonus Depreciation (50%)</th>
<th>Year</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deductions (000s)</td>
<td></td>
<td>600</td>
<td>160</td>
<td>96</td>
<td>57.5</td>
<td>57.5</td>
<td>29</td>
<td>1000</td>
</tr>
<tr>
<td>Tax Benefit ($\tau = 35%$)</td>
<td></td>
<td>210</td>
<td>56</td>
<td>33.6</td>
<td>20.2</td>
<td>20.2</td>
<td>10</td>
<td>350</td>
</tr>
</tbody>
</table>

Source: 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?

(a) Past Estimates

Hassett and Hubbard (2002) range

Source: Zwick-Mahon (2014)
Reconciliation with past estimates: financing constraints?

(b) Estimates by Firm Size, Bonus Sample

Hassett and Hubbard (2002) range

Compustat

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: $\tau^{DIV}$ is equivalent to $\tau^{INC}$

- For compactness, ignore depreciation, uncertainty (so that $\rho = r$, fixed world interest rate), changing capital prices, and adjustment costs. Firm chooses $K$ such that:
  
  $$
  (1 - \tau^{DIV}) (1 - \tau^{INC}) F'(K) = r
  $$
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 $\tau^{DIV}$ affect value but not investment:
  \[
  (1 - \tau^{DIV}) (1 - \tau^{INC}) F'(K) = (1 - \tau^{DIV}) r
  \]

- Firm retains cash for investment ($PASTPROFITS - PAYOUTS$) up to point where $(1 - \tau^{INC}) F'(PASTPROFITS - PAYOUTS) = r$, regardless of $\tau^{DIV}$

- Change in $\tau^{DIV}$ 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 $\tau^{DIV}$ 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

Percent of Top 3807 Firms

Quarter

Source: Chetty-Saez (2005), updated through 2006
Effect of 2003 dividend tax cut on dividend-paying fraction

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

Quarter

Percent of Top 3807 Firms

Source: Chetty-Saez (2005), updated through 2006
Heterogeneity suggestive of agency problems

Source: Chetty-Saez (2005)

Graduate Public Economics

Actual Capital Taxation – Investment

98 / 338
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

![Graph showing U.S. Corporate Investment: NIPA Private Fixed Non-residential Investment. The graph indicates a significant change in corporate investment levels around 1992 when there was no dividend tax cut. There is another significant increase in investment levels from 2000 to 2002, with a partial return to the previous levels by 2008. The source of the data is Yagan (2015).]
After incorporating, a corporation elects either C or S tax status

<table>
<thead>
<tr>
<th></th>
<th>Tax rate on annual income</th>
<th>Tax rate on dividends</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-corporations (treatment)</td>
<td>35%</td>
<td>15%</td>
</tr>
<tr>
<td>S-corporations (control)</td>
<td>35%</td>
<td>0%</td>
</tr>
</tbody>
</table>

S-corporations: < 100 non-institutional investors, one stock class

Operate in same narrow industries and at the same scale throughout United States → common trends
Example: Retail hardware chains

- Largest hardware chain
  - C-corporation

- Third-largest hardware chain
  - S-corporation

Example: Retail hardware chains

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

NAICS 1: Agriculture & Forestry
NAICS 2: Construction & Mining
NAICS 3: Manufacturing
NAICS 4: Retail & Wholesale Trade
NAICS 5: Information & Professional Services
NAICS 6: Health Care
NAICS 7: Entertainment, Food, & Hotels
NAICS 8: Other Services

Zero effects on investment and employee compensation

Effects constant across firm size distribution

Immediate financial response confirms relevance/salience

- 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)
One explanation: New view is correct and most firms fund marginal investments out of retained earnings (e.g. median firm is 22 years old) → 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 → 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 → domestic parent enjoys interest deductions while foreign sub pays little tax on interest earnings
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

The Share of Profits Made Abroad in US Corporate Profits

Source: Zucman (2014)
Rising importance of earnings booked abroad

The Share of Tax Havens in US Corporate Profits Made Abroad

% of US corporate profits made abroad

- Singapore
- Bermuda (and other Caribbean)
- Luxembourg
- Switzerland
- Netherlands
- Ireland

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

US Corporate Profits Retained in Tax Havens

Source: Zucman (2014)
• 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 intentionally mobile, so savings taxes reduce U.S. capital accumulation, wages, and GDP!”
  - (Vice versa for Democrats)
Motivation: Low savings rates among non-rich

Saez-Zucman (2014)

Source: Saez-Zucman (2014)
Motivation: Forced savings and elderly poverty

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)

How do taxes and other policy affect savings?

- Institutions: 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
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

Participation in non-Social-Security pension plans

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): $V_{NT} = (1 - \tau_0^{PERSONAL})(1 + r)^T$
  - Regular taxable account:
    $V_{NT} = (1 - \tau_0^{PERSONAL})(1 + r \left(1 - \tau^{DIV}\right))^T$

- 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:

$$
\max_{c_1, c_2} u (c_1, c_2) \\
\text{s.t. } c_1 + \frac{c_2}{1 + r} \leq z_1
$$

• With capital tax, budget constraint becomes:

$$
c_1 + \frac{c_2}{1 + r (1 - \tau_k)} \leq z_1
$$

• 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 $c_2$

- Substitution 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. $\tau_k = -0.5$)
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

Stable ratio of pension contributions to wages


[Bar chart showing the ratio of pension contributions to wages from 1975 to 1999.]

Legend:
- Private Contributions/Private W&S
- Total Contributions/Total W&S

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 $t / (1 − t)$
- 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

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

Source: Madrian-Shea (2001)
Effect of automatic enrollment on contribution rate

Distribution of contribution rates: Company B

Default contribution rate under automatic enrollment

Source: Madrian-Shea (2001)
- 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 sum 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


Note: $1 \equiv 6$ DKr

\[ \Delta \text{Subsidy} = -14\% \]

Impact of 1999 capital pension subsidy reform

Diff-in-Diff: $\mu^L = -2.449$

Complete crowd-out

Active vs. passive savers


Percent of Individuals

Percent Change in Capital Pension Contributions

• 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.

Active vs. passive savers

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

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)
Motivation: Modern governments do social insurance

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

1953

- National Defense: 69.4%
- Other: 21.6%
- Social Security: 3.6%
- Income Security: 5%
- Health: 0.4%

2008

- National Defense: 20.7%
- Other: 34.8%
- Social Security: 20.7%
- Income Security: 14.5%
- Health: 9.4%

Source: Chetty; Office of Management and Budget, historical tables
Motivation: Developed governments do social insurance

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:
    \[
    \max_{\{c_w, l_w\}} \int u(c_w, l_w) f(w) \, dw \\
    \text{s.t. } \{c_w, l_w\} = \arg \max u(wl_w - T(wl_w), l_w) \quad \forall w \quad \text{(IC)} \\
    \text{and } \int T(wl_w) f(w) \, dw = 0 \quad \text{(RC)}
    \]
  
- 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 (Miryees 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)
- Adverse 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

$0 if highest total quarterly earnings < $2,871 ($220/wk)

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


- 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:

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

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

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

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

s.t. $e(b) t = (1 - e(b)) b$
Second Best Problem

**Problem**

*Optimal Social Insurance*

\[
\max_b V(b, t(b)) \\
\text{s.t. } e(b)t(b) = (1 - e(b))b \\
e(b) = \arg \max_e e \cdot u(A + w_h - t) + (1 - e) \cdot u(A + w_l + b) - \psi(e)
\]

- Formally equivalent to an optimal Ramsey tax problem with state-contingent taxes
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
At an interior optimum, the optimal benefit rate must satisfy

\[ \frac{dV}{db}(b^*) = 0 \]

To calculate this derivative, write \( V(b) \) as

\[ V(b) = \max_e e u(A + w_h - t(b)) + (1 - e) u(A + w_l + b) - \psi(e) \]

Since \( f_n \) has been optimized over \( e \), Envelope Thm. implies:

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

Can ignore \( \frac{\partial e}{\partial b} \) terms because of agent optimization.
• Exploiting f.o.c.’s from agent optimization particularly useful in more complex models

• Kaplan (2009): unemployed youth move back in with their parents.
  
  • How does this affect optimal UI?

• Kaplan takes a structural approach and estimates a dynamic model of the decision to move back home
Suppose moving home raises consumption by $H$ and has a cost $g(H)$:

$$V(b) = \max_{e,H} eu\left(A + w_h - t(b)\right)$$

$$+ (1 - e)\left[u(A + w_l + b + H) - g(H)\right] - \psi(e)$$

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 ($H$).
The government’s UI budget constraint implies

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

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

Setting \( \frac{dV(b)}{db} = 0 \) 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:
  2. Shimer and Werning (2007): reservation wages
Consumption-Based Formula

1. Write marginal utility gap using a Taylor expansion

\[ u'(c_l) - u'(c_h) \approx u''(c_h)(c_l - c_h) \]

2. Defining coefficient of relative risk aversion \( \gamma = \frac{-u''(c)c}{u'(c)} \), we can write

\[
\frac{u'(c_l) - u'(c_h)}{u'(c_h)} \approx -\frac{u''}{u'} c_h \frac{\Delta c}{c} \\
= \gamma \frac{\Delta c}{c}
\]  \hspace{1cm} (1)

3. Gap in marginal utilities is a function of curvature of utility (risk aversion) and consumption drop from high to low states
The optimal unemployment benefit level $b^*$ satisfies

$$
\gamma \frac{\Delta c}{c} (b^*) \approx \frac{\varepsilon_{1-e,b}}{e}
$$

where

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

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

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

---

Diagram:

Period $t$

- $S_t$
- $1 - S_t$
- $C_t^u$
- $C_t^e = C_{t+1}^e = ...$
- $S_{t+1}$
- $1 - S_{t+1}$
- $C_{t+1}^e$
- $C_{t+1}^u$
• 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:

\[ \partial s_t / \partial b_t = -u'(c_t^u) / \psi''(s_t) \]
Effect of benefits on durations can be decomposed into two terms:

\[ \frac{\partial s_t}{\partial b_t} = \frac{\partial s_t}{\partial A_t} - \frac{\partial s_t}{\partial w_t} \]

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

and \( \frac{\partial s_t}{\partial w_t} = \frac{u'(c_t^{e})}{\psi''(s_t)} > 0 \) ("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
\[ \frac{\partial s_t}{\partial A_t} = \{u'(c^e_t) - u'(c^u_t)\} / \psi''(s_t) \geq 0 \]

\[ \frac{\partial s_t}{\partial w_t} = u'(c^e_t) / \psi''(s_t) > 0 \]

\[ \Rightarrow \frac{\partial s_t}{\partial A_t} = \frac{\text{LIQ}}{\text{MH}} = \frac{u'(c^u_t) - u'(c^e_t)}{u'(c^e_t)} \]

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

\[ \frac{u'(c^u_t) - u'(c^e_t)}{u'(c^e_t)} = \frac{\varepsilon_{1-s_t,b}}{s_t} \]

- Combining yields formula that depends solely on duration elasticities:

\[ \frac{\partial s^*_t}{\partial A_t} = \frac{\varepsilon_{1-s_t,b}}{s_t} \]

\[ \frac{\partial s^*_t}{\partial b_t} - \frac{\varepsilon_{1-e,A}}{\varepsilon_{1-e,b,A} - \varepsilon_{1-e,A}} = \frac{\varepsilon_{1-s,b}}{s} \]
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)
Figure 3a

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

Wilcoxon Test for Equality: p = 0.01

Source: Chetty 2008
Figure 3b

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

Fraction Unemployed

Mean rep. rate = .48
Mean rep. rate = .53

Wilcoxon Test for Equality: p = 0.04

Weeks Unemployed

Avg. UI benefit below mean
Avg. UI benefit above mean

Source: Chetty 2008
Figure 3c

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

Wilcoxon Test for Equality: p = 0.69

Source: Chetty 2008
Figure 3d

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

Wilcoxon Test for Equality: $p = 0.43$

Mean rep. rate = .52

Mean rep. rate = .43

Source: Chetty 2008
Figure 6a

Effect of Severance Pay on Durations: Below Median Net Wealth

Source: Chetty 2008
Figure 6b
Effect of Severance Pay on Durations: Above Median Net Wealth

Source: Chetty 2008
Use elasticities to compute ratio of liquidity effect to total (liquidity + moral hazard) effect ($\approx 0.6$) 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

![Graph showing UI Tax Rate (%) vs Benefit Ratio (100*UI Benefits Paid/Payroll). The graph compares Washington’s UI Tax Schedule to a Perfect Experience Rating.](image)
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

Source: Schmieder-Wacther-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-Wacther-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-Wacther-Bender (2012)
Outline: Adverse selection and health insurance

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 $\alpha_2$ in loss state

- Individuals vary only in probability $p_i$ of incurring loss

- Diminishing marginal utility (not assumed in used car application) → 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
  \[
  V_i(\alpha) = (1 - p_i)u(E_1 - \alpha_1) + p_iu(E_2 + \alpha_2)
  \]

- 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.
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 $\alpha$.
  2. **Separating**: high-risk types choose a contract $\alpha_H$ while low-risk types choose a different contract $\alpha_L$. 
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 $(1 - p_i)E_1 + p_iE_2$ regardless of the state.
Equilibrium with Perfect Information

\[ \text{Slope} = \frac{1-p}{p} \]

\[ MRS_{12} = \frac{u'(c_1)(1-p)}{u'(c_2)p} \]

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
• **Result #1:** no pooling equilibrium exists

• If H and L types are pooled in a contract $\alpha$, low-risk types lose money in expectation.

• Zero-profit condition requires $\alpha_2 = \frac{1-\bar{p}}{\bar{p}} \alpha_1$ but $\bar{p} > p_L$.

  • Low-risk type gets fewer dollars in state 2 than he should if the insurance were fair for him.

• Creates an opportunity for a new insurer to enter and “pick off” low-risk types by offering slightly less insurance at a better price: higher $c_1$, lower $c_2$.

  • Only low-risk types switch, because they value $c_1$ more.
No Pooling Equilibrium with Asymmetric Information

Source: Rothschild and Stiglitz 1976
Rothschild-Stiglitz: Second Best Solution

- **Result #2**: in a separating eq, Type H obtains full insurance and Type L is under-insured

- **Intuition**: in any sep. eq., both types are getting actuarially fair insurance because of the zero-profits condition
  
  - For H, no cost to firm in providing full ins (worst that can happen is that L will join the pool, raising profits)
  
  - But for L, full ins. would create an incentive for H to buy this (cheaper) policy, forcing firm into negative profits

- Incentive constraints always bind downward—“no distortion at the top” result in standard asymmetric info. models

- In eq., L gets as much 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

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

- 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


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


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**

<table>
<thead>
<tr>
<th>EXPLANATORY VARIABLE</th>
<th><strong>ESTIMATES FROM HAZARD MODEL OF MORTALITY AFTER PURCHASING AN ANNUITY</strong></th>
<th><strong>ESTIMATES FROM LINEAR PROBABILITY MODEL OF PROBABILITY OF DYING WITHIN FIVE YEARS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Compulsory Market</strong> (1)</td>
<td><strong>Voluntary Market</strong> (2)</td>
</tr>
<tr>
<td>Index-linked</td>
<td>−.839*** (0.217)</td>
<td>−.894** (0.358)</td>
</tr>
<tr>
<td>Escalating</td>
<td>−1.085*** (0.113)</td>
<td>−1.497*** (0.253)</td>
</tr>
<tr>
<td>Guaranteed</td>
<td>.019 (0.029)</td>
<td>.216*** (0.060)</td>
</tr>
<tr>
<td>Capital-protected</td>
<td>... (0.051)</td>
<td>.056</td>
</tr>
<tr>
<td>Payment (£100s)</td>
<td>−.003*** (0.0006)</td>
<td>.001** (0.0004)</td>
</tr>
<tr>
<td>Male Annuitant</td>
<td>.640*** (0.039)</td>
<td>.252*** (0.051)</td>
</tr>
<tr>
<td>Observations</td>
<td>38,362</td>
<td>3,692</td>
</tr>
<tr>
<td>Number of deaths in</td>
<td>6,311</td>
<td>1,944</td>
</tr>
</tbody>
</table>

*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 unraveled
   - 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 $\psi_i$

- To max social welfare, not desirable for those with high $\psi_i$ 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
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
Motivation: One-third of U.S. gov. spending is subnational

Share of total government spending

- Federal
- State
- Local

Source: Gruber (2013)
Motivation: Subnational govs do different things

Source: Gruber (2013)
Motivation: Subnational govs tax property and sales

Source: Gruber (2013)

[Chart showing sources of revenue for 2008]
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 equilibrium 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-\(\alpha\) type (high taxes with great schools) while others do opposite
  - Consumers of each type move to city satisfying MRS = MRT: \(\alpha v'(g_c/n_c) = 1\)
  - Allocation is same as if consumers bought publicly provided good on private market!
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:
  \[
  \max_{c \in C} z - t_c + \alpha_i v(g_c)
  \]

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

- Free-rider and coordination problem without differentiated taxes
  - A single large city can be second-best
  - But for sufficiently wide variation in \(\alpha\), another “entrepreneurial” (Bewley 1981) city can pick off the lowest-\(\alpha\) 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 $i = \tau^\text{PROP}_c \times \text{ASSESSEDHOMEVALUE}_i$

- 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 Map
Village of Winnetka

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 jurisdictions had higher pre-existing tax rates → 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 → $7 increase in house price)

- At interest rate of 12-15%, which implies full capitalization ($1/.135 ≈ 7$)

- 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”? 
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

Effect of living in 2007 in a severe-fluctuation CZ (pp)

Year

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 → need to pool pre-treatment years and post-treatment years in main DD specification.
- Should still show year-by-year version → separate identification demonstration (common trends) from inference (main specification).
Year-by-year estimates in pooled DD setting

• 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
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 → “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
  - $\text{TOT} = \text{ITT} - \text{[cumulative effect of bond passage on fewer subsequent approvals]}$
Source: Cellini-Ferreira-Rothstein (2010)
CFR: ITT

Source: Cellini-Ferreira-Rothstein (2010)
Source: Cellini-Ferreira-Rothstein (2010)
CFR: TOT only somewhat larger than ITT

Source: Cellini-Ferreira-Rothstein (2010)
CFR vs. Rosen

- Rosen: Property taxes ↓ → House prices ↑↑ ⇒ Had been way over-spending
  - Looks like Tiebout failure: $1 in property taxes value near $0 (relative to $1 Tiebout benchmark)

- CFR: Property taxes ↑ → House prices ↑ ⇒ 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) $\rightarrow$ 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
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).
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

What do you think is the worst tax—that is, the least fair—federal income tax, federal Social Security tax, state income tax, state sales tax, local property tax, other/don't know?

1972
- Local property tax: 37%
- State sales tax: 16%
- Federal income tax: 23%
- State income tax: 12%
- Other/don't know: 12%

2005
- Local property tax: 39%
- State sales tax: 15%
- Federal income tax: 20%
- State income tax: 10%
- Social Security tax: 11%
- Other/don't know: 5%

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

<table>
<thead>
<tr>
<th>Unit</th>
<th>Central</th>
<th>State</th>
<th>Localities</th>
</tr>
</thead>
<tbody>
<tr>
<td>State government</td>
<td>24%</td>
<td>75%</td>
<td>1%</td>
</tr>
<tr>
<td>Local government</td>
<td>4%</td>
<td>30%</td>
<td>66%</td>
</tr>
</tbody>
</table>

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)
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 → 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
  - Paternalism: 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


- RHS vars reflect direct tax effects and endogenous responses

- 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)
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-amenity 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?
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

Motivation: Geographically concentrated poverty/race

The same data, aggregated by community area and shown with solid colors.

Motivation: Geographically concentrated recessions

Source: Yagan (2016)
Maximum Duration of Unemployment Insurance by State

Note: Map includes regular benefits, all tiers of EUC and EB. The Virgin Islands has 73 weeks of UI and Puerto Rico has 89 weeks.
*States with fewer than 26 weeks of regular benefits have proportionally fewer weeks of federal benefits available for those who file for UI after the reduction took effect. Please see the table on page 3 for a fuller explanation of the benefits available in each state.
Source: CBPP analysis of Department of Labor Employment and Training Administration data. Data from March 18, 2012.

Source: CBPP (2012)
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)

- 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:
  \[ r_j = g_j (L_j) \]
• 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:

\[ A_1 - A_2 = (w_2 - r_2) - (w_1 - r_1) \]

- 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)

- 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)

Source: Yagan (2016), replicating Blanchard-Katz

Diagram showing the deviation from trend and aggregate (in %, pp) over years after a shock. The graph plots employment (%), population (%), and employment rate (pp) over a 10-year period. The source is mentioned in the document.
• 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")
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
• Workers now have mean-zero idiosyncratic taste for local amenities $\varepsilon_{ij}$:

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

• 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:

$$L_1 = F \left[ (w_1 - r_1 + A_1) - (w_2 - r_2 + A_2) \right]$$
• 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)
Let $\bar{\xi} \equiv F^{-1}(L^*_1)$ denote marginal mover’s relative preference for 2

Consumer surplus:

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

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

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 → rent in city 1 rises by $1 → only city 1 landlords benefit (city 2 landlords lose)

- Extreme case B: Inelastic housing supply → rent in city 1 rises by $1 → 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
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
• 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)
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

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 quantities (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

Ex. Place-based policies after economic shocks

• U.S. extends state unemployment insurance durations when state unemployment rate reaches thresholds (26 weeks → 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)
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 → rents to tax

- Suarez-Serrato-Zidar (2016): 40% borne by firm owners, 25-30% by landowners, 30-35% by workers
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)

F-test all leads are 0 has p-value = 0.92
F-test all lags are 0 has p-value = 0.036

Cumulative Effect no leads
Cumulative Effect w/ leads
Long Difference Point Estimate
95% Confidence Interval

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