Econ 2450B, Lecture 1: Basics of Welfare Estimation

Nathaniel Hendren

Harvard and NBER

Spring, 2015

Goal of Public Finance

- Study the interaction between the government and the economy
- Implicitly, always a normative component
  - Should we increase taxes?
  - Should we spend more on education, roads, etc.?
  - Should we change the mix of taxes (e.g. commodity vs. property vs. capital vs. income taxes)?
- Requires notion of "should"
  - Use economics to formalize notions of welfare

Welfare Measurement

- Measurement of welfare:
  - Consumer surplus (Marshall 1890)
  - Compensating variation and Equivalent variation (Hicks 1939, 1940, 1941; Kaldor 1939)
  - Takeaway of this literature
    - Income affects make definitions difficult
    - Interpersonal comparisons are very difficult (Boadway 1974)
    - But, envelope theorem is useful for analyzing small welfare changes

Welfare Impact of Policy Changes

- Common approach in PF: consider small policy changes
- How do we think about the welfare impacts of small policy changes?
  - Marginal excess burden (Harberger 1964, Feldstein 1999, Kleven and Kreiner 2005)
  - Marginal willingness to pay (Mayshar 1990, Slemrod and Yitzhaki 1996, 2001, Kleven and Kreiner 2006, Hendren 2014)
- Set up a general economic model to define these (Following Hendren 2014)

Setup

- Individual i chooses:
- Goods x_i and labor supply l_i
  - Could be vectors of goods/labor supply activities
- Government chooses:
  - Publicly provided goods and services, G, at marginal cost c
  - Taxes on goods and labor supply: τ_i^x and τ_i^l
  - Transfers T_i
  - Non-linear taxes?

Utility function

- Individuals have utility function

$$u_i(x, l, G)$$

  - Rules out externalities
  - Production: Goods are produced linearly with one unit of labor supply

$$(1+\tau_i^x)\, x_i \leq \left(1-\tau_i^l\right) l_i + T_i$$

  - Rules out:
    - Spillovers
    - GE effects (see pset)

Maximization

- Indirect utility function

$$V_i\left(\tau_i^x, \tau_i^l, G\right) = max\ u_i\left(x, l, G\right)$$

s.t.

$$\left(1+\tau_i^x\right) x_i \leq \left(1-\tau_i^l\right) l_i + T_i$$

- Lagrange multiplier λ_i is marginal utility of income

Social Welfare

- Social welfare function

$$W\left(\left\{\tau_i^x, \tau_i^l, G_i\right\}_i\right) = \sum_i \psi_i V_i\left(\tau_i^x, \tau_i^l, G_i\right)$$

- Bergson (1938)-Samuelson (1947)
- Does ψ depend on things other than utility? (Saez and Stantcheva 2013)
- Does it matter that we assume weights are linear?
  - Local changes...

Policy Changes

- Define a "Policy Path" to trace out changes to government policy, $P(\theta)$:
- For any $\theta \in (-\epsilon, \epsilon)$

$$P(\theta) = \left\{ \left\{ \hat{\tau}^{l}_{ij}(\theta) \right\}_j, \left\{ \hat{\tau}^{x}_{ij}(\theta) \right\}_j, \hat{T}_i(\theta), \widehat{\mathbf{G}}_\mathbf{i}(\theta) \right\}_{i,}$$

  - Two assumptions (Draw Picture):
    1. $\theta = 0$ is status quo:
       $$\left\{ \left\{ \hat{\tau}^{l}_{ij}(0) \right\}, \left\{ \hat{\tau}^{x}_{ij}(0) \right\}, \hat{T}_i(0), \widehat{\mathbf{G}}_\mathbf{i}(0) \right\}_{i,} = \left\{ \left\{ \tau^{l}_{ij} \right\}, \left\{ \tau^{x}_{ij} \right\}, T_i, \mathbf{G}_\mathbf{i} \right\}_i$$
    2. $P(\theta)$ is continuously differentiable in θ
       - $\frac{d\hat{\tau}^{x}_{ij}}{d\theta}$, $\frac{d\hat{\tau}^{l}_{ij}}{d\theta}$, $\frac{d\hat{T}_i}{d\theta}$, and $\frac{d\hat{G}_{ij}}{d\theta}$ exist and are continuous in θ
- Should the government follow the policy path and increase θ?
  - Need to measure how welfare changes with θ
  - First, start with the positive questions...

Positive Analysis: Agent's Behavior and Government Budget

- Agents optimally choose $\mathbf{x}_i$ and $\mathbf{l}_i$ facing policy $P(\theta)$
  - $\widehat{\mathbf{x}}_i(\theta) = \{\hat{x}_{ij}(\theta)\}_j$ and $\widehat{\mathbf{l}}_i(\theta) = \{\hat{l}_{ij}(\theta)\}_j$
    - These are "potential outcomes" in world $P(\theta)$
- Net government resources towards individual i,

$$\hat{t}_i(\theta) = \sum_{j=1}^{J_G} c_j^G \hat{G}_{ij}(\theta) + \hat{T}_i(\theta) - \sum_{j=1}^{J_X} \hat{\tau}_{ij}^x(\theta)\hat{x}_{ij}(\theta) - \sum_{j=1}^{J_L} \hat{\tau}_{ij}^l(\theta)\hat{l}_{ij}(\theta)$$

- Budget neutrality would be $\sum_i \frac{d\hat{t}_i}{d\theta} = 0 \quad \forall \theta$
  - $\frac{d\hat{t}_i}{d\theta}$ captures distributional impact
- Behavioral response affects budget

$$\frac{d}{d\theta}\left(\sum_{j=1}^{J_X} \hat{\tau}_{ij}^x(\theta)\hat{x}_{ij}(\theta) + \sum_{j=1}^{J_L} \hat{\tau}_{ij}^l(\theta)\hat{l}_{ij}(\theta)\right) = \underbrace{\left(\sum_j^{J_X} \frac{d\hat{\tau}_{ij}^x}{d\theta} x_{ij} + \sum_j^{J_L} \frac{d\hat{\tau}_{ij}^l}{d\theta} l_{ij}\right)}_{\substack{\text{Mechanical Impact} \\ \text{on Govt Revenue}}} + \underbrace{\left(\sum_j^{J_X} \tau_{ij}^x \frac{d\hat{x}_{ij}}{d\theta} + \sum_j^{J_L} \tau_{ij}^l \frac{d\hat{l}_{ij}}{d\theta}\right)}_{\substack{\text{Behavioral Impact} \\ \text{on Govt Revenue}}}$$

Normative Analysis: Marginal Willingness to Pay for Policy

- Normative questions:
  - WTP: How much are people willing to pay to move along the policy path?
  - MEB: How much additional revenue could the government get if the policy change is implemented but utility is held constant using individual specific lump-sum transfers
- Person i's marginal willingness to pay to move along the policy path

$$\frac{\frac{d\hat{V}_i}{d\theta}\big|_{\theta=0}}{\lambda_i}$$

  - Money metric utility measure
  - Equivalent to marginal EV and marginal CV
    - Why?
  - Will define MEB later
    - Why is this different from EV/CV? (think about whose dollars we are measuring)

Characterization of Marginal Willingness to Pay for Policy

- The envelope theorem (Draw Picture) implies:

$$\frac{\frac{d\hat{V}_i}{d\theta}|_{\theta=0}}{\lambda_i} = \sum_{j=1}^{J_G} \frac{\frac{\partial u_i}{\partial G_{ij}}}{\lambda_i} \frac{d\hat{G}_{ij}}{d\theta} + \frac{dT_i}{d\theta} - \sum_{j}^{J_X} \frac{d\hat{\tau}_{ij}^{x}}{d\theta} x_{ij} - \sum_{j}^{J_L} \frac{d\hat{\tau}_{ij}^{l}}{d\theta} l_{ij}$$

- Now, substitute:

$$\frac{d\hat{T}_i}{d\theta} = \frac{d\hat{t}_i}{d\theta} - \sum_{j=1}^{J_G} c_j^G \frac{d\hat{G}_{ij}}{d\theta} + \frac{d}{d\theta}\left(\sum_{j=1}^{J_X} \hat{\tau}_{ij}^{x}(\theta)\,\hat{x}_{ij}(\theta) + \sum_{j=1}^{J_L} \hat{\tau}_{ij}^{l}(\theta)\,\hat{l}_{ij}(\theta)\right)$$

Characterization of MWTP

- Behavioral responses matter in keeping track of net resources

$$\frac{\frac{d\hat{V}_i}{d\theta}\big|_{\theta=0}}{\lambda_i} = \underbrace{\frac{d\hat{t}_i}{d\theta}}_{\text{Net Resources}} + \underbrace{\sum_{j=1}^{J_G}\left(\frac{\frac{\partial u_i}{\partial G_{ij}}}{\lambda_i} - c_j^G\right)\frac{d\hat{G}_{ij}}{d\theta}}_{\substack{\text{Public Spending/} \\ \text{Mkt Failure}}} + \underbrace{\left(\sum_j^{J_X}\tau_{ij}^x\frac{d\hat{x}_{ij}}{d\theta} + \sum_j^{J_L}\tau_{ij}^l\frac{d\hat{l}_{ij}}{d\theta}\right)}_{\substack{\text{Behavioral Impact} \\ \text{on Govt Revenue}}}$$

where the RHS is evaluated at $\theta = 0$.

  - Behavioral responses matter to the extent to which individuals impose resource costs for which they don't pay

- If government taxation is only wedge between social and private costs, a single causal effect is sufficient
  - Impact on government revenue is sufficient for all behavioral responses

Marginal Excess Burden (MEB)

- Can define MEB/MDWL in this framework
  - Let **v** denote a vector of pre-specified utilities (e.g. status quo <–> "equivalent variation" MEB in Auerbach and Hines 2002)
    - Define an augmented policy path:

$$P^v = \left\{ \left\{ \hat{\tau}_{ij}^l(\theta) \right\}_j, \left\{ \hat{\tau}_{ij}^x(\theta) \right\}_j, \hat{T}_i(\theta) + \hat{C}_i(\theta; v), \widehat{\mathbf{G}}_\mathbf{i}(\theta) \right\}_i$$

      where $\hat{C}_i(\theta; v)$ holds utilities constant at **v**.
  - MEB is defined as

$$MEB_i^{v_i} = \frac{d\hat{t}_i^v}{d\theta}|_{\theta=0}$$

    - Measures additional revenue government could obtain if it implements the policy but then holds people's utility constant using individual-specific lump-sum transfers
    - Depends on compensated elasticities (by definition)
    - Conceptually deals with the budget constraint
    - But, MWTP is generally positive for budget-negative policies!

Motivating a Particular MVPF Measure

- Many real-world policies are not budget neutral
  - Common to "adjust for the MCPF"
- There are a lot of different definitions (see Ballard and Fullerton, 1992; Dahlby, 2008)
- One definition is particularly useful: no need to decompose any causal effects into income and substitution effects

Defining the MVPF

- Suppose $P_1(\theta)$ and $P_2(\theta)$ are two non-budget neutral policies
  - Marginal cost to govt of $\int_i \frac{d\hat{t}_i^{P_1}}{d\theta} di$ and $\int_i \frac{d\hat{t}_i^{P_2}}{d\theta} di$
  - Marginal social welfare of $\int_i \eta_i \frac{\frac{d\hat{V}_i^{P_1}}{d\theta}|_{\theta=0}}{\lambda_i} di$ and $\int_i \eta_i \frac{\frac{d\hat{V}_i^{P_2}}{d\theta}|_{\theta=0}}{\lambda_i} di$
- Define MVPF as in Mayshar (1990), Dahlby (1998), Slemrod and Yitzhaki (1996, 2001), Kleven and Kreiner (2006)
- Benefit-cost ratio for each policy

$$MVPF_P^{\hat{i}} = \frac{\int_i \frac{\eta_i}{\eta_{\hat{i}}} \frac{\frac{d\hat{V}_i^P}{d\theta}|_{\theta=0}}{\lambda_i} di}{\int_i \frac{d\hat{t}_i^P}{d\theta} di} = \frac{\text{"}BENEFIT\text{"}}{\text{"}COST\text{"}}$$

  - measured in units of $\hat{i}$ income

Properties

- MVPF is policy-specific – can be computed for any non-budget neutral policy
    - Does not require decompositions into income and substitution effects
- Comparisons of MVPF correspond to comparisons of social welfare
    - Welfare impact of budget-neutral policy with more P_2 and less P_1

$$\frac{\frac{dW}{d\theta}}{\eta_i} = MVPF^i_{P_2} - MVPF^i_{P_1}$$

    - Allocating money to high MVPF from low MVPF policies increases social welfare
        - But need to keep units i the same

Comparisons Using Okun's Bucket

- In general, MVPF requires weighting by social marginal utilities of income
- Formula can be simplified if η_i is constant within the set of beneficiaries
  - Beneficiaries of P_1 have equal social marginal utility of income η_1
    - $MVPF^1_{P_1}$ is marginal benefit to beneficiaries, normalized by govt cost
  - Beneficiaries of P_2 have equal social marginal utility of income η_2
    - $MVPF^2_{P_2}$ is marginal benefit to beneficiaries, normalized by govt cost
- Increasing spending on P_1 and decrease spending on P_2 increases welfare iff

$$\frac{\eta_1}{\eta_2} \geq \frac{MVPF^2_{P_2}}{MVPF^1_{P_1}}$$

  - Differs from MEB comparisons across policies which requires modified social welfare weights that include income effects (Diamond and Mirrlees 1971)

Simplified Formulas

- Simplification #1: Assume beneficiaries have same η_i
  - Compute MVPF in units of beneficiaries' income
- Simplification #2: Suppose policy either effects market or non-market transfers
  - [Market Goods/Transfers] $P(\theta)$ increases mechanical transfers/subsidies by $\$\theta$

$$MVPF = \frac{1}{\frac{1}{|I|}\int_{i\in I}\frac{d\hat{t}_i^P}{d\theta}di}$$

    - $\frac{1}{|I|}\int_{i\in I}\frac{d\hat{t}_i^P}{d\theta}di = 1 + FE$ is cost of providing \$1 mechanical income
  - [Non-Market Goods] $P(\theta)$ increases public goods/services by $\$\theta$

$$MVPF = \frac{\frac{\frac{\partial u}{\partial G}}{\lambda}}{\frac{1}{|I|}\int_{i\in I}\frac{d\hat{t}_i^P}{d\theta}di}$$

    - Multiply by WTP for G relative to income, $\frac{\frac{\partial u}{\partial G}}{\lambda}$

Applications

- Use existing causal effects to calculate MVPF for various policy changes
  - Top marginal tax rate increase
    - Many studies summarized in Saez et al (2012)
  - EITC Generosity
    - Many studies summarized in Hotz and Scholz (2003), Chetty et al (2013)
  - Food Stamps
    - Hoynes and Schanzenbach (2012)
  - Job Training
    - RCT of Job Training Partnership Act (Bloom et al 1997)
  - Section 8 Housing Vouchers
    - Lotteried access to Section 8 in Illinois (Jacob and Ludwig 2012)

Top Tax Rate Increases

- Large literature studying causal impact of top tax rate increases / decreases
  - Saez, Slemrod, and Giertz (2012) provide review
    - Many estimates of causal effect of changes to top income tax rate
    - Tax-weighted taxable income elasticity
  - Suggests 25-50% of mechanical revenue lost (lots of disagreement/uncertainty!)
    - Fiscal cost is \$0.50-\$0.75 for \$1 in transfer
  - Suggests MVPF of \$1.33-\$2

$$MVPF = \frac{1}{1-.25} = 1.33$$

EITC Expansions

- Large literature studying causal impact of EITC expansions (Hotz and Scholz 2003, Chetty et al 2013)
  - Intensive + extensive calculations suggest fiscal cost of EITC is ∼14% higher because of labor supply impacts
  - Fiscal cost is \$1.14 for \$1 in mechanical EITC benefits
  - Suggests MVPF of \$0.88

$$MVPF = \frac{1}{1+.14} = 0.88$$

Food Stamps

- Hoynes and Schanzenbach (2012) use variation across counties in introduction of food stamp program (1960-70s)
- Use data from 1968-78 PSID
- Compare labor supply over counties across time

$$\begin{aligned} y_{ict} &= \alpha + \delta FSP_{ct} + \eta_c + \lambda_t + \mu_{st} \\ &\quad + X_{it}\beta + \sigma CB60_c * t + \gamma TRANSFERS_{ct} \end{aligned}$$

- Standard DD with controls for 1960 census controls * time trends

Results

- Find large but imprecise decrease in labor earnings of $2,943 (can't reject zero)
  - Assume 20% marginal tax rate -> $588.60 impact on government budget
- Average household transfer: $1,153.25
- Total cost is $1,153.25+$588.60=$1,741.85.
- MVPF:

$$\frac{1}{\frac{1}{|I|}\int_{i\in I}\frac{d\hat{t}_i^P}{d\theta}di} = \frac{1,153.25}{1741.85} = 0.66$$

- Food stamps are "in-kind": $\frac{\frac{\partial u}{\partial G}}{\lambda} \neq 1$
  - May be that $\frac{\frac{\partial u}{\partial G}}{\lambda} < 1$ because goods are in kind
    - Smeeding (1982) estimates 0.97; Moffitt (1989) estimates ~1
    - Whitmore (2002) estimates 0.80 for marginal/distorted recipients
- Assuming food stamps valued as cash, MVPF is 0.66

Concerns with Food Stamp Results

- Causal effects very imprecise
- Also, causal effect in 1970 = causal effect now?
- How do we value non-market goods?
  - Generally requires structural modeling assumptions... (Whitmore 2002)
    - How much do people distort their consumption
    - How much does this affect their WTP (small distortion satisfies envelope theorem)

Job Training

- Job Training Partnership Act of 1982 provided job training services to low income youth and adults
- Bloom et al (1997) report results from RCT (I focus on adult women impact)
    - Increased tax collection of $236
    - Reduction in welfare benefits (AFDC) $235
        - $471 net increase in government budget from behavioral responses
    - Marginal cost of providing the training is $1,381
        - Cost net of fiscal externality is $910
    - MVPF is 1.52 if program costs are valued at its costs

Job Training

- No estimates of $\frac{\frac{\partial u}{\partial G}}{\lambda}$ for the program
  - Bloom et al (1997) implicitly assume earnings is fully valued
    - This is far too common!
    - When is this OK?
  - Earnings increase of \$1,683 for marginal cost of \$1,381 -> $\frac{\frac{\partial u}{\partial G}}{\lambda} = 1.22$
    - Suggests MVPF of 1.85 if increase was entirely productivity
  - But could be MVPF $= 0$ if no one valued it

Section 8 Housing Vouchers

- Section 8 is largest low-income housing program in US
  - Provides vouchers to low-income households (see MTO experiment, etc.)
- Jacob and Ludwig (2012) exploit excess applications in Illinois
  - Allocated via lottery
  - Estimate significant impact on labor supply and welfare take-up
    - Earnings decrease implies fiscal externality of $129 per voucher
    - Welfare programs increase sum to $432 (mostly medicaid)
    - But vouchers are a lot of money ($8,400/yr)
    - Voucher cost $1.05 for every $1 of vouchers

$$MVPF = 0.95\frac{\frac{\partial u}{\partial G}}{\lambda}$$

Section 8 Housing Vouchers

- Reeder (1985) suggests \$1 vouchers valued at $\frac{\frac{\partial u}{\partial G}}{\lambda} = 0.83$
  - People consume too much house!
- Suggests MVPF of 0.79 for housing vouchers

Summary

Policy	$\frac{\frac{\partial u}{\partial G}}{\lambda}$	$\frac{1}{\frac{1}{\lvert I \rvert}\int_i \frac{d\hat{t}_i}{d\theta} di}$	*MVPF*
Top Tax Rate	1	1.33 - 2	1.33 - 2
EITC Expansion	1	0.88	0.88
Food Stamps	0.8 - 1	0.66	0.53 - 0.66
Job Training	0 - 1.22	1.52	0 - 1.85
Housing Vouchers	0.83	0.95	0.79

- Taking $MVPF^{TopTax} = 1.33$, increasing EITC and top tax rate desireable iff

$$\frac{\eta^{Rich}}{\eta^{Poor}} \leq \frac{.88}{1.33} = 0.66$$

- \$0.66 to a poor person or \$1 to a rich person?
- Question: What about MEB comparisons?

Takeaways

- Need causal effect of policy in question
  - Is this what we used?
  - ATE/ATT/ITT?
- Also need WTP for non-market goods
  - This is the hard part!
- Aggregations across people require social welfare weights
  - How to identify?
  - Surveys (Saez Stantcheva 2013)
  - Inverse Optimum (Later...)

Optimal Policy vs. MVPF

- How does the welfare framework we have covered so far relate to optimal policy formulas?
  - We have been asking "what marginal policy changes from the status quo are worth are most valuable (highest MVPF)?"
  - Optimal policy formulas ask "In an ideal world, what would the optimal policy look like (given our SWF)?"
- MVPF is evaluated around the status quo
  - Hence, requires local causal effects defined around the status quo–observed elasticities
  - What is the marginal WTP for a policy change?
  - What is the additional deadweight loss from a policy change? (well-defined?)

Optimal Policy vs. MVPF

- Optimal policy is thought experiment: "Given our SWF, What is the optimal ___...":
  - Tax rate/subsidy
  - Public goods provision
  - "...given our constraints
- Status quo is irrelevant
  - But when behavioral responses are important, must be evaluated around the optimal policy, not the status quo
- Early optimal policy:
  - Samuelson (1954) -> public goods
  - Ramsey (1927) -> taxes

Optimal Public Goods (Samuelson 1954)

- Before turning to taxes, consider public goods
- Pure public goods:
  - Non-rival: My consumption doesn't prevent your consumption (No Congestion)
  - Non-excludable: Provider can't prevent consumption by those who don't pay
- Private Goods benefit one individual i at a time:
  - Total consumed units of X $= \sum_i x_i$
- Public Goods benefit several individuals simultaneously
  - Each person consumes G but G may be the sum of all contributions by others:

$$G = \sum_j G_j$$

    where G_j is the contribution by individual j towards the public good

    - Note this differs slightly from the notation above where G_i is the publicly provided good consumed by individual i.

Public Goods

- Why might the free market under-provide public goods?
    - Free-riding
    - Public goods create positive externalities, individuals underprovide
- Two main research questions:
    - 1. What is the optimal level of PGs?
    - 2. How are (and can) PGs provided in practice?
- Provision problems
    - Free-rider models (BBV, in section)
    - Optimal "warm glow" offsets the free-riding externality
    - Will government crowd-out private provision? Do we care?

Optimal Public Goods (Samuelson 1954)

- First Welfare Theorem: Any market equilibrium is Pareto Optimal
  - With public goods, this fails
  - Samuelson (1954) derives condition for a Pareto Optimum (may require transfers)
- Consider First Welfare Theorem setup:
  - Individuals indexed by i, two goods, X and G
  - Utility functions $U^i(x_i, G_i)$, standard budget constraint
  - c is the dollar cost of producing G. (Normalize price of x to 1 so $\frac{p_G}{p_x} = c$)
- Condition for private optimality
  - $\frac{U_G(x_i, G_i)}{U_x(x_i, G_i)} = c \iff MRS_i = MRT \ \forall i$

Optimal Public Goods: Failure of FWT

- Now, suppose G is public
  - So each person purchases G_i, but values $G = \sum_i G_i$
  - Utility is $U(x_i, G) = U(x_i, G_i + \sum_{j \neq i} G_j)$
- Condition for private optimality
  - Still $\frac{U_G(x_i,G)}{U_x(x_i,G)} = c \iff MRS_i = MRT \forall i$
- But social value of additional spending on G is $\sum \eta_i \frac{U_G(x_i,G)}{U_x(x_i,G)}$
- Private choice of G ignores external benefit!
  - Private equilibrium is Pareto Inferior: there is an allocation that can make everyone better

Optimal Public Goods

- Now assume G is only provided by the government
  - Worry about crowd out?
- Conder a policy path θ that just provides a Public Good funded by Lump-Sum Transfers (like Sameulson's planner would)
  - $\frac{dG}{d\theta} = 1$, at marginal cost $c = 1$
  - $\frac{dt_i}{d\theta} = 0$ (budget neutral)
  - Each person pays $\frac{dT_i}{d\theta}$
- What is the condition under which we should have more G? Optimality Condition:

$$\frac{dW}{d\theta} = 0$$

Optimal Public Goods (Samuelson 1954)

- Case 1: $\eta_i = \eta_j = \eta$. Then,

$$\frac{dW}{d\theta} = \sum_i \left(\frac{\frac{\partial u_i}{\partial G}}{\lambda_i} - \frac{dT_i}{d\theta} \right)$$

which equals zero if and only if the sum of MWTP for G out of own income equals the marginal cost ($1):

$$\sum_i \frac{\frac{\partial u_i}{\partial G}}{\lambda_i} = \sum_i \frac{dT_i}{d\theta} = 1$$

- Is this justified by the Pareto principle?

Optimal Public Goods (Samuelson 1954)

- Case 2: Pareto implementation. Use Lindahl pricing: $\frac{dT_i}{d\theta} = -\frac{\frac{\partial u_i}{\partial G}}{\lambda_i}$.
  - Each person pays their WTP. Then,

$$\frac{\frac{d\hat{V}_i}{d\theta}}{\lambda_i} = \frac{\frac{\partial u_i}{\partial G}}{\lambda_i} + \frac{dT_i}{d\theta} = 0$$

which is budget feasible if and only if the taxes collected are greater than the cost:

$$\sum_i \frac{\frac{\partial u_i}{\partial G}}{\lambda_i} = \sum_i \frac{dT_i}{d\theta} \geq 1$$

- Problems with Lindahl price? Need each person to pay their (privately known) WTP.

Stiglitz-Dasgupta-Atkinson-Stern MCPF

- Suppose we need to finance with linear tax on labor.
- Consider representative agent model
  - NOTE: As we will see, it's a bad idea to assume away lump-sum taxes in representative agent models...this makes the results largely useless and even mis-leading.
- Tax on labor, τ
- We have $\frac{dV}{dG} = 0$ if and only if:

$$\frac{\frac{\partial u}{\partial G}}{\lambda} = 1 + \tau \left(-\frac{dl}{dG} \right)$$

The fiscal externality, $FE = \tau \left(-\frac{dl}{dG} \right)$ is known as the "Marginal Cost of Public Funds" by Atkinson and Stern (1974) and Stiglitz and Dasgupta (1971).

Stiglitz-Dasgupta-Atkinson-Stern MCPF

- MCPF$=\tau\left(-\frac{dI}{dG}\right)$
    - Intuition: the fiscal externality from a budget neutral policy that increases spending on G by \$1 financed by an increase in τ.
- Two components:

$$\frac{dI}{d\theta}=\frac{dI}{d\tau}\frac{d\tau}{d\theta}+\frac{dI}{dG}\frac{dG}{d\theta}$$

- Here, the MCPF is defined as a component of the welfare impact of a budget-neutral policy
    - Differs from "MVPF" in Hendren (2013) or "MCPF" in Kleven and Kreiner (2006), who define MCPF/MVPF as the welfare impact per dollar spent on a non-budget neutral policy.

MVPF, MEB, MCPF...Yikes!

- A lot of different definitions of welfare measures / costs of taxation / etc:
- MVPF: Marginal welfare impact per dollar of government spending
  - Defined for non-budget neutral policies
  - Can be compared across policies to form hypothetical budget neutral policies ($MVPF^{P_1} - MVPF^{P_2}$ is the welfare impact of a budget neutral policy that spends money on P_1 financed with money from P_2).
  - Depends on causal effects of policy on government budget (to accurately measure costs)
- MEB: Marginal additional revenue the govt can collect if it implements the policy but holds all utilities constant using individual-specific lump-sum taxation
  - Depends on compensated elasticities
  - Can't be aggregated to social weflare without adding back in the income effects (Diamond and Mirrlees 1971).

MVPF, MEB, MCPF...Yikes!

- MCPF of Atkinson-Stern-Stiglitz-Dasgupta
  - Atkinson and Stern (1974)
  - Dasgupta and Stiglitz (1971)
  - Fiscal externality component of broader budget-neutral policy that simultaneously increases G financed by τ.
- MCPF of Harberger, Pigou, Browning
  - Defines MCPF as fiscal externality of MEB experiment of spending more on G financed by τ but holding utilities constant using individual-specific lump-sum taxation.
    - Depends on compensated response.
    - Implicitly compares to world with lump-sum taxation

Nathan's Commentary...

- This literature is fairly complicated – terms are often not used in a consistent manner
- My suggestion (others may disagree):
  - Just construct the benefit-cost ratio (MVPF) and compare across policies.
  - It's simpler, allows for easier conceptual comparisons across policies, and aggregates across people using social marginal utilites of income (Okun's bucket)
- If you have a non-budget neutral policy and you've constructed people's MWTP for that policy, you don't need to "adjust for the marginal cost of public funds"; rather, you need to think of your estimate as a marginal value of public funds that can be compared to other policies.

Pigouvian Externalities

- Now suppose choosing x causes an externality, $E(\bar{x})$, where $\bar{x}$ is the aggregate choice of x in the population
- To make things simple, retain the representative agent framework but assume choosing x doesn't incorporate the effect on $\bar{x}$ and thus the externality.
- Utility

$$u(x, l, G, E(x))$$

- MWTP:

$$\frac{\frac{d\hat{V}}{d\theta}}{\lambda} = \frac{d\hat{t}}{d\theta} + \left(\frac{\frac{\partial u}{\partial G}}{\lambda} - c\right)\frac{dG}{d\theta} + \tau^x \frac{d\hat{x}}{d\theta} + \tau^l \frac{d\hat{l}}{d\theta} + \frac{dE}{d\theta}\frac{\frac{\partial u}{\partial E}}{\lambda}$$

where $\frac{\frac{\partial u}{\partial E}}{\lambda}$ is the MWTP for E and $\frac{dE}{d\theta}$ is the causal (not compensated) impact on E.

Pigouvian Tax

- Assume budget neutrality, no public goods, and no tax on labor (does this matter?)
- Note that

$$\frac{dE}{d\theta} = \frac{dx}{d\theta}\frac{dE}{dx}$$

so

$$\frac{\frac{d\hat{V}}{d\theta}}{\lambda} = \left(\tau^x + \frac{dE}{dx}\frac{\frac{\partial u}{\partial E}}{\lambda}\right)\frac{d\hat{x}}{d\theta}$$

- Pigouvian tax:

$$\tau^{PIGOU} = -\frac{dE}{dx}\frac{\frac{\partial u}{\partial E}}{\lambda}$$

MCPF and Double Dividend

- With public goods

$$\frac{\frac{d\hat{V}}{d\theta}}{\lambda} = \left(\frac{\frac{\partial u}{\partial G}}{\lambda} - c \right) \frac{dG}{d\theta} + \left(\tau^x + \frac{dE}{dx} \frac{\frac{\partial u}{\partial E}}{\lambda} \right) \frac{d\hat{x}}{d\theta}$$

- Aside: Stiglitz-Dasgupta-Atkinson-Stern definition of MCPF is just the FE
- Double dividend: taxing x yields a 'cheaper' MCPF because it also deals with externality
- Your exercise: Show this is true iff $\tau^x < \tau^{PIGOU}$. What if $\tau^x > \tau^{PIGOU}$

Optimal Taxation in Ramsey (1927)

- Ramsey (1927): How should commodities be taxed to raise revenue, $R > 0$.
  - Modeled by Diamond and Mirrlees (1971)
- Key result: Tax-weighted Hicksian price derivatives are equated across goods
  - "Inverse elasticity rule": tax goods with smaller compensated behavioral responses

Setup

- Representative Agent (drop i subscripts).
- Commodities, x_k, indexed by k
- Government imposes taxes on commodities, τ_k.
- Necessary condition for optimality

$$\frac{d\hat{V}_P}{d\theta}|_{\theta=0} = 0$$

  for all feasible policy paths P.
- Optimal tax would be lump-sum of size R
  - Assumed to not exist

Commodity Tax Variation

- Consider policy $P(\theta)$ that changes commodity taxes (e.g. lowers tax on good 1 and raises tax on good 2)
- Budget neutral: $\frac{d\hat{t}}{d\theta} = 0$
- No change in public goods
- So, optimality condition only involves behavioral response:

$$\sum_k \hat{\tau}_k \frac{d\hat{x}_k}{d\theta}|_{\theta=0} = 0$$

Hicksian Elasticity

- Diamond and Mirrlees (1971): At the optimum, expand the behavioral response using the Hicksian demands, x_k^h,

$$\frac{dx_k}{d\theta} = \frac{\partial x_k^h}{\partial \tau_1}\frac{d\tau_1}{d\theta} + \frac{\partial x_k^h}{\partial \tau_2}\frac{d\tau_2}{d\theta}$$

  - Additional term, $\frac{\partial x_k^h}{\partial u}\frac{dV_p}{d\theta}$, but this vanishes at the optimum.
- Optimality condition is given by

$$\sum_k \tau_k \frac{\partial x_k^h}{\partial \tau_1}\frac{d\tau_1}{d\theta} = \sum_k \tau_k \frac{\partial x_k^h}{\partial \tau_2}\left(-\frac{d\tau_2}{d\theta}\right)$$

  - Tax-weighted Hicksian responses are equated across the tax rates
    - Inverse elasticity rule
- What are the needed elasticities?

Inverse Elasticity Rule

- Assume cross elasticities are zero:

$$BC = x_1 \frac{d\tau_1}{d\theta} + \tau_1 \frac{dx_1}{d\theta} + x_2 \frac{d\tau_2}{d\theta} + \tau_2 \frac{dx_2}{d\theta} = 0$$

so

$$x_1 \left(1 + \frac{\tau_1}{x_1}\frac{\partial x_1^h}{\partial \tau_1}\right)\frac{d\tau_1}{d\theta} = x_2 \left(1 + \frac{\tau_2}{x_2}\frac{\partial x_2^h}{\partial \tau_2}\right)\left(-\frac{d\tau_2}{d\theta}\right)$$

- And optimality implies

$$x_1 \left(\frac{\tau_1}{x_1}\frac{\partial x_1^h}{\partial \tau_1}\right)\frac{d\tau_1}{d\theta} = x_2 \left(\frac{\tau_2}{x_2}\frac{\partial x_2^h}{\partial \tau_2}\right)\left(-\frac{d\tau_2}{d\theta}\right)$$

Inverse Elasticity Rule

- So
$$\left(\frac{\tau_1}{x_1}\frac{\partial x_1^h}{\partial \tau_1}\right) = \left(\frac{\tau_2}{x_2}\frac{\partial x_2^h}{\partial \tau_2}\right) = \kappa$$
- Translating to price (1+tau) instead of tax (tau) elasticities:
$$\frac{\tau_j}{1+\tau_j}\epsilon^h_{j,(1+\tau_j)} = \kappa$$
Or
$$\frac{\tau_j}{1+\tau_j} = \frac{\kappa}{\epsilon^h_{j,(1+\tau_j)}}$$
which is the "inverse elasticity rule".

Key Result: Inverse Elasticity Rule

- Main result of Ramsey model: Inverse elasticity rule
- Key Assumptions:
  - Representative agent
  - No lump sum taxation

Optimal Taxation of Production

- Diamond and Mirrlees (1971) also consider the issue of production efficiency.
- Commodities, x_k, indexed by k, transformed into one another (produced) by firms and government
- Producer prices p_k, Consumer prices q_k
  - Tax is wedge $\tau_k = q_k - p_k$
- Consumer i solves $\max u_i(\mathbf{x})$ s.t. $\sum q_k x_k \leq 0$
  - Defines consumer (final) demand for each commodity $x_k^i(\mathbf{q})$
  - and indirect utility $V_i(\mathbf{q}) = u(\mathbf{x}^i(\mathbf{q}))$
- Note: Consumers are the ones endowed with the initial commodity supply
- Endowments allow them to exchange, consumers are on budget constraint

Firm side

- Price-taking firms j transform commodities
- Production possibilites represented by input output function $f^j(\mathbf{y}) = 0$
  - for example, $y_1 = y_2^{.3} * y_3^{.7} \iff y_1 - (-y_2^{.3}) * (-y_3^{.7}) = 0$
  - Can turn $y_2$and y_3 into y_1 (or vice versa, depending of domain)
  - Negative arguments are inputs, positives are outputs

Firm side: CRS Production

- Assumption: constant returns to scale
- Then each firm can produce "as much" or "as little" as desired in fixed proportions
  - Together, many CRS firms define an aggregate production function $f(\mathbf{y}) = 0$
  - No profits for any firm (otherwise infinite production) in equilibrium
  - $\mathbf{p} \cdot \mathbf{y}^{\mathbf{j}} = 0$ must hold in equilibrium, and thus $\mathbf{p} \cdot \mathbf{y} = \mathbf{p} \cdot \left(\sum \mathbf{y}_{\mathbf{j}}\right) = 0$
- Under CRS, behavior of many optimizing firms same as one aggregate firm

Firm side: Firm Objective

- Objective: Choose point on frontier to maximize output prices - input prices
- $$\max \mathbf{p} \cdot \mathbf{y} \text{ s.t. } f(y) = 0$$
- Optimality condition: $\frac{\partial f}{\partial y_k} = p_k \iff MRT = \frac{\frac{\partial f}{\partial y_k}}{\frac{\partial f}{\partial y_{k'}}} = \frac{p_k}{p'_k}$
    - Why can we ignore lagrange multiplier on $f(y) = 0$ condition? Because we can normalize the units of f to be in terms of one of the commodities...see Diamond-Mirrlees (1971).

Govt

- D&M think of Gov't as a planner with a distributive objective but:
  - Can't just pick point on PPF
  - Must deal with consumers through market place using uniform prices
  - Uses:
    - a.) linear commodity taxes to set prices and
    - b.) public production to adjust quantities above and beyond what private sector does given prices
- Public production follows PPF given by $g(z) \leq 0$

Objective

- What is the objective here?
  - redistribution–different than Ramsey, since no revenue requirement
- Why would commodity taxes help with no lump sum transfers?
  - differential wealth levels are due to endowment differences
  - Commodity taxes target:
    - Different tastes
    - Value of endowment
  - But commodity taxes cause DWL

Objective

- Solve

$$\max_{q,p,z} \sum_i W(V_i(\mathbf{q})) \text{ s.t. } \sum_i x_k^i(\mathbf{q}) = y_k(\mathbf{p}) + z_k \text{ , } f(\mathbf{y}) = \mathbf{0}, \text{ and } g(\mathbf{z}) = 0$$

- Lagrangian

$$\max_{q,p,z} \sum_i W(V_i(\mathbf{q})) + \sum_k \lambda_k (y_k(\mathbf{p}) + z_k - \sum_i x_k^i(\mathbf{q})) + \gamma^f f(\mathbf{y}(\mathbf{p})) + \gamma^g g(\mathbf{z})$$

Objective

- Production-side and consumer-side variables are additively separable

$$\max_{q,p,z} \underbrace{\sum_i W(V_i(\mathbf{q})) - \sum_k \lambda_k \sum_i x_k^i(\mathbf{q})}_{consumption} + \underbrace{\sum_k \lambda_k (y_k(\mathbf{p}) + z_k) + \gamma^f f(\mathbf{y}(\mathbf{p})) + \gamma^g g(\mathbf{z})}_{production}$$

- Note that FOC for producer prices and government production depend on W only through the shadow value of an endowment unit of k.
- Also, choice of p *directly* implements y, so we can choose y directly

$$\max_{q,y,z} \underbrace{\sum_i W(V_i(\mathbf{q})) - \sum_k \lambda_k \sum_i x_k^i(\mathbf{q})}_{consumption} + \underbrace{\sum_k \lambda_k (y_k + z_k) + \gamma^f f(\mathbf{y}) + \gamma^g g(\mathbf{z})}_{production}$$

Result

- [FOC y_k]$\lambda_k = \gamma^f \frac{\partial f}{\partial y_k}$
- [FOC g_k]$\lambda_k = \gamma^g \frac{\partial g}{\partial z_k}$
- Taking ratio, for any social welfare objective, it must be the case that:

$$\frac{\frac{\partial g}{\partial z_k}}{\frac{\partial g}{\partial z_{k'}}} = \frac{\frac{\partial f}{\partial y_k}}{\frac{\partial f}{\partial y_{k'}}} = \frac{p_k}{p_{k'}}$$

- The government's decision to intervene in the economy is independent of the objective. MRTs are always equalized, and the only wedge is between consumer and producer prices. Production-side and consumer-side variables are additively separable

Mirrlees Model

- Mirrlees (1971)
  - Formalizes tradeoff between equity and efficiency as an information problem for the government
    - Equity and the size of the pie are necessarily related...
- Saez (2001) shows Mirrlees model can be interpreted through empirical elasticities
  - Spawned huge empirical literature on optimal taxation
- Key difference: Redistribution as rationale for taxing income instead of lump-sum taxation
  - Would like individual-specific lump-sum taxes (efficient!)
  - But, hard to make individual-specific taxes (information constraints)

Setup

- Individuals choose effort/income to maximize utility
- Government doesn't observe your latent productivity
- Can only tax you based on observed income, not latent productivity / effort
- Taxing income -> affects effort
  - Efficiency / equity tradeoff

Setup

- Individuals consume c and earn y
- Individuals are heterogeneous and have utility $u(c, y; \theta)$
  - e.g. θ indexes cost of effort
  - Popular specifications:
    - $u(c, y; \theta) = u\left(c, \frac{y}{\theta}\right) = c - v\left(\frac{y}{\theta}\right)$ where $\frac{y}{\theta}$ is "effective effort"
    - High θ implies higher productive (can earn y with lower utility cost).

Government's Problem

- Government maximizes Bergson-Samuelson SWF:

$$\max \int v(\theta)\,\psi(\theta)\,d\mu(\theta)$$

  where $v(\theta)$ is individual θ's utility and $\psi(\theta)$ are a set of welfare weights.
- If government could observe θ, second welfare theorem applies: Tax people based on θ
  - Individual-specific lump sum taxes for each θ
- Key insight of Mirrlees: government can't observe θ
  - Can observe income, y
  - Can choose tax function $T(y)$ so that individuals have budget constraint
- Individuals earn y are taxed $T(y)$:

$$c \leq y - T(y)$$

- Partial equilibrium assumption: choice of y by type θ doesn't affect earnings capabilities of type θ'

Individual's Problem

- Individuals choose y subject to tax schedule $T(y)$
- Substituting $c = y - T(y)$, individuals choose earnings to maximize utility:

$$\max_{y} u(y - T(y), y; \theta)$$

- If u is concave, yields FOC:

$$u_c\left(1 - T'(y(\theta))\right) = u_y$$

or

$$\frac{u_y}{u_c} = 1 - T'(y(\theta))$$

  - MRS equated to wages
  - Taxes distort earningsy decisions
  - Leads to choice $y(\theta)$ and utility $v(\theta)$

$$v(\theta) = u(y(\theta) - T(y(\theta)), y(\theta); \theta)$$

Government's Constrained Problem

- Government chooses function $T(y)$ subject to the constraint that individuals choose $y(\theta)$ when facing $T(y)$

$$\max \int v(\theta)\,\psi(\theta)\,d\mu(\theta)$$

s.t.

$$\begin{aligned} v(\theta) &= max_y \; u(y - T(y), y; \theta) \quad [IC] \\ y(\theta) &= argmax \; u(y - T(y), y; \theta) \end{aligned}$$

and

$$\int T(y(\theta))\,d\mu(\theta) \leq 0 \quad [RC]$$

- Note we assume the government does not face a participation constraint

Solution: Two Approaches

- General solution is difficult (Mirrlees 1971)
- Two general approaches: Calculus of Variations vs. Hamiltonian
  - Hamiltonian works well if θ has a nice structure (e.g. uni-dimensional)
  - Calculus of variations useful for characterizing necessary conditions for optimum, but sufficiency is difficult
    - More closely aligned to empirical quantities
    - Loosely, calculate MVPF for variations in tax schedule and set welfare impact equal to zero
- Saez (2001) takes an empirical approach varying $T'(y)$ function
  - Saez (2001) provides very nice formulas
  - Start with a simpler version: optimal top tax rate

Calculus of Variations Approach: Top Tax Rate Changes

- Suppose τ is tax rate on income above $\bar{y}$. What is the optimal τ? (Saez 2001)
  - Simpler question: what is the revenue maximizing tax rate?
    - Assume no social value of income on the rich
- Total revenue from tax on incomes above $\bar{y}$:

$$R = \int_{y \geq \bar{y}} \tau \left(y - \bar{y}\right) f\left(y\right) dy$$

  where y (and it's pdf, $f\left(y\right)$) is an endogenous response to the tax rate, τ
- Marginal revenue from increasing τ:

$$\frac{dR}{d\tau} = \underbrace{\left(E\left[y|y \geq \bar{y}\right] - \bar{y}\right)\left(1 - F\left(\bar{y}\right)\right)}_{\text{Mechanical}} + \tau \underbrace{\frac{d}{d\tau}\int_{y \geq \bar{y}} \left(y - \bar{y}\right) f\left(y\right) dy}_{\text{Behavioral}}$$

  where $F\left(y\right)$ is the c.d.f. of the income distribution
- The revenue-maximizing tax rate is:

Behavioral Response

- Under some assumptions can write the response using "structural" elasticities:

$$\frac{d}{d\tau}\int_{y\geq\bar{y}}(y-\bar{y})\,f(y)\,dy=\int_{y\geq\bar{y}}\frac{dy}{d\tau}f(y)\,dy$$

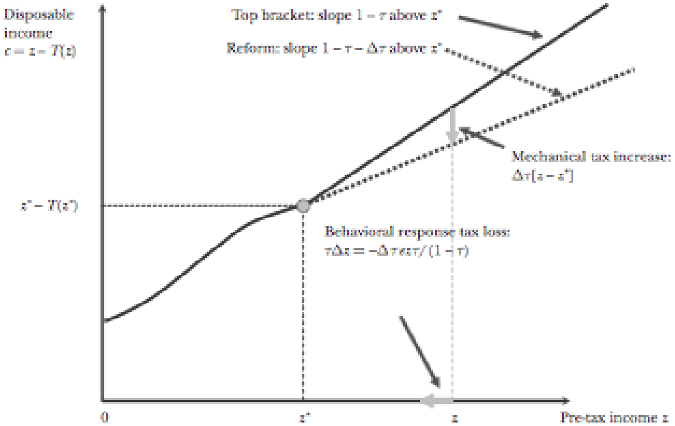
  - e.g. Participation margin? Discrete shifts in labor supply?
- What is $\frac{dy}{d\tau}$?
  - Use utility theory (Saez 2001; Diamond and Saez 2011)
  - Uncompensated Elasticity

$$\zeta^{u}=\frac{1-\tau}{y}\frac{\partial y}{\partial(1-\tau)}$$

  - Assume no income effects (see Saez 2001 for broader derivation)

Graphically (Diamond and Saez 2011)

Figure 1
Optimal Top Tax Rate Derivation



Behavioral Response

Then,

$$\begin{aligned}\frac{dy}{d\tau} &= -\frac{\partial y}{\partial (1-\tau)} d\tau \\ \frac{dy}{d\tau} &= -\zeta^u \frac{y}{1-\tau}\end{aligned}$$

So

$$\begin{aligned}-\int_{y \geq \bar{y}} \frac{dy}{d\tau} f(y)\, dy &= \int_{y \geq \bar{y}} \zeta^u \frac{y}{1-\tau} f(y)\, dy \\ &= \frac{\zeta^u}{1-\tau} \int_{y \geq \bar{y}} y f(y)\, dy \\ &= \frac{\zeta^u}{1-\tau} (1 - F(\bar{y})) E[y | y \geq \bar{y}]\end{aligned}$$

where ζ^u is the income-weighted elasticity and $y^m = E[y | y \geq \bar{y}]$

- Note the elasticities are defined around the optimum

Top Tax Formula

- Top tax rate satisfies:

$$\tau = \frac{1-\tau}{\zeta^u} \frac{E\left[y|y \geq \bar{y}\right] - \bar{y}}{E\left[y|y \geq \bar{y}\right]}$$

or

$$\tau = \frac{1}{1 + a\zeta^u}$$

where

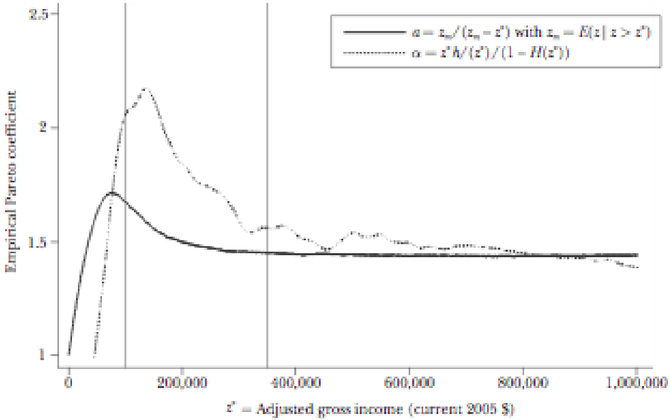
$$a = \frac{E\left[y|y \geq \bar{y}\right] - \bar{y}}{\bar{y}}$$

is the pareto parameter

Pareto Tails

Figure 2

Empirical Pareto Coefficients in the United States, 2005



Top Tax Rate

If $\zeta^u = 0.3$ and $a = 1.5$, then top tax rate is

$$\tau = \frac{1}{1 + 1.5 * 0.3} = 0.69$$

If $\zeta^u = 0.5$, then

$$\tau = 57\%$$

If $\zeta^u = 0.1$, then

$$\tau = 87\%$$

What is the size of the behavioral response?

Top Taxable Income Elasticity

Much literature on this...See Saez, Slemrod, and Giertz (2012 JEL review)

- DD with tax reforms reforms:
  - Reagan tax cuts
    - Feldstein (1995)
  - Clinton tax increases (OBRA 93)
    - Goolsbee 2000
    - Giertz 2009
- Pooled variation
  - Gruber and Saez 2002 (JpubEc)
- Kink analysis
  - Saez 2002
- Difficulty: Top incomes are NOISY...precision is difficult.

Classic Study: Feldstein 1995 JPE

TABLE 2

Estimated Elasticities of Taxable Income with Respect to Net-of-Tax Rates

Taxpayer Groups Classified by 1985 Marginal Rate	Net of Tax Rate (1)	Adjusted Taxable Income (2)	Adjusted Taxable Income Plus Gross Loss (3)
	Percentage Changes, 1985–88		
1. Medium (22–38)	12.2	6.2	6.4
2. High (42–45)	25.6	21.0	20.3
3. Highest (49–50)	42.2	71.6	44.8
	Differences of Differences		
4. High minus medium	13.4	14.8	13.9
5. Highest minus high	16.6	50.6	24.5
6. Highest minus medium	30.0	65.4	38.4
	Implied Elasticity Estimates		
7. High minus medium		1.10	1.04
8. Highest minus high		3.05	1.48
9. Highest minus medium		2.14	1.25

Shortcomings

- Feldstein 1995 is a seminal article, highlighting that people may respond to taxation and this affects how we should think about tax design
- But, there are a few shortcomings:
  - 57 observations in the top income bracket!
    - Where are the standard errors?
  - Is panel data better than a repeated cross section?
    - Mean reversion issues?
    - Conditions on tax group in 1985...is that the right counterfactual?

Gruber and Saez (2002, JPubEc)

- Gruber and Saez 2002 fix the mean reversion issue
  - Use variation in tax rates from 1979-1990
- Consider model where income $y_{it} = y_{it}^0 (1 - \tau_{it})^{\epsilon}$ where ϵ is the taxable income elasticity
- Taking logs:

$$log\left(\frac{y_{it+3}}{y_{it}}\right) = \alpha + \epsilon log\left(\frac{1-\tau_{it+3}}{1-\tau_{it}}\right) + \eta_{it}$$

  - There are two reasons people face higher taxes:
    - They choose to earn more
    - For a given earnings, policy yields different tax rates
    - Hence, $1 - \tau_{it+3}$ is correlated with η_{it} because τ_{it+3} is a function of y_{it+3}! Need an instrument to isolate policy changes!
  - IV using $\tau_{it+3}^P = T'_{t+3}(y_{it})$: Instrument for marginal tax rate in $t+3$ if the only thing that changed was the tax policy (not y!).
  - Yields 0.3-0.6 estimate of ETI.

Top Tax Rate

- Goodsbee 2002 JPE
- Uses panel data on corporate compensation
- Shows taxable income declines after OBRA93
    - Short run elasticity > 1 (consistent with Feldstein 1995)
- But it's all short term
    - One-year elasticity 0-.4
    - Changes in timing of compensation around the introduction of OBRA
- Entirely driven by changes in the exercising of stock options around the time of OBRA!
    - Increased 1992 taxable income, decreased 1993 taxable income

Are behavioral responses to top tax rates real vs avoidance?

- Large literature on tax avoidance / tax evasion
- Feldstein (1999)
  - OK if y is taxable income
- Chetty (2009)
  - What if evasion has externalities / internalities?
- Evasion effects the optimal design
  - Sandmo 1981
  - Slemrod and Yitzhaki (2002 Handbook)
  - Dina Pomerantz (2013 JMP) VAT...

General Solution

- We ask: what are the properties of $T(y)$ that maximize the government's objective?
- If $T(y)$ is optimal, then should be indifferent to small changes in $T(y)$
- Calculus of Variations approach: Vary $T(y)$.
- Consider calculus of variations in $T(y)$
  - Define $\hat{T}(y; y^*, \epsilon, \eta)$ by

$$\hat{T}(y; y^*, \epsilon, \eta) = \begin{cases} T(y) & \text{if } y \notin \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right) \\ T(y) - \eta & \text{if } y \in \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right) \end{cases}$$

  - $\hat{T}$ provides η additional resources to an ϵ-region of individuals earning between $y^* - \epsilon/2$ and $y^* + \epsilon/2$.
  - Given $\hat{T}$, individual of type θ chooses $\hat{y}(y^*, \epsilon, \eta; \theta)$ that maximizes utility
    - Intuitively, some people who earn near y^* might move away from y^* because the government is taxing them more (or move towards y^* if $\eta < 0$)

Causal effects vs. IC constraints

- Define choice of income, y, in environment with ϵ and η by

$$\hat{y}\left(\theta; y^*, \epsilon, \eta\right) = argmax \quad u\left(y - \hat{T}\left(y; y^*, \epsilon, \eta\right), y; \theta\right)$$

- Why might an individual change choice of $\hat{y}$?
- Why do we care about these changes?
  - Impact on government revenue!
  - Do individuals internalize this impact on government revenue? NO.
- Where are the IC constraints?
  - Embedded in $\hat{y}$ function - we substitute the maximization program into the resource constraint and assume observed behavior maximizes the IC constraint
  - Therefore, we need causal effects of policy changes instead of a full solution to the programming problem.

Government Revenue Impact

- Given choices $\hat{y}\left(y^*, \epsilon, \eta; \theta\right)$, government revenue is given by

$$\hat{q}\left(y^*, \epsilon, \eta\right) = \frac{1}{\Pr\left\{y\left(\theta\right) \in \left[y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2}\right]\right\}} \int_\theta \hat{T}\left(\hat{y}\left(\theta; y^*, \epsilon, \eta\right); y^*, \epsilon,$$

(normalized by the number of mechanical beneficiaries).

  - Note $\hat{q}\left(y^*, 0, \eta\right) = \hat{q}\left(y^*, \epsilon, 0\right) = 0$ for all ϵ and η

Social Welfare

- Recall social welfare:

$$W = \int v(\theta)\,\psi(\theta)\,d\mu(\theta)$$

- Define social welfare $\hat{W}(y^*, \epsilon, \eta)$ to be social welfare under $\hat{T}(y; y^*, \epsilon, \eta)$
- Let $g(\theta)$ denote the social marginal utility of income for type θ:

$$g(\theta) = \frac{dW}{dy_\theta} = \lambda(\theta)\,\psi(\theta)$$

where λ is the individual's marginal utility of income
- So, g is the impact on social welfare of giving type θ an additional $1.
- Ratios of g are Okun's bucket (Okun (1975))

$$\frac{g(\theta_1)}{g(\theta_2)} = 2$$

implies indifferent to $1 to type θ_1 relative to $2 to type θ_2

Welfare Impact

- Welfare impact of increasing η
- Can use envelope theorem:
  - Marginal welfare impact given by mechanical loss in income weighted by social marginal utility of income:

$$\frac{dW}{d\eta}|_{\eta=0} = \int g(\theta) 1 \left\{ y \in \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2} \right) \right\} d\mu(\theta)$$

  - Note this assumes partial equilibrium (no welfare impact of changing taxes on those not directly affected)
- Marginal cost given by

$$\Pr\left\{ y(\theta) \in \left[y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2} \right] \right\} \frac{d\hat{q}(y^*, \epsilon, \eta)}{d\eta}|_{\eta=0}$$

- Welfare impact per dollar of government budget:

$$\frac{\frac{1}{\Pr\left\{ y(\theta) \in \left[y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2} \right] \right\}} \int g(\theta) 1 \left\{ y \in \left(y^* - \frac{\epsilon}{2}, y^* + \frac{\epsilon}{2} \right) \right\} d\mu(\theta)}{\frac{d\hat{q}(y^*, \epsilon, \eta)}{d\eta}|_{\eta=0}}$$

Optimality Condition

- Calculus of Variations implies welfare impact per dollar should be equated for all y^* and all $\epsilon > 0$!
  - At the optimum, the government is indifferent to small variations in $T(y)$
  - Otherwise wouldn't be at an optimum!
- General optimality condition
  - Let $q(y) = \lim_{\epsilon \to 0} \left(\frac{d\hat{q}(y,\epsilon,\eta)}{d\eta}|_{\eta=0} \right) = 1 + FE(y)$
  - Marginal revenue gained from imposing \$1 of welfare loss on individuals earning y

Optimality as equating MVPFs

- Consider policy of giving money to y_1:

$$MVPF_{y_1} = \frac{E\left[g\left(\theta\right)|y\left(\theta\right) = y_1\right]}{q\left(y_1\right)} = \frac{E\left[g\left(\theta\right)|y\left(\theta\right) = y_1\right]}{1 + FE\left(y_1\right)}$$

- Necessary condition for optimality: Indifferent to giving more money to y_1 vs. y_2.:

$$MVPF_{y_1} = MVPF_{y_2}$$

or

$$\frac{E\left[g\left(\theta\right)|y\left(\theta\right) = y_1\right]}{1 + FE\left(y_1\right)} = \frac{E\left[g\left(\theta\right)|y\left(\theta\right) = y_2\right]}{1 + FE\left(y_2\right)}$$

- Must be indifferent to budget-neutral manipulations of resources
    - Relative cost given by impact of small variations in tax schedule

Income and Substitution Effects

- Can write $FE(y)$ using behavioral elasticities...
- For linear tax rates,

$$E\left[g(\theta)\,|y(\theta)=y^*\right] = 1 - \underbrace{\frac{\tau(y^*)}{1-\tau(y^*)}\zeta(y^*)}_{\text{Income Effect}} + \underbrace{\frac{\tau(y)}{1-\tau(y)}\frac{d}{dy}|_{y=y^*}\left[\frac{yf(y)}{f(y^*)}\epsilon^c(y)\right]}_{\text{Substitution Effect}}$$

Optimal income tax schedule depends on behavioral distortions

Simplification: Diamond/Mirrlees/Saez formula

- Assume:
  - No income effects: $\zeta = 0$
  - Constant compensated elasticity ϵ^c
  - g' is constant conditional on y
- Then

$$\frac{g(y)-1}{\epsilon^c} = \underbrace{\frac{\tau(y)}{1-\tau(y)} \frac{d}{dy}\Big|_{y=y^*} \left[\frac{yf(y)}{f(y^*)}\right]}_{\text{Substitution Effect}}$$

- Some intuition: In regions where the optimal tax is linear:

$$\frac{1-\tau}{\tau} = \frac{\epsilon}{1-g(y)}\alpha$$

  where $\alpha = -1 - \frac{dlog(f(y))}{dlog(y)}$ is the elasticity of the income distribution (constant if Pareto)
- Tax is higher when elasticity is lower, g is lower, alpha is lower.

Laffer Tax

- Suppose $g' = 0$
- Then,

$$\frac{\tau}{1-\tau} = \frac{1}{\epsilon\alpha}$$

or

$$\tau\epsilon\alpha = 1 - \tau$$

or

$$\tau = \frac{1}{1+\epsilon\alpha}$$

which is Diamond-Saez 2011 formula for revenue maximizing top tax rate

General Mirrlees Formula

- Optimal Tax solves (τ linear)

$$\frac{g(y)-1}{\epsilon} f(y) = \frac{d}{dy}\Big|_{y=y^*}\left[\frac{\tau(y)}{1-\tau(y)} y f(y)\right]$$

- Fundamental Thm of Calculus:

$$\left[\lim_{\tilde{y}\to\infty} \frac{\tau(y)}{1-\tau(y)} \frac{yf(y)}{f(y^*)}\right] - \frac{\tau(y)}{1-\tau(y)} y f(y) = \int_y^\infty \frac{g(\tilde{y})-1}{\epsilon} f(\tilde{y})\, d\tilde{y}$$

- Generally, $\lim_{\tilde{y}\to\infty} \frac{\tau(y)}{1-\tau(y)} \frac{yf(y)}{f(y^*)} = 0$ (e.g. if f is pareto, $f \propto y^{-\alpha-1}$)
- So

$$\frac{\tau(y)}{1-\tau(y)} y f(y) = \int_y^\infty \frac{1-g(\tilde{y})}{\epsilon} f(\tilde{y})\, d\tilde{y}$$

General Mirrlees Formula (Diamond 1998)

- Re-writing:

$$\frac{\tau(y)}{1-\tau(y)}\alpha\epsilon = 1 - G(y)$$

where

$$G(y) = \frac{1}{1-F(y)}\int_y^\infty g(\tilde{y})\, f(\tilde{y})\, d\tilde{y}$$

is the average social marginal utilities on those earning more than y

$$\alpha(y) = \frac{yf(y)}{1-F(y)}$$

is the local Pareto parameter of the income distribution

General Mirrlees Formula (Diamond 1998)

- Re-writing:

$$\frac{\tau(y)}{1-\tau(y)} = \frac{1-G(y)}{\alpha\epsilon}$$

- Title of Diamond (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates"

Optimal Top Tax Rate is Zero?

- Wisdom prior to Saez (2001): Mirrlees Optimal Tax program implies "no tax at the top".
  - Celebrated result of Mirrlees (1971)
- Suppose income distribution is bounded, $y \leq \bar{y}$.
- Should the marginal tax rate be positive at $\bar{y}$?
  - Suppose it is.
  - Lower the tax rate slightly at $\bar{y} + \epsilon$ −> individuals may increase earnings to $\bar{y}$
    - Increases tax revenue as long as marginal tax rate on earnings $\bar{y}$ to $\bar{y} + \epsilon$ is positive
  - No individuals above $\bar{y}$ -> no one will decrease their earnings
    - Therefore, optimality requires the tax on earnings $= 0$ from $\bar{y}$ to $\bar{y} + \epsilon$

Optimal Top Tax Rate is Zero?

- Top tax rate $= 0$ only applies to the person who earns the maximum
- But, also applies for some distributions:
  - Suppose $\alpha(y) \to \infty$.
  - Recall $\alpha(y) = -1 - \frac{dlog(f)}{dlog(y)} = -1 - \frac{yf'(y)}{f(y)}$
  - If f'/f is constant, then $\alpha \to -\infty$, so that $\tau \to 0$
    - e.g. y exponential $\lambda e^{-\lambda y}$ -> $\lambda^2 e^{-\lambda y}$, so that

$$y\frac{f'(y)}{f(y)} = y\lambda$$

      which blows up.
  - But if f is Pareto, then α is constant
    - Turns out Pareto is good approximation (Saez 2001) -> optimal tax rate is not zero at the top.
    - Need tax return data to know this!

Participation Effects

- Implicitly, we assumed behavioral responses were continuous
  - Local income and substitution effects govern cost from behavioral responess
- But what if people enter the labor force?
  - Saez 2002
  - Kleven and Kreiner 2006

Mirrlees Model assumes no general equilibrium

[this section walks through proof of optimal tax with GE effects a la RS QJE 2013 roy model...]

- Mirrlees model assumes no general equilibrium effects of tax changes
  - Individuals θ_1's choice of y doesn't affect the tradeoffs faced by θ_2
  - Own-earnings impact of θ_1 earning y is same as GDP impact of θ_1 increasing y
- Assume there exists a utility function

$$u\left(c, y, \theta\right)$$

  where $y = f\left(l; \theta\right)$
- But, what if choice of labor effort of type 1 affects earnings of type 2?

Two Sources of GE Bias

- "Trickle down" GE effects
  - Rich people work harder -> poor people make more money?
  - Build businesses, "job creators", etc.
  - Rothschild and Scheuer (QJE 2013)
- "zero-sum" GE effects
  - Rich people earn more money -> they steal it from poor people
  - Rent-seeking / investment banking
  - Rothschild and Scheuer (Econometrica 2013)
- The presence of such GE effects changes the optimal income tax schedule

Trickle Down/Up Economics

- Suppose earnings of y given by

$$y(\theta) = f\left(\{l(\theta)\}_{\theta}\right)$$

  where l is effort of type θ
  - Allows for arbitrary inter-connectedness of production/earnings
- Suppose govt icnreases transfers to those earning y by η dollars
  - Has direct effect on those earning y
  - Behavioral responses can directly affect earnings of those with earnings away from y
- How do we value these impacts?
  - Envelope Theorem

Envelope Theorem

- Utility
$$u\left(c,l,\theta\right)$$
- Earnings of type θ
$$y\left(\theta\right)=w\left(\theta\right)l\left(\theta\right)$$
- Suppose we change the top marginal income tax rate, τ^{top}
- Impact on wages of type θ is $\frac{dw}{d\tau^{top}}$
- Welfare impact on type θ not directly affected by the transfer is given by
$$\frac{dV}{d\tau^{top}}=\frac{dw}{d\tau^{top}}l=\frac{dy\left(\theta\right)}{d\tau^{top}}|_{l}$$
earnings impact holding behavioral responses constant

Welfare Impact

- Combining with the impact of the transfer, we have an aggregate welfare impact from the tax policy giving transfers near y as:

$$\frac{dW}{d\tau^{top}}|_{\eta=0} = g'_{top} + \underbrace{\int g'(\tilde{y}) \frac{dy}{d\tau^{top}}|_{I} dy}_{\text{Externalities}}$$

  where g'_{top} is the average social marginal utility of income on the rich (continue to assume $g'_{top} = 0$)
- Trickle down:

$$\int g'(\tilde{y}) \frac{dy}{d\tau^{top}}|_{I} dy > 0$$

  Giving money to rich people increases incomes of poor:
- But could go the other way (e.g. fixed job slots):

$$\int g'(\tilde{y}) \frac{dy}{d\tau^{top}}|_{I} dy < 0$$

Optimality condition with externalities

- Externalities impact direct welfare impact of tax changes
- But, now behavioral responses are also more complicated...
  - Maybe the poor offset the mechanical impact on earnings by increased labor supply?
  - Effects fiscal cost of tax change
- Total revenue from increasing τ:

$$dR = \underbrace{E\left[y|y \geq \bar{y}\right] - \bar{y}}_{\text{Mechanical}} + \underbrace{\tau^{top} \frac{dE\left[y|y \geq \bar{y}\right]}{d\tau^{top}}}_{\text{Behavioral}} + \underbrace{\int_{y<\bar{y}} T'\left(y\right) \frac{dy}{d\tau^{top}}}_{\text{Trickle-down Rev Impact}}$$

assuming continuously differentiable responses

Optimal Tax Rate

- Normalize g' so that g' is 1 for a policy that provides income across the distribution (dW=dR+dW
- Optimality implies

$$\tau^* = \underbrace{\frac{E\left[y|y \geq \bar{y}\right] - \bar{y}}{-\frac{dE[y|y\geq\bar{y}]}{d\tau}}}_{\text{Original Formula}} + \underbrace{\frac{\int_{y<\bar{y}} T'\left(y\right) \frac{dy}{d\tau^{top}}}{-\frac{dE[y|y\geq\bar{y}]}{d\tau}}}_{\text{Fiscal Externality}} - \underbrace{\frac{\int g'\left(\tilde{y}\right) \frac{dy}{d\tau^{top}}|_I dy}{-\frac{dE[y|y\geq\bar{y}]}{d\tau}}}_{\text{Direct Externality}}$$

- Suppose fiscal externality small:
  - Trickle down externalities -> lower top marginal tax rate
  - Fixed jobs -> higher top marginal tax rate
- Could the fiscal externality overwhelm?

Inverse Optimum Program

- Bourguignon and Spadaro (2012), Werning (2007), Jacobs, Jongen, and Zoutman (2013; 2014), Lockwood and Weinzierl (2014), Hendren (2014)
- Traditional "optimal tax": Start with primitives, derive optimal tax rates
- Inverse optimum approach: Start with status quo, derive implicit weights, $g(y)$.

Inverse Optimum Program: A Crazy Approach?

- You may think this is crazy (why would I like Harry Reid's preferences!?). But, it is not! There are three key reasons to like this type of an approach even if you don't think govts are optimizing:
  - Can be used to search for Pareto improvements of the tax schedule (Werning 2007, Bourguignon and Spadaro 2012).
  - For other policies besides the tax schedule, can use Kaldor-Hicks efficiency criterion, as opposed to subjective social preference
    - Weights correspond to the "correct" Kaldor Hicks corrections for the cost of transfers (Hendren 2014)
  - Although the inverse optimum approach is generically "not invertible" and the social welfare interpretation breaks down in multi-dimensional heterogeneity, the measure of these costs as "marginal costs" for the Kaldor-Hicks corrections remains quite general

Bourguignon and Spadaro

- "Tax-Benefit Revealed Social Preferences"
- We can identify $G(y)$ from:

$$\frac{\tau(y)}{1-\tau(y)} = \frac{1-G(y)}{\alpha(y)\epsilon}$$

where $\alpha(y) = \frac{yf(y)}{1-F(y)}$ and $G(y) = E\left[g(\tilde{y})\,|\tilde{y} \geq y\right]$. Re-writing:

$$G(y) = 1 + \alpha(y)\epsilon\frac{T'(y)}{1-T'(y)}$$

- And $g(y) = G'(y)$

Bourguignon and Spadaro

- Use data from France
  - Marginal tax rates
  - Wage distribution from survey data
  - Find $G'(y) < 0$ for high values of earnings
- Suggests implicit welfare weights are negative at the top

Bourguignon and Spadaro

- Concerns:
  - Use of HH data for wage distribution -> thin upper tail (declining $\alpha(y)$)
    - Mechanically generates optimal zero tax rate at top (as in Diamond 98)
  - Treatment of heterogeneity
    - Different tax rates for same wage...how to construct $T'(y)$ from survey data? See heterogeneity section below.

Werning (2009)

- Pareto Efficient Income Taxation
- Characterize when is a tax schedule constrained inefficient?
- Local Laffer Effects

$$g(y) = 1 - \frac{T'(y)}{1 - T'(y)} \epsilon \alpha(y) \geq 0$$

where $\alpha = -1 - \frac{dlog(f(y))}{dlog(y)}$ (derivative of α above...).

- Pareto efficiency test:

$$\frac{T'(y)}{1 - T'(y)} \epsilon \alpha(y) \leq 1$$

Generalization to multiple dimensions

- Suppose people have characteristics X (multi-dimensional). Obtain taxes/transfers $T(X)$.
- Construct $FE(X) = \lim_{\epsilon \to 0} FE(X, \epsilon)$, where $FE(X^*, \epsilon) = \frac{d}{d\eta} q(X, \eta, \epsilon) - 1$ is the fiscal externality from giving a tax cut to those with values of $X \in N_\epsilon(X^*)$
- Test:

$$FE(X) < -1$$

- Local Laffer Effects...
- But, testing $FE(X) < -1$ for all X is hard!
  - Think of all the possible values of X in reality! All of people's observable characteristics – very high dimensional!

Resolving interpersonal comparisons

- Difficult to find Laffer effects. Otherwise, need to resolve interpersonal comparisons
  - $g(\theta)$ can describe social preferences...but what weights to use?
- Kaldor and Hicks: can winner's hypothetically compensate losers through transfers?
  - Generally interpreted as lump-sum transfers (individual specific)
- Setup: Consider policy path P; construct WTP out of own income, $s(\theta) = \frac{dV}{\lambda}$
- Assume budget neutral policy for now
- In principle, $s(\theta) < 0$ for some >0 for others.
- KH test: does there exist transfers $t(\theta)$ s.t. everyone is better off.
- True iff

$$\underbrace{\int s(\theta)\, d\mu(\theta)}_{\text{Aggregate Surplus}} > 0$$

Lump sum vs. distortionary transfers

- KH motivates aggregate surplus (aka efficiency) as a normative criterion
- Mirrlees 71 observation: Transfers aren't free!
- Hendren (2014): Imagine the transfers occur through the income tax schedule
  - See also Coate (2000) and Kaplow (2006; 2008).
- Imagine taxing back the benefits to the government through changes in the tax schedule
- For now, assume benefits are homogeneous conditional on income $s(y)$

Multi-D heterogeneity

- Benefits $s(y)$. We tax them back s.t. government revenue is 1+FE per \$1 taxed:

$$(1 + FE(y))\, s(y)$$

- Can get positive govt revenue from spending the \$1 if

$$\int (1 + FE(y))\, s(y)\, dF(y) \geq 0$$

  If holds, then can do the policy, tax the benefits, and redistribute the surplus to everyone!

  - Benefit-based taxation (Musgrave).

- Motivates using weights $1 + FE(y)$ or

$$g(y) = \frac{1 + FE(y)}{E[1 + FE(y)]}$$

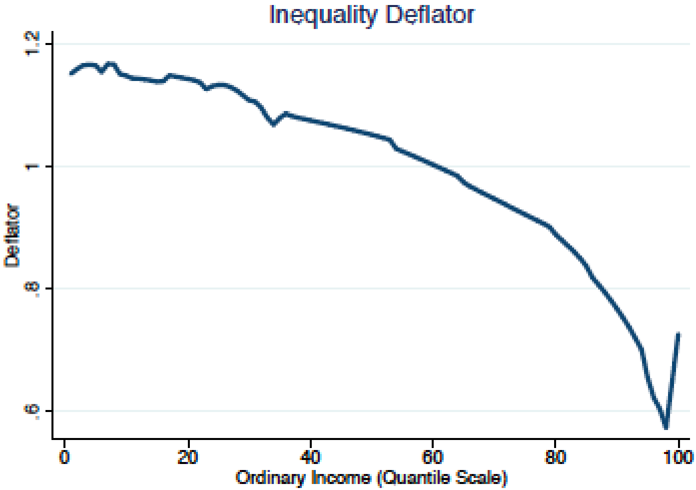
Two reasons to use these weights...

- $g(y)$ characterizes cost of providing \$1 of welfare to those earning y through modifications in the tax schedule.
  - Give higher weight to those more costly to reach!
- Two reasons to use these weights:
  - Positive (Hendren 2014): Can augment policy with benefit tax to make Pareto improvement
    - Prefer the policy by the (potential) Pareto principle
  - Normative (Bourguignon and Spadaro): $g(y)$ are welfare weights that rationalize status quo as optimum
    - $g(y)$ rationalizes status quo as optimal (assuming $g(y) > 0$).

Application to U.S. data (Hendren 2014)

- Hendren (2014) estimates $FE(y)$ by calibrating elasticities
- Estimates T' and α using universe of income tax returns
  - Accounts for covariance between T' and α
  - Contrasts with existing approaches assuming single tax rate schedules (e.g. Bourguignon and Spadaro, Jacobs et. al.)

Application to U.S. data (Hendren 2014)



Multi-D heterogeneity

- What to do if there is heterogeneity conditional on income
  - in earnings y
  - in surplus $s(\theta)$

Commodity versus Income Taxation

- Mirrlees model assumes only income tax
  - What about commodity taxes? Or other taxes?
- Diamond-Mirrlees (1971, AER) calculates optimal commodity taxes in world with no lump-sum taxation
  - Leads to inverse elasticity rule
  - Consider policy variation around an optimum that solely changes commodity taxes:

$$\frac{dV^P}{d\theta} = 0$$

which implies

$$\sum_k \hat{\tau}_k \frac{d\hat{x}_k}{d\theta} = 0$$

At an optimum, budget-neutral policy changes have no utility impacts -> compensated elasticities.

- But, suppose there was lump-sum taxation -> optimal distortionary tax is zero.
  - Need to nest commodity taxes in model with rational desire to avoid

Commodity Taxation Revisited

- Does commodity taxation have a role if we have a nonlinear income tax (with lump-sum)?
  - Need to put commodity taxes into Mirrlees (1971) framework
  - Atkinson and Stiglitz (1976) JPubEc
    - Follow Kaplow (2006, JPubEc) for a simple proof

Kaplow (2006)

- Setup
- Individuals choose commodities $\{c_1, c_2 ...\}$ and labor effort, l
- Maximize utility function

$$u_h(c_1, c_2, ..., l) = \tilde{u}_h(g(c_1, ...), l)$$

  - **Key assumption:** g is the same across people
- Subject to budget constraint

$$\sum (p_i + \tau_i) c_i \leq wl - T(wl)$$

  where w is an individual's wage (heterogeneous in population)
- wl is earnings and $T(wl)$ is the (nonlinear) tax on earnings

Statement

- Suppose there is a commodity tax
$$\frac{p_i + \tau_i}{p_j + \tau_j} \neq \frac{p_i}{p_j}$$
for some i and j
- Can welfare be improved by re-setting $\tau_i = \tau_j = 0$ and suitably augmenting the tax schedule T?
  - Atkinson-Stiglitz/Kaplow: YES.
- Define $V(\tau, T, wl)$ to be
$$V(\tau, T, wl) = max\ v(c_1, c_2, ...)$$
$$s.t. \sum (p_i + \tau_i) c_i \leq wl - T(wl)$$
- V is the value of the consumption argument of the utility function – holds independent of labor effort l!
  - Consumption allocations don't reveal any information about labor supply type w conditional on wl.

Proof

- Define intermediate environment:
  - Start with commodity taxes τ
  - Define new taxes at zero $\tau_i^* = 0$
  - Augment the tax schedule
    - Define T^* to offset the impact on utility so that utility is held constant in this intermediate world
  - Specifically, T^* satisfies

$$V(\tau, T, wl) = V(\tau^*, T^*, wl)$$

for all wl

Proof (Cont'd)

- Lemma 1: Every type w chooses the same level of labor effort under τ^*, T^* as under τ, T.
- Proof:
  - Note that

    $$U(\tau, T, w, l) = u(V(\tau, T, wl), l) = u(V(\tau^*, T^*, wl), l) = U(\tau^*, T^*, w$$

    so that utility is the same in both environments for a given individual for any choice of l.
  - Therefore, the l that maximizes utility in the original world maximizes utility in the intermediate world

Proof Cont'd

- Lemma 2: The augmented world raises more revenue than the original world
- Proof:
  - Will show that no individual in the intermediate regime can afford the original consumption vector
    - Implies they pay more taxes in intermediate regime
  - Suppose type w can afford original vector
    - Then she strictly prefers a different vector because of change in relative price
    - Implies intermediate environment is strictly better off -> contradicting definition of intermediate environment holding utilities constant

Proof Cont'd

- Why does this imply aggregate tax revenue is higher in the intermediate environment?
- We have:
$$\sum (p_i) c_i > wl - T^* (wl)$$
for all wl (note $\tau^* = 0$)
- Budget constraint in initial regime implies
$$\sum_i (p_i + \tau_i) c_i = wl - T (wl)$$
so that
$$\sum_i p_i c_i = -\sum_i \tau_i c_i + wl - T (wl)$$
- So that
$$-\sum_i \tau_i c_i + wl - T (wl) > wl - T^* (wl)$$
or
$$T^* (wl) > \sum \tau_i c_i + T (wl)$$
- QED

Proof Cont'd

- So, the intermediate world generates more tax revenue and holds utility constant
- Now, generate a third world that gives ϵ benefits to everyone through lowering the tax schedule
- Implies everyone better off.

Implications of Atkinson Stiglitz

- Incredibly powerful
- Implies zero capital taxes in the standard model
- Nests the celebrated "production efficiency" theorem of Diamond and Mirrlees (1971)

Capital Taxes

- Suppose

$$U(c_1, c_2, ..., l) = u(c_1) - v(l_1) + \beta[u(c_2) - v(l_2)] + ...$$

- With budget constraint

$$\sum_i (p_i + \tau_i) c_i \leq \sum_i w_i l_i$$

- So

$$g(c_1, c_2, ...) = u(c_1) + \beta u(c_2) + ...$$

- Implies no distortion in relative price of c_1 and c_2
  - You should prove extension to case with l_i instead of just l.
- What if more productive types have higher preferences for bequests?

Production Efficiency

- Diamond and Mirrlees (1971) show a surprising result:
- Suppose C is produced with a bunch of intermediate inputs, x_i

$$C = f(x_1, ..., x_n)$$

- Question: would you ever want to tax these inputs?
- Answer: No!

$$u(x, l) = U(C(x), l)$$

- The production function for C is the same for all people
  - Weak separability holds
  - Implies no taxes on intermediate inputs

When does Atkinson-Stiglitz Fail?

- Mirrlees information logic:
  - When commodity choices have desirable information about type conditional on earnings!
- What constitutes "desirable information"? (Saez 2002 JPubEc)
  - Information about social welfare weights: Society likes people that consume x_1 more than x_2 conditional on earnings
    - Implement subsidy on good x_1 financed by tax on x_2
    - First order welfare gain (b/c of difference in social welfare weights)
    - Second order distortionary cost starting at $\tau = 0$
  - Information about latent productivity: More productive types like x_1 more than x_2 conditional on earnings
    - e.g. x_1 is books; x_2 is surf boards
    - Then, tax the goods rich people like but reduce the marginal tax rate
    - Leads to increase in earnings!
    - Depends on covariance

Back to Diamond and Mirrlees (1971) Optimal Commodity Taxation

- Diamond and Mirrlees (1971) derive conditions for optimal commodity taxation
- Consider model without an intercept in the tax schedule (i.e. no lump-sum transfers)
- Result: Taxes on goods inverse to their behavioral responses
    - Inverse elasticity rule
- **My view: This is arguably the most mis-understood result in all of public economics.**
    - Because intercept is ruled out, introduces desire to tax inelastic goods because they replicate the lump-sum tax.
    - But with lump-sum tax, this desire goes away
    - Optimal commodity taxes depend on whether commodity choice provides systematic information about latent productivity and allows for a relaxation of the income distribution

Education Policy

Should we subsidize education?

- Bovenberg and Jacobs (2005 JPubEc); Stantcheva (2013, JMP)
- Setup:
  - l is labor effort (unobserved)
  - y is an individual's production (observed)
  - θ is an individual's type (unobserved)
  - h is human capital investment (observed)
  - Arbitrary production (for now):

$$y = f(h, l, \theta)$$

  - Utility

$$U(c, l, h, \theta)$$

Education Policy

- Question: How should we tax/subsidize human capital in addition to nonlinear income tax $T(y)$?
- Special Case: Can we apply Atkinson-Stiglitz theorem?
- Case 1: h is pure consumption and has no impact on production
  - If $U = \tilde{U}(g(c,h), l, \theta)$, then AS implies no tax/subsidy for education
- Case 2: High earning people like human capital more than poor people:

$$\frac{g_c^{Poor}}{g_h^{Poor}} > \frac{g_c^{Rich}}{g_h^{Rich}}$$

  - Tax it (allows lower marginal tax on earnings)

- Case 3: h has productive value and no utility value
  - Follow Bovenberg and Jacobs (2005, JPubEc)
- Production function

$$y = l\phi\left(h\right)\theta = lh^{\beta}\theta$$

- Utility

$$U = c - \frac{l^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}}$$

- This is more complicated...turn to Hamiltonian approach

Hamiltonian Approach

- Common method for solving uni-dimensional screening problems: Use a Hamiltonian
  - Original Approach by Mirrlees (1971)
- WLOG?, government chooses menu of $\{c(\theta), l(\theta), h(\theta)\}_\theta$ to maximize social welfare:

$$\int v(\theta)\psi(\theta)\,d\theta$$

  - Will sometimes switch l and y...
- Subject to IC constraints and aggregate resource constraints

Switch to utility space

- Common approach to using the hamiltonian: switch to utility space and solve for utility

$$c\left(v,y,h\right)=v+\frac{\left(\frac{y}{\theta\phi(h)}\right)^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}}$$

so given v, l, and h, c is the level of consumption that delivers utility v

  - Helpful to have quasilinear utility here...

IC and Resource Constraint

- Start with IC constraint:
  - Define

$$\hat{v}\left(\theta,\hat{\theta}\right)=u\left(c\left(\hat{\theta}\right),y\left(\hat{\theta}\right),h\left(\hat{\theta}\right);\theta\right)=c\left(\hat{\theta}\right)-\frac{\left[\frac{y\left(\hat{\theta}\right)}{\phi\left(h\left(\hat{\theta}\right)\right)\theta}\right]^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}}$$

    utility of type θ who says they are type $\hat{\theta}$
  - IC constraint is

$$v\left(\theta\right)=max_{\hat{\theta}}\hat{v}\left(\theta,\hat{\theta}\right)\quad\forall\theta$$

  - Each type prefers truth-telling
- Resource Constraint:

$$\int T\left(h\left(\theta\right),y\left(\theta\right)\right)=\int\left(y\left(\theta\right)-c\left(\theta\right)-h\left(\theta\right)\right)d\theta\geq 0$$

First Order Approach

- Under a single crossing assumption, the global incentive constraints can be replaced with local incentive constraints
- Local IC constraints described by envelope theorem:

$$
\begin{aligned}
v'(\theta) &= \frac{\partial u}{\partial \theta} = -\left(\frac{y(\theta)}{\phi(h(\theta))}\right)^{1+\frac{1}{\epsilon}} \frac{\frac{d}{d\theta}\theta^{-\left(1+\frac{1}{\epsilon}\right)}}{1+\frac{1}{\epsilon}} \\
&= -\left(\frac{y(\theta)}{\phi(h(\theta))}\right)^{1+\frac{1}{\epsilon}} \frac{\frac{d}{d\theta}\theta^{-\left(1+\frac{1}{\epsilon}\right)}}{1+\frac{1}{\epsilon}} \\
&= -\left(\frac{y(\theta)}{\phi(h(\theta))}\right)^{1+\frac{1}{\epsilon}} \theta^{\frac{1}{\epsilon}} \\
&= \frac{1}{\theta}\left(\frac{y(\theta)}{\phi(h(\theta))\,\theta}\right)^{1+\frac{1}{\epsilon}}
\end{aligned}
$$

- Note that $v'(\theta) > 0$. Implies more productive types must get higher utility...

Hamiltonian

- Hamiltonian:
  - Think of θ as "time"
  - $v(\theta)$ is the state variable (we have a constraint for $v'(\theta)$)
  - control variables (aka co-state variables): $h(\theta)$, $y(\theta)$, and $c(\theta)$

Hamiltonian

$$H = v(\theta)\psi(\theta) - \gamma_{IC}(\theta)\left[\frac{1}{\theta}\left(\frac{y(\theta)}{\phi(h(\theta))\theta}\right)^{1+\frac{1}{\epsilon}}\right] + \gamma_{RC}\left[y(\theta) - h(\theta) - c(y\right.$$

- Why is this useful? Well, we're done. We can solve for the optimal subsidy on human capital...
- At the optimum, $\frac{\partial H}{\partial f(X)} = 0$ for all ctsly diff functions of control variables $f(X)$.
    - So, substitute back $l(\theta)$ instead of $y(\theta)$.

    $$H = v(\theta)\psi(\theta) - \gamma_{IC}(\theta)\frac{1}{\theta}l^{1+\frac{1}{\epsilon}} + \gamma_{RC}\left[\theta l(\theta)\phi(h(\theta)) - h(\theta) - c(\theta l(\theta\right.$$

    - Now, take derivative wrt h holding l and v constant:

    $$\frac{\partial H}{\partial h} = \gamma_{RC}\left[\theta l(\theta)\phi'(h(\theta)) - 1 - \frac{dc}{dh}|_{l,v}\right] = 0$$

    where $\frac{dc}{dh}|_{l,v} = 0$

How general is this?

- Why is the tax precisely zero?
  - Because taxing it provides no help in screening (h does not help in IC constraint)
  - Is this general?
- No: Depends on shape of incentive constraints (Stantcheva 2013).
- Consider general case: $y(\theta) = \phi(h, \theta)$:

$$\hat{v}\left(\theta, \hat{\theta}\right) = c\left(\hat{\theta}\right) - \frac{\left[\frac{y\left(\hat{\theta}\right)}{\phi\left(h\left(\hat{\theta}\right), \theta\right)}\right]^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}}$$

Then,

$$v'(\theta) = \left[\frac{y\left(\hat{\theta}\right)}{\phi\left(h\left(\hat{\theta}\right), \theta\right)}\right]^{\frac{1}{\epsilon}} \frac{y(\theta)\frac{\partial\phi}{\partial\theta}}{\phi^2} = \left[\frac{y\left(\hat{\theta}\right)}{\phi\left(h\left(\hat{\theta}\right), \theta\right)}\right]^{1+\frac{1}{\epsilon}} \frac{\frac{\partial\phi}{\partial\theta}}{\phi}$$

- Note that $\frac{\frac{\partial\phi}{\partial\theta}}{\phi}$ does NOT depend on h when $\phi = h\theta$. But, more generally it does.