

Lectures 2 and 3: Optimal Income Taxation

Stefanie Stantcheva

Fall 2019

GOALS OF THESE LECTURES

1) Understand the **core optimal income tax model**: linear and nonlinear taxes in the Saez (2001) framework.

General method, intuitive, sufficient statistics.

2) Introduce the mechanism design approach of **Mirrlees (1971)**.

Incentive compatibility, optimal control.

With and without income effects.

3) Extensions: Migration and rent-seeking

4) Should commodity taxes be used in addition to income taxes?

Atkinson-Stiglitz Theorem

OPTIMAL TAXATION: SIMPLE MODEL WITH NO BEHAVIORAL RESPONSES

Utility $u(c)$ strictly increasing and concave

Same for everybody where c is after tax income.

Income is z and is fixed for each individual, $c = z - T(z)$ where $T(z)$ is tax on z . z has density distribution $h(z)$

Government maximizes **Utilitarian** objective:

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

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

SIMPLE MODEL WITH NO BEHAVIORAL RESPONSES

Form lagrangian: $L = [u(z - T(z)) + \lambda \cdot T(z)] \cdot h(z)$

First order condition (FOC) in $T(z)$:

$$0 = \frac{\partial L}{\partial T(z)} = [-u'(z - T(z)) + \lambda] \cdot h(z) \Rightarrow u'(z - T(z)) = \lambda$$

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

$\Rightarrow c = \bar{z} - E$ where $\bar{z} = \int zh(z)dz$ average income.

100% marginal tax rate. Perfect equalization of after-tax income.

Utilitarianism with decreasing marginal utility leads to perfect egalitarianism [Edgeworth, 1897]

ISSUES WITH SIMPLE MODEL

1) **No behavioral responses:** Obvious missing piece: 100% redistribution would destroy incentives to work and thus the assumption that z is exogenous is unrealistic

⇒ Optimal income tax theory incorporates behavioral responses (Mirrlees REStud '71): **equity-efficiency trade-off**

2) **Issue with Utilitarianism:** Even absent behavioral responses, many people would object to 100% redistribution [perceived as confiscatory]

⇒ Citizens' views on fairness impose **bounds** on redistribution.

The issue is the restricted nature of social preferences that can be captured by most social welfare functions.

We will discuss preferences for redistribution in another lecture! For now we remain agnostic about the " g_i ".

MIRRLEES OPTIMAL INCOME TAX MODEL

We will solve the Mirrleesian model later. For now, let's look at the spirit of optimal tax evolution.

1) **Standard labor supply model:** Individual maximizes $u(c, l)$ subject to $c = wl - T(wl)$ where c consumption, l labor supply, w wage rate, $T(\cdot)$ nonlinear income tax \Rightarrow taxes affect labor supply

2) **Individuals differ in ability w , private information w** distributed with density $f(w)$.

3) **Govt social welfare maximization:** Govt maximizes

$$SWF = \int G(u(c, l))f(w)dw$$

($G(\cdot) \uparrow$ concave) subject to

(a) budget constraint $\int T(wl)f(w)dw \geq E$ (multiplier λ)

(b) individuals' labor supply l depends on $T(\cdot)$

MIRROLEES MODEL RESULTS

Optimal income tax trades-off redistribution and efficiency (as tax based on w only not feasible)

$\Rightarrow T(.) < 0$ at bottom (transfer) and $T(.) > 0$ further up (tax) [full integration of taxes/transfers]

Mirrlees formulas complex, only a couple fairly general results:

- 1) $0 \leq T'(\cdot) \leq 1$, $T'(\cdot) \geq 0$ is non-trivial (rules out EITC) [Seade '77]
- 2) Marginal tax rate $T'(\cdot)$ should be zero at the top (if skill distribution bounded) [Sadka '76-Seade '77]
- 3) If everybody works and lowest $w/l > 0$, $T'(\cdot) = 0$ at bottom

HISTORY: BEYOND MIRRLEES

Mirrlees '71 had a huge impact on information economics: models with asymmetric information in contract theory

Discrete 2-type version of Mirrlees model developed by Stiglitz JpubE '82 with individual FOC replaced by Incentive Compatibility constraint [high type should not mimick low type]

Till late 1990s, Mirrlees results not closely connected to empirical tax studies and little impact on tax policy recommendations

Since late 1990s, Diamond AER'98, Piketty '97, Saez ReStud '01 have connected Mirrlees model to practical tax policy / empirical tax studies

[new approach summarized in Diamond-Saez JEP'11 and Piketty-Saez Handbook'13]

INTENSIVE LABOR SUPPLY CONCEPTS

$$\max_{c, z} u(c, z) \text{ subject to } c = z \cdot (1 - \tau) + R$$

Imagine a linearized budget constraint: R is virtual income (why virtual?) and τ marginal tax rate.

FOC in $c, z \Rightarrow (1 - \tau)u_c + u_z = 0 \Rightarrow$ Marshallian labor supply
 $z = z(1 - \tau, R)$

$$\text{Uncompensated elasticity } \epsilon^u = \frac{(1 - \tau)}{z} \frac{\partial z}{\partial (1 - \tau)}$$

$$\text{Income effects } \eta = (1 - \tau) \frac{\partial z}{\partial R} \leq 0$$

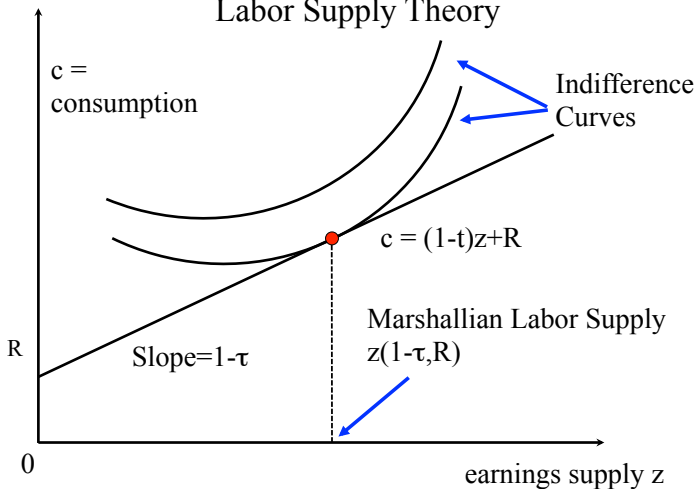
INTENSIVE LABOR SUPPLY CONCEPTS (II)

Substitution effects: Hicksian labor supply: $z^c(1 - \tau, u)$ minimizes cost needed to reach u given slope $1 - \tau \Rightarrow$

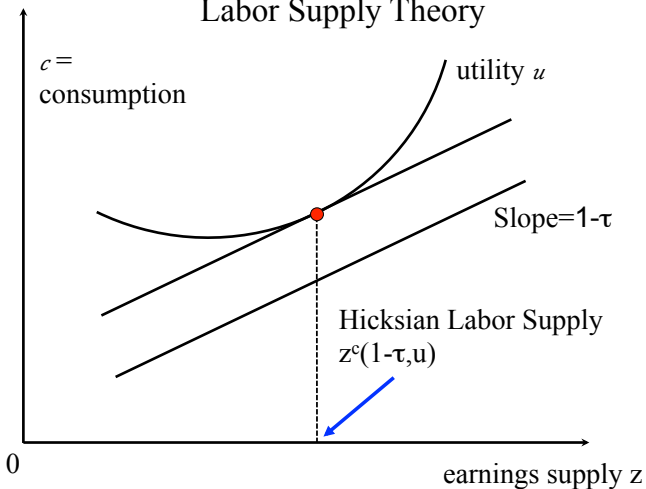
$$\text{Compensated elasticity } \varepsilon^c = \frac{(1 - \tau)}{z} \frac{\partial z^c}{\partial (1 - \tau)} > 0$$

$$\text{Slutsky equation } \frac{\partial z}{\partial (1 - \tau)} = \frac{\partial z^c}{\partial (1 - \tau)} + z \frac{\partial z}{\partial R} \Rightarrow \varepsilon^u = \varepsilon^c + \eta$$

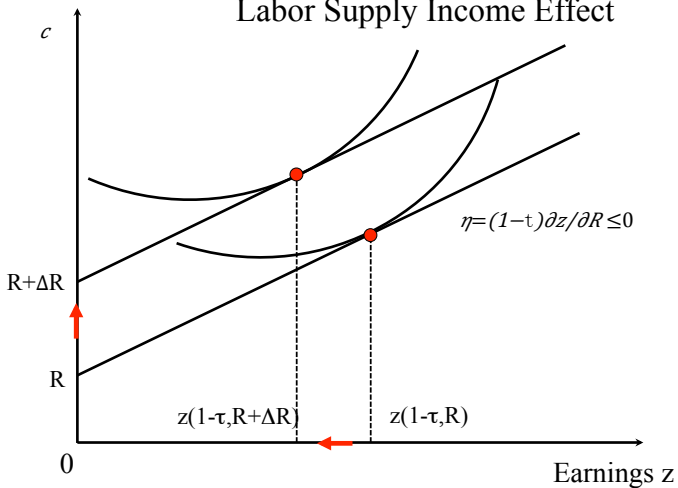
Labor Supply Theory



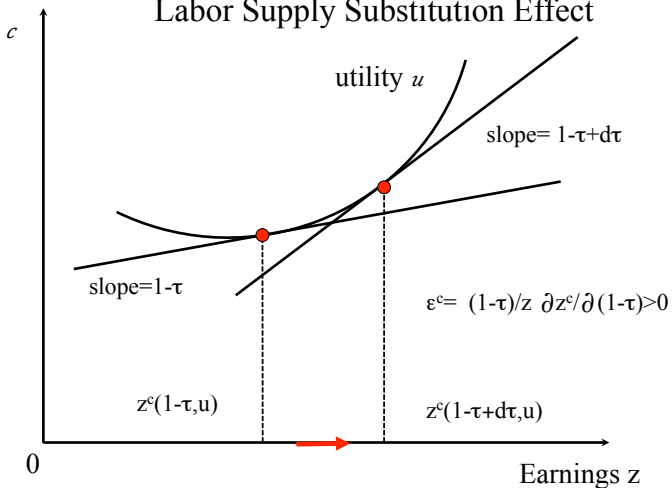
Labor Supply Theory



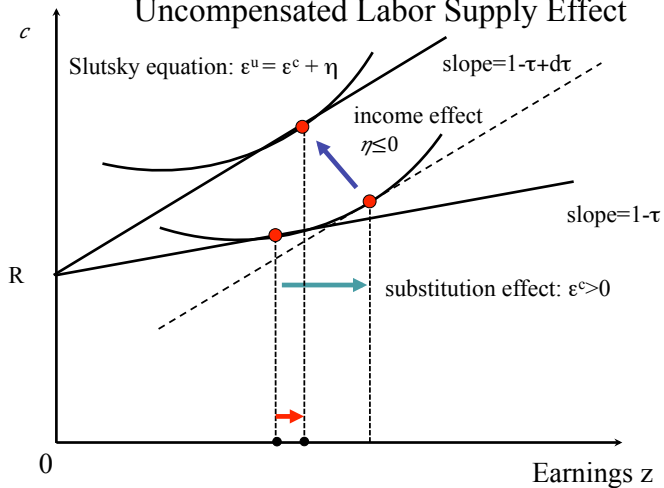
Labor Supply Income Effect



Labor Supply Substitution Effect



Uncompensated Labor Supply Effect



Labor Supply Effects of Taxes and Transfers

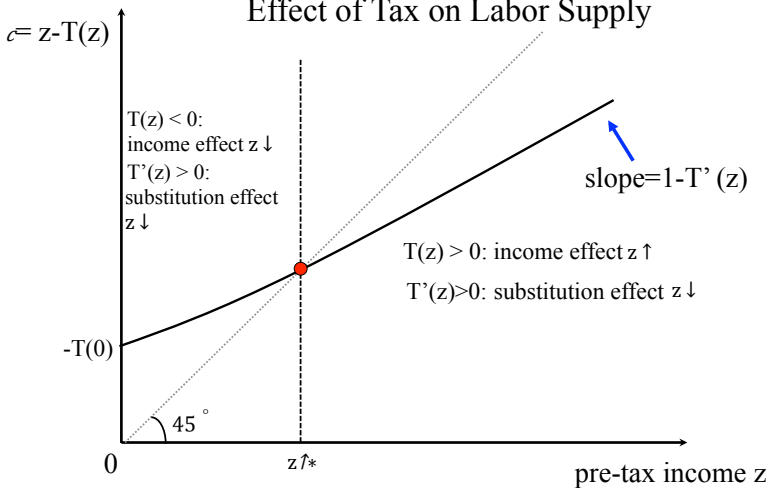
Taxes and transfers change the slope $1 - T'(z)$ of the budget constraint and net disposable income $z - T(z)$ (relative to the no tax situation where $c = z$)

Positive MTR $T'(z) > 0$ reduces labor supply through substitution effects

Net transfer ($T(z) < 0$) reduces labor supply through income effects

Net tax ($T(z) > 0$) increases labor supply through income effects

Effect of Tax on Labor Supply



WELFARE EFFECT OF SMALL TAX REFORM

Indirect utility: $V(1 - \tau, R) = \max_z u((1 - \tau)z + R, z)$ where R is virtual income intercept

Small tax reform: $d\tau$ and dR :

$$dV = u_c \cdot [-zd\tau + dR] + dz \cdot [(1 - \tau)u_c + u_z] = u_c \cdot [-zd\tau + dR]$$

Envelope theorem: no effect of dz on V because z is already chosen to maximize utility ($(1 - \tau)u_c + u_z = 0$)

$[-zd\tau + dR]$ is the mechanical change in disposable income due to tax reform

Welfare impact of a small tax reform is given by u_c times the money metric mechanical change in tax

WELFARE EFFECT OF SMALL TAX REFORM (II)

!! Remains true of any nonlinear tax system $T(z)$

Just need to look at $dT(z)$, mechanical change in taxes, or dT_i for agent i .

dV_i = Welfare impact is $-u_c dT(z_i)$.

When is the welfare impact not just the mechanical change in disposable income?

Envelope Theorem: For a constrained problem

$$V(\theta) = \max_x F(x, \theta) \quad \text{s.t.} \quad c \geq G(x, \theta)$$

$$V'(\theta) = \frac{\partial F}{\partial \theta}(x^*(\theta), \theta) - \lambda^*(\theta) \frac{\partial G}{\partial \theta}(x^*(\theta), \theta)$$

SOCIAL WELFARE FUNCTIONS (SWF)

Welfarism = social welfare based solely on individual utilities

Any other social objective will lead to Pareto dominated outcomes in some circumstances (Kaplow and Shavell JPE'01) Why?

Most widely used welfarist SWF:

1) Utilitarian: $SWF = \int_i u^i$

2) Rawlsian (also called Maxi-Min): $SWF = \min_i u^i$

3) $SWF = \int_i G(u^i)$ with $G(\cdot) \uparrow$ and concave, e.g., $G(u) = u^{1-\gamma}/(1-\gamma)$
(Utilitarian is $\gamma = 0$, Rawlsian is $\gamma = \infty$)

4) General Pareto weights: $SWF = \int_i \mu_i \cdot u^i$ with $\mu_i \geq 0$ exogenously given

SOCIAL MARGINAL WELFARE WEIGHTS

Key sufficient statistics in optimal tax formulas are **Social Marginal Welfare Weights** for each individual:

Social Marginal Welfare Weight on individual i is $g_i = G'(u^i)u_c^i / \lambda$ (λ multiplier of govt budget constraint) measures \$ value for govt of giving \$1 extra to person i

No income effects $\Rightarrow \int_i g_i = 1$: giving \$1 to all costs \$1 (population has measure 1) and increase SWF (in \$ terms) by $\int_i g_i$

g_i typically depend on tax system (endogenous variable)

Utilitarian case: g_i decreases with z_i due to decreasing marginal utility of consumption

Rawlsian case: g_i concentrated on most disadvantaged (typically those with $z_i = 0$)

OPTIMAL LINEAR TAX RATE: INDIVIDUAL PROBLEM

Disposable income (consumption): $c = (1 - \tau) \cdot z + R$ with τ linear tax rate and R demogrant funded by taxes τZ with Z aggregate earnings

Population of size one (continuum) with heterogeneous preferences $u^i(c, z)$ [differences in earnings ability are built in utility function]

Individual i chooses z to maximize $u^i((1 - \tau) \cdot z + R, z)$ labor supply choices $z^i(1 - \tau, R)$ aggregate to economy wide earnings $Z(1 - \tau) = \int_i z^i$ (are a function of the net-of-tax-rate).

Tax Revenue $R(\tau) = \tau \cdot Z(1 - \tau)$ is inversely U-shaped with τ :
 $R(\tau = 0) = 0$ (no taxes) and $R(\tau = 1) = 0$ (nobody works): called the Laffer Curve

OPTIMAL LINEAR TAX RATES: PLAN

Let's look at:

- 1) The optimal linear tax formula (on all income $z \in [0, \infty)$).
- 2) The revenue-maximizing rate (special case).
- 3) The top revenue-maximizing tax rate (nests previous case if top bracket starts at $z = 0$).

OPTIMAL LINEAR TAX RATE: FORMULA

Government chooses τ to maximize

$$\int_i G[u^i((1-\tau)z^i + \tau Z(1-\tau), z^i)]$$

Govt FOC (using the envelope theorem as z^i maximizes u^i):

$$0 = \int_i G'(u^i)u_c^i \cdot \left[-z^i + Z - \tau \frac{dZ}{d(1-\tau)} \right],$$

$$0 = \int_i G'(u^i)u_c^i \cdot \left[(Z - z^i) - \frac{\tau}{1-\tau} eZ \right],$$

First term $(Z - z^i)$ is mechanical redistributive effect of $d\tau$, second term is efficiency cost due to behavioral response of Z

\Rightarrow we obtain the following optimal linear income tax formula

$$\tau = \frac{1 - \bar{g}}{1 - \bar{g} + e} \quad \text{with} \quad \bar{g} = \frac{\int g_i \cdot z_i}{Z \cdot \int g_i}, \quad g_i = G'(u^i)u_c^i$$

OPTIMAL LINEAR TAX RATE: FORMULA

$$\tau = \frac{1 - \bar{g}}{1 - \bar{g} + e} \quad \text{with} \quad \bar{g} = \frac{\int g_i \cdot z_i}{Z \cdot \int g_i}, \quad g_i = G'(u^i) u_c^i$$

$0 \leq \bar{g} < 1$ if g_i is decreasing with z_i (social marginal welfare weights fall with z_i).

\bar{g} low when (a) inequality is high, (b) $g^i \downarrow$ sharply with z^i

Formula captures the equity-efficiency trade-off **robustly** ($\tau \downarrow \bar{g}$, $\tau \downarrow e$)

Rawlsian case: $g_i \equiv 0$ for all $z_i > 0$ so $\bar{g} = 0$ and $\tau = 1/(1 + e)$

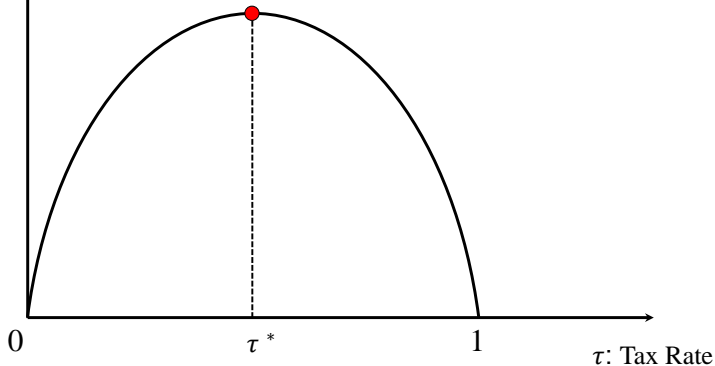
Rawlsian optimum = top of Laffer curve if $\min_i u^i$ agent earns $z_i = 0$.

Laffer Curve

Tax
Revenue
R

$$R = \tau \cdot Z(1 - \tau)$$

$$\tau^* = \frac{1}{1 + e} \text{ with } e = \frac{1 - \tau}{Z} \cdot \frac{dZ}{d(1 - \tau)}$$



OPTIMAL TOP INCOME TAX RATE (SAEZ '01)

Consider constant MTR τ above fixed z^* . Goal is to derive optimal τ

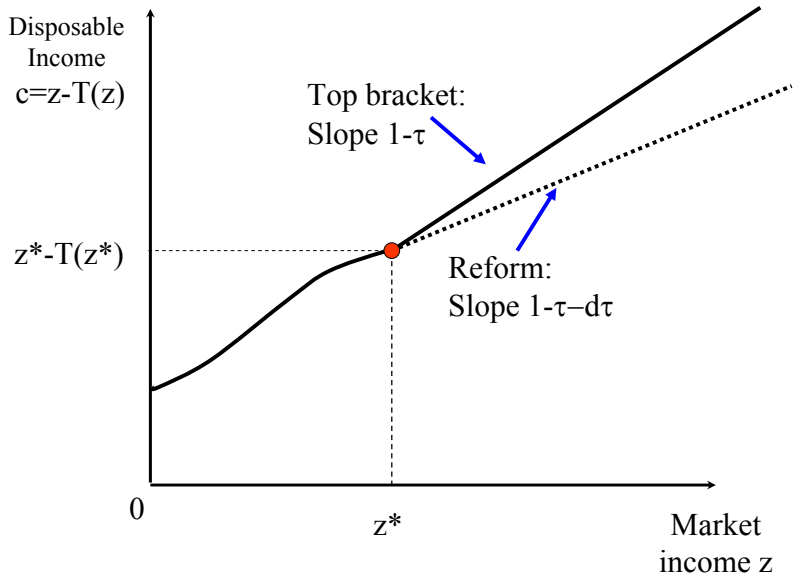
Assume w.l.o.g there is a continuum of measure one of individuals above z^*

Let $z(1 - \tau)$ be their average income [depends on net-of-tax rate $1 - \tau$], with elasticity $e = [(1 - \tau)/z] \cdot dz/d(1 - \tau)$

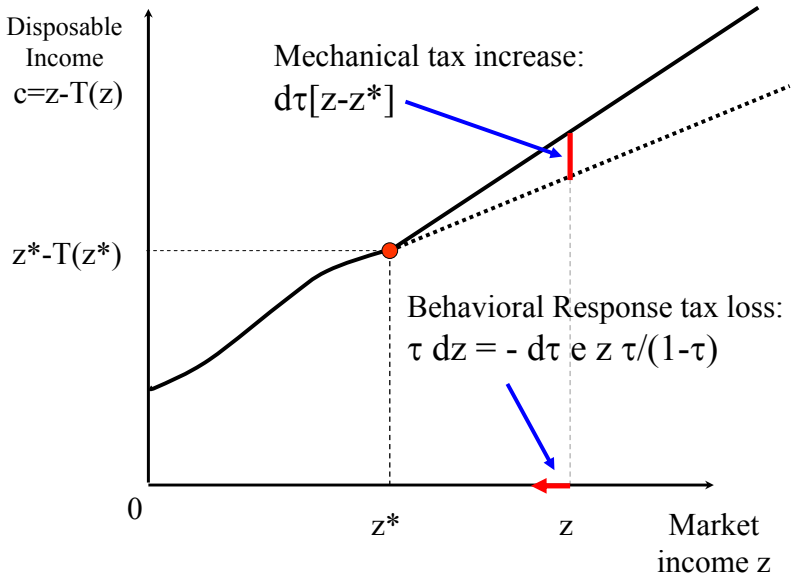
! Careful, what is e ?

Note that e is a mix of income and substitution effects (see Saez '01)

Optimal Top Income Tax Rate (Mirrlees '71 model)



Optimal Top Income Tax Rate (Mirrlees '71 model)



OPTIMAL TOP INCOME TAX RATE

Consider small $d\tau > 0$ reform above z^* .

1) **Mechanical increase** in tax revenue:

$$dM = [z - z^*]d\tau$$

2) **Welfare effect**:

$$dW = -\bar{g}dM = -\bar{g}[z - z^*]d\tau$$

where \bar{g} is the social marginal welfare weight for top earners

3) **Behavioral response** reduces tax revenue:

$$\begin{aligned} dB &= \tau \cdot dz = -\tau \frac{dz}{d(1-\tau)} d\tau = -\frac{\tau}{1-\tau} \cdot \frac{1-\tau}{z} \frac{dz}{d(1-\tau)} \cdot zd\tau \\ &\Rightarrow dB = -\frac{\tau}{1-\tau} \cdot e \cdot zd\tau \end{aligned}$$

OPTIMAL TOP INCOME TAX RATE

$$dM + dW + dB = d\tau \left[(1 - \bar{g})[z - z^*] - e \frac{\tau}{1 - \tau} z \right]$$

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

$$\frac{\tau}{1 - \tau} = \frac{(1 - \bar{g})[z - z^*]}{e \cdot z}$$

$$\tau = \frac{1 - \bar{g}}{1 - \bar{g} + a \cdot e} \quad \text{with} \quad a = \frac{z}{z - z^*}$$

Optimal $\tau \downarrow \bar{g}$ [redistributive tastes]

Optimal $\tau \downarrow$ with e [efficiency]

Optimal $\tau \downarrow a$ [thinness of top tail]

OPTIMAL LINEAR RATES: RECAP

1) The optimal linear tax formula (on all income $z \in [0, \infty)$):

$$\tau^* = \frac{1 - \bar{g}}{1 - \bar{g} + e} \quad \text{with} \quad \bar{g} = \frac{\int g_i \cdot z_i}{Z \cdot \int g_i}, \quad g_i = G'(u^i) u_c^i$$

2) The revenue-maximizing rate (special case if $\bar{g} = 0$, i.e., if $g_i = 0$ for all $z_i \neq 0$).

$$\tau^R = \frac{1}{1 + e}$$

3) The top revenue-maximizing tax rate (equal to τ^* if $z^* = 0$).

$$\tau^{top} = \frac{1 - \bar{g}}{1 - \bar{g} + a \cdot e} \quad \text{with} \quad a = \frac{z}{z - z^*}$$

SUFFICIENT STATS FORMULA

Pause for a bit: did we say anything about underlying characteristics of people?

Note how general the formula is!

Sufficient statistics, observables only.

ZERO TOP RATE RESULT

Suppose top earner earns z^T

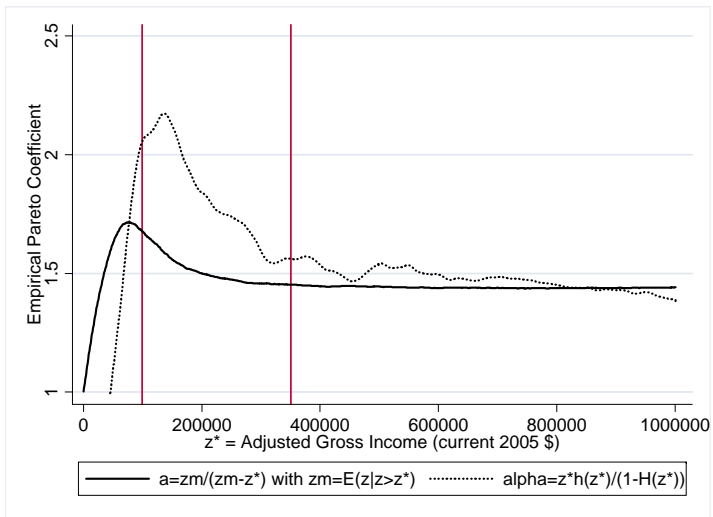
When $z^* \rightarrow z^T \Rightarrow z \rightarrow z^T$

$$dM = d\tau[z - z^*] \ll dB = d\tau \cdot e \cdot \frac{\tau}{1 - \tau} z \quad \text{when } z^* \rightarrow z^T$$

Intuition: extra tax applies only to earnings above z^* but behavioral response applies to full $z \Rightarrow$

Optimal τ should be zero when z^* close to z^T (Sadka-Seade zero top rate result) but result applies only to top earner

Top is uncertain: If actual distribution is finite draw from an underlying Pareto distribution then expected revenue maximizing rate is $1/(1 + a \cdot e)$ (Diamond and Saez JEP'11)



Source: Diamond and Saez JEP'11

OPTIMAL TOP INCOME TAX RATE

Empirically: $a = z/(z - z^*)$ very stable above $z^* = \$400K$

Pareto distribution $1 - F(z) = (k/z)^\alpha$, $f(z) = \alpha \cdot k^\alpha / z^{1+\alpha}$, with α Pareto parameter

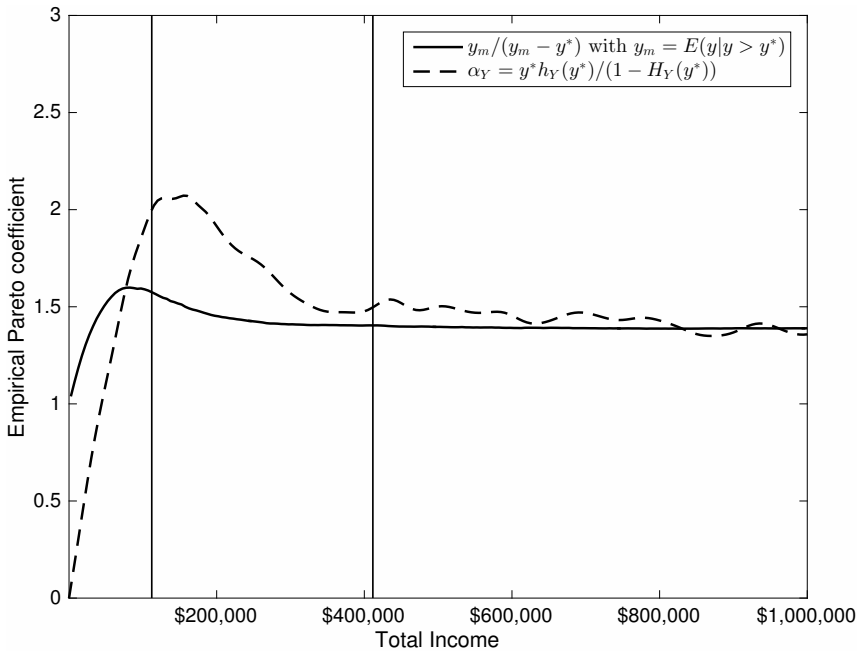
$$z(z^*) = \frac{\int_{z^*}^{\infty} sf(s) ds}{\int_{z^*}^{\infty} f(s) ds} = \frac{\int_{z^*}^{\infty} s^{-\alpha} ds}{\int_{z^*}^{\infty} s^{-\alpha-1} ds} = \frac{\alpha}{\alpha - 1} \cdot z^*$$

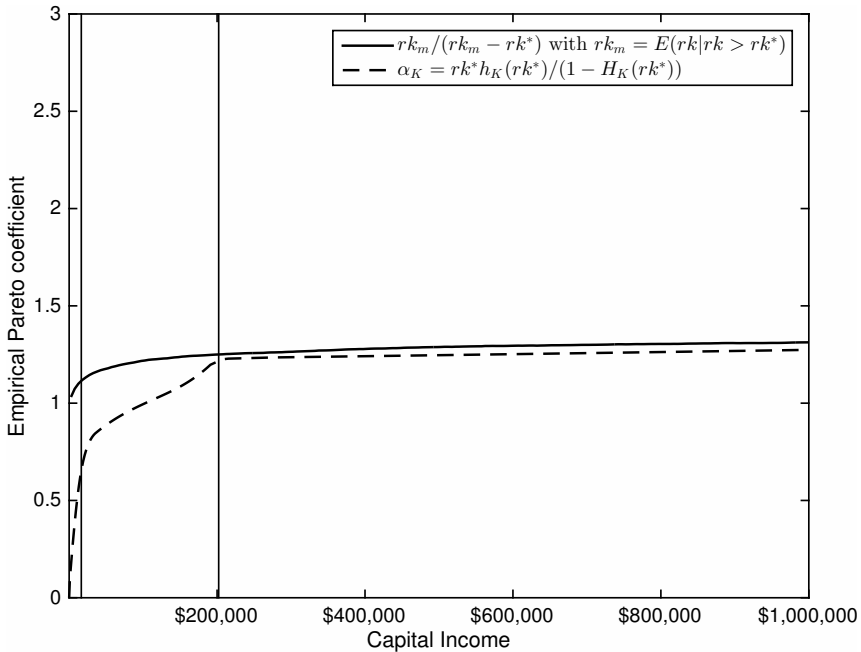
$\alpha = z/(z - z^*) = a$ measures *thinness* of top tail of the distribution

Empirically $a \in (1.5, 3)$, US has $a = 1.5$, Denmark has $a = 3$

$$\tau = \frac{1 - \bar{g}}{1 - \bar{g} + a \cdot e}$$

Only difficult parameter to estimate is e





TOP TAX REVENUE MAXIMIZING TAX RATE

Utilitarian criterion with $u_c \rightarrow 0$ when $c \rightarrow \infty \Rightarrow \bar{g} \rightarrow 0$ when $z^* \rightarrow \infty$

Rawlsian criterion (maximize utility of worst off person) $\Rightarrow \bar{g} = 0$ for any $z^* > \min(z)$

In the end, \bar{g} reflects the value that society puts on marginal consumption of the rich

$\bar{g} = 0 \Rightarrow$ Tax Revenue Maximizing Rate $\tau = 1/(1 + a \cdot e)$ (upper bound on top tax rate)

Example: $a = 2$ and $e = 0.25 \Rightarrow \tau = 2/3 = 66.7\%$

Laffer linear rate is a special case with $z^* = 0$, $z^m/z^* = \infty = a/(a - 1)$ and hence $a = 1$, $\tau = 1/(1 + e)$

EXTENSIONS AND LIMITATIONS

- 1) Model includes only intensive earnings response. Extensive earnings responses [entrepreneurship decisions, migration decisions] \Rightarrow Formulas can be modified
 - 2) Model does not include **fiscal externalities**: part of the response to $d\tau$ comes from **income shifting** which affects other taxes \Rightarrow Formulas can be modified
 - 3) Model does not include **classical externalities**: (a) charitable contributions, (b) positive spillovers (trickle down) [top earners underpaid], (c) negative spillovers [top earners overpaid]
- Classical general equilibrium effects on prices are NOT externalities and do not affect formulas [Diamond-Mirrlees AER '71, Saez JpubE '04]

GENERAL NON-LINEAR INCOME TAX $T(z)$

(1) Lumpsum grant given to everybody equal to $-T(0)$

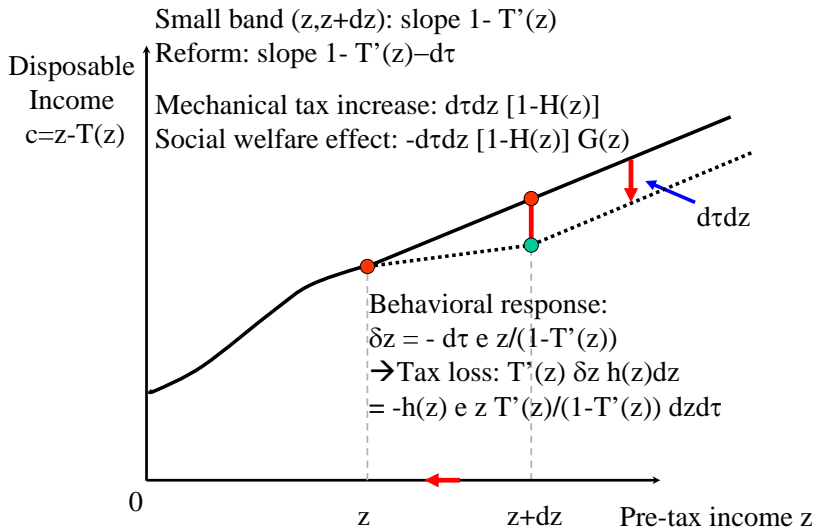
(2) Marginal tax rate schedule $T'(z)$ describing how (a) lump-sum grant is taxed away, (b) how tax liability increases with income

Let $H(z)$ be the income CDF [population normalized to 1] and $h(z)$ its density [endogenous to $T(\cdot)$]

Let $g(z)$ be the social marginal value of consumption for taxpayers with income z in terms of public funds [formally $g(z) = G'(u) \cdot u_c / \lambda$]: no income effects $\Rightarrow \int g(z)h(z)dz = 1$

Redistribution valued $\Rightarrow g(z)$ decreases with z

Let $G(z)$ the *average* social marginal value of c for taxpayers with income above z [$G(z) = \int_z^\infty g(s)h(s)ds / (1 - H(z))$]



GENERAL NON-LINEAR INCOME TAX

Assume away income effects $\varepsilon^c = \varepsilon^u = e$ [Diamond AER'98 shows this is the key theoretical simplification]

Consider small reform: increase T' by $d\tau$ in small band z and $z + dz$

Mechanical effect $dM = dzd\tau[1 - H(z)]$

Welfare effect $dW = -dzd\tau[1 - H(z)]G(z)$

Behavioral effect: substitution effect δz inside small band $[z, z + dz]$:
$$dB = h(z)dz \cdot T' \cdot \delta z = -h(z)dz \cdot T' \cdot d\tau \cdot z \cdot e_{(z)} / (1 - T')$$

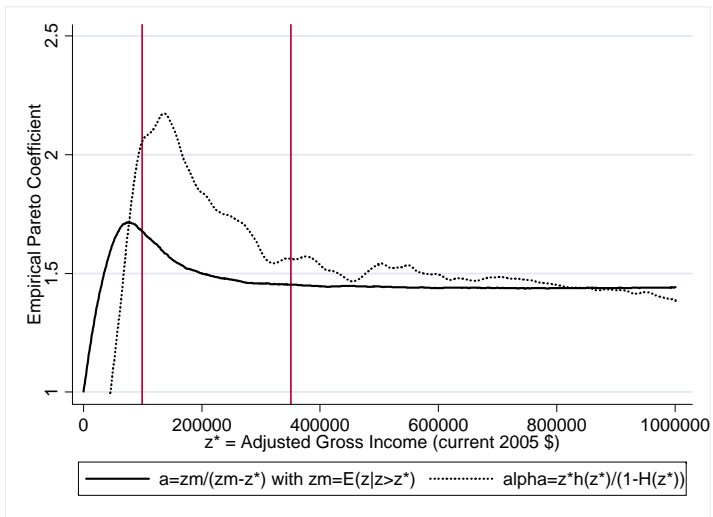
Optimum $dM + dW + dB = 0$

GENERAL NON-LINEAR INCOME TAX

$$T'(z) = \frac{1 - G(z)}{1 - G(z) + \alpha(z) \cdot e_{(z)}}$$

- 1) $T'(z)$ decreases with $e_{(z)}$ (elasticity efficiency effects)
- 2) $T'(z)$ decreases with $\alpha(z) = (zh(z))/(1 - H(z))$ (local Pareto parameter)
- 3) $T'(z)$ decreases with $G(z)$ (redistributive tastes)

Asymptotics: $G(z) \rightarrow \bar{g}$, $\alpha(z) \rightarrow a$, $e_{(z)} \rightarrow e \Rightarrow$ Recover top rate formula
 $\tau = (1 - \bar{g}) / (1 - \bar{g} + a \cdot e)$



Source: Diamond and Saez JEP'11

Negative Marginal Tax Rates Never Optimal

Suppose $T' < 0$ in band $[z, z + dz]$

Increase T' by $d\tau > 0$ in band $[z, z + dz]$: $dM + dW > 0$ and $dB > 0$
because $T'(z) < 0$

⇒ Desirable reform

⇒ $T'(z) < 0$ cannot be optimal

EITC schemes are not desirable in Mirrlees '71 model

MIRRLEES MODEL

The difference to before: we need to specify the *structural primitives*.

Key simplification is the lack of income effects (Diamond, 1998). We look into income effects next time.

Individual utility: $c - v(l)$, l is labor supply.

Skill n is exogenously given, equal to marginal productivity. Earnings are $z = nl$.

Density is $f(n)$ and CDF $F(n)$ on $[0, \infty)$.

Entry into contract theory/mechanism design here: The government does not observe **skill**. Tax is based on income z , $T(z)$.

What happens if we had a tax $T(n)$ available?

Why did we not talk about this in the earlier derivations? Did we ignore the incentive compatibility constraints?

Elasticity of labor to taxes

Recall we derive elasticities on the linearized budget set. If marginal tax rate is τ , labor supply is: $l = l(n(1 - \tau))$. Why the $n(1 - \tau)$? Why *only* $n(1 - \tau)$?

FOC of the agent for labor supply:

$$n(1 - \tau) = v'(l)$$

Totally differentiate this (key thing: skill is fixed!)

$$d(n(1 - \tau)) = v''(l)dl$$

$$\Rightarrow e = \frac{dl}{d(n(1 - \tau))} \frac{(1 - \tau)n}{l} = \frac{(1 - \tau)n}{lv''(l)} = \frac{v'(l)}{lv''(l)}$$

Is this compensated? uncompensated?

Direct Revelation Mechanism and Incentive Compatibility

We want to max social welfare and have exogenous revenue requirement (non transfer-related E).

We imagine a direct revelation mechanism. Every agent comes to government, reports a type n' . We assign allocations as a function of the report. $c(n')$, $z(n')$, $u(n')$. Why are we not assigning labor $l(n')$?

What are the constraints in this problem?

Feasibility (net resources sum to zero): $\int_n c_n f(n) dn \geq n l_n f(n) dn - E$.

Incentive compatibility:

Direct Revelation Mechanism and Incentive Compatibility

We want to max social welfare and have exogenous revenue requirement (non transfer-related E).

We imagine a direct revelation mechanism. Every agent comes to government, reports a type n' . We assign allocations as a function of the report. $c(n')$, $z(n')$, $u(n')$. Why are we not assigning labor $l(n')$?

What are the constraints in this problem?

Feasibility (net resources sum to zero): $\int_n c_n f(n) dn \geq n l_n f(n) dn - E$.

Incentive compatibility:

$$c(n) - v \left(\frac{z(n)}{n} \right) \geq c(n') - v \left(\frac{z(n')}{n} \right) \quad \forall n, n'$$

That's a lot of constraints!

Envelope Theorem and First order Approach

Replace the infinity of constraints with agents' first-order condition. If we take derivative of utility wrt type n at *truth-telling*

$$\frac{du_n}{dn} = \left(c'(n) - \frac{z'(n)}{n} v' \left(\frac{z(n)}{n} \right) \right) \frac{dn'}{dn} + \frac{z(n)}{n^2} v' \left(\frac{z(n)}{n} \right)$$

What if report is optimally chosen?

Envelope condition:

$$\frac{du_n}{dn} = \frac{l_n v'(l_n)}{n}$$

Will replace infinity of constraints.

Is necessary, but what about sufficiency?

Full Optimization Program

$$\max_{c_n, u_n, z_n} \int_n G(u_n) f(n) dn \quad \text{s.t.} \quad \int_n c_n f(n) dn \leq \int_n n l_n f(n) dn - E$$

$$\text{and s.t. } \frac{du_n}{dn} = \frac{l_n v'(l_n)}{n}$$

State variable: u_n .

Control variables: l_n , with $c_n = u_n + v(l_n)$.

Why am I suddenly saying l_n is a control?

Use optimal control.

Hamiltonian and Optimal Control

The Hamiltonian is:

$$H = [G(u_n) + p \cdot (nl_n - u_n - v(l_n))]f(n) + \phi(n) \cdot \frac{l_n v'(l_n)}{n}$$

p : multiplier on the resource constraint.

$\phi(n)$: multiplier on the envelope condition ("costate"). Depends on n !

FOCs:

$$\frac{\partial H}{\partial l_n} = p \cdot [n - v'(l_n)]f(n) + \frac{\phi(n)}{n} \cdot [v'(l_n) + l_n v''(l_n)] = 0$$

$$\frac{\partial H}{\partial u_n} = [G'(u_n) - p]f(n) = -\frac{d\phi(n)}{dn}$$

Transversality: $\lim_{n \rightarrow \infty} \phi(n) = 0$ and $\phi(0) = 0$.

Rearranging the FOCs

Take the integral of the FOC wrt u_n to solve for $\phi(n)$:

$$-\phi(n) = \int_n^{\infty} [p - G'(u_m)] f(m) dm$$

Integrate this same FOC over the full space, using transversality conditions:

$$p = \int_0^{\infty} G'(u_n) f(m) dm$$

What does this say?

How can we make the tax rate appear? Use the agent's FOC.

$$n - v'(l_n) = nT'(z_n)$$

Obtaining the Optimal Tax Formula

Recall that $e = \frac{(1 - T'(z_n))n}{l v''(l)}$

Rearranging the last term in the FOC for l_n :

$$[v'(l_n) + l_n v''(l_n)]/n = [1 - T'(z_n)][1 + 1/e]$$

Let $g_m \equiv G'(u_m)/p$ be the marginal social welfare weight on type m .

Then, the FOC for l_n becomes:

$$\frac{T'(z_n)}{1 - T'(z_n)} = \left(1 + \frac{1}{e}\right) \cdot \left(\frac{\int_n^\infty (1 - g_m) dF(m)}{nf(n)}\right)$$

This is the Diamond (1998) formula.

What is different from the previously derived formula *à la* Saez (2001)?

Let's go from types to observable income

How do we go from type distribution to income distribution?

Under linearized tax schedule, earnings are a function $z_n = nl(n(1 - \tau))$.

How do earnings vary with type?

$$\frac{dz_n}{dn} = l + (1 - \tau)n \frac{dl}{d(m(1 - \tau))} = l_n \cdot (1 + e)$$

(intuition?)

Let $h(z)$ be the density of earnings, with CDF $H(z)$. The following relation must hold:

$$h(z_n)dz_n = f(n)dn$$

$$f(n) = h(z_n)l_n(1 + e) \Rightarrow ng(n) = z_nh(z_n)(1 + e)$$

Let's substitute income distributions for type distributions in the formula. 44/11

Optimal Tax Formula with No Income Effects

$$\begin{aligned}\frac{T'(z_n)}{1 - T'(z_n)} &= \left(1 + \frac{1}{e_{(z_n)}}\right) \left(\frac{\int_n^\infty (1 - g_m) dF(m)}{nf(n)}\right) \quad (\text{primitives}) \\ &= \frac{1}{e_{(z_n)}} \left(\frac{1 - H(z_n)}{z_n h(z_n)}\right) \cdot (1 - G(z_n)) \quad (\text{incomes})\end{aligned}$$

where:

$$G(z_n) = \frac{\int_n^\infty g_m dF(m)}{1 - F(n)} = \frac{\int_{z_n}^\infty g_m dH(z_m)}{1 - H(z_n)}$$

is the average marginal social welfare weight on individuals with income above z_n (change of variables to income distributions in last equality).

Rearrange, use definition of Pareto parameter $\alpha(z) = (zh(z))/(1 - H(z))$ to get same formula as before:

$$T'(z) = \frac{1 - G(z)}{1 - G(z) + \alpha(z) \cdot e_{(z)}}$$

Recap:

“Mechanism design approach” requires you to specify *primitives* (utility function, uni-dimensional heterogeneity) as done in Mirrlees (1971).

“Sufficient stats approach” captures arbitrary heterogeneity conditional on z as long as well-behaved elasticities.

Yield same formula if can make the link between types and income distributions.

$$\begin{aligned}\frac{T'(z_n)}{1 - T'(z_n)} &= \left(1 + \frac{1}{e_{(z_n)}}\right) \left(\frac{\int_n^\infty (1 - g_m) dF(m)}{nf(n)}\right) \quad (\text{primitives}) \\ &= \frac{1}{e_{(z_n)}} \left(\frac{1 - H(z_n)}{z_n h(z_n)}\right) \cdot (1 - G(z_n)) \quad (\text{incomes})\end{aligned}$$

NUMERICAL SIMULATIONS

$H(z)$ [and also $G(z)$] endogenous to $T(\cdot)$. Calibration method (Saez Restud '01):

Specify utility function (e.g. constant elasticity):

$$u(c, z) = c - \frac{1}{1 + \frac{1}{e}} \cdot \left(\frac{z}{n}\right)^{1 + \frac{1}{e}}$$

Individual FOC $\Rightarrow z = n^{1+e}(1 - T')^e$

Calibrate the exogenous skill distribution $F(n)$ so that, using **actual** $T'(\cdot)$, you recover **empirical** $H(z)$

Use Mirrlees '71 tax formula (expressed in terms of $F(n)$) to obtain the optimal tax rate schedule T' .

NUMERICAL SIMULATIONS

$$\frac{T'(z(n))}{1 - T'(z(n))} = \left(1 + \frac{1}{e}\right) \left(\frac{1}{nf(n)}\right) \int_n^\infty \left[1 - \frac{G'(u(m))}{\lambda}\right] f(m) dm,$$

Iterative Fixed Point method: start with T'_0 , compute $z^0(n)$ using individual FOC, get $T^0(0)$ using govt budget, compute $u^0(n)$, get λ using $\lambda = \int G'(u)f$, use formula to estimate T'_1 , iterate till convergence

Fast and effective method (Brewer-Saez-Shepard '10)

NUMERICAL SIMULATION RESULTS

$$T'(z) = \frac{1 - G(z)}{1 - G(z) + \alpha(z) \cdot e_{(z)}}$$

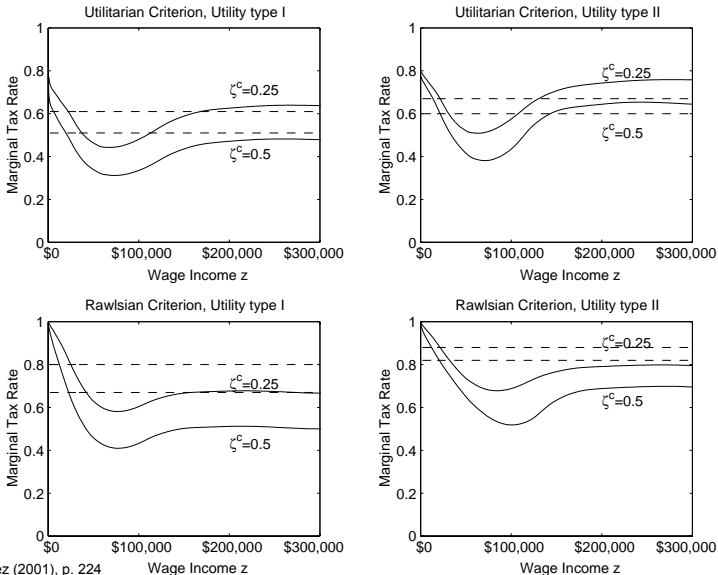
Take utility function with e constant

2) $\alpha(z) = (zh(z))/(1 - H(z))$ is inversely U-shaped empirically

3) $1 - G(z)$ increases with z from 0 to 1 ($\bar{g} = 0$)

⇒ Numerical optimal $T'(z)$ is U-shaped with z : reverse of the general results $T' = 0$ at top and bottom [Diamond AER'98 gives theoretical conditions to get U-shape]

FIGURE 5 – Optimal Tax Simulations



Source: Saez (2001), p. 224

EXTENSION 1: MIGRATION EFFECTS

Tax rates may affect migration (evidence on this next time).

Migration issues may be particularly important at the top end (brain drain).

Some theory papers (Mirrlees '82, Lehmann-Simula QJE'14). Here:
Simplified Mirrlees (1982) model.

Earnings z are fixed, conditional on residence.

$P(c|z)$ is number of residents earning z when disposable income is c , with
 $c = z - T(z)$.

Consider small tax reform $dT(z)$ for those earning z .

What is migration responding to? Marginal taxes?

ELASTICITY OF MIGRATION TO TAXES

Mechanical effect net of welfare is: $M + W = (1 - g(z))P(c|z)dT$.

Why? Where is utility effect of changing country induced by taxes?

Migration responds to average taxes (or total taxes, since income fixed).

$$\eta_m(z) = \frac{\partial P(c|z)}{\partial c} \frac{z - T(z)}{P(c|z)}$$

Fiscal cost of raising taxes by $dT(z)$ is: $B = -\frac{T(z)}{z - T(z)} \cdot P(c|z) \cdot \eta_m$

Optimal tax is where $M + W + B = 0$:

$$\frac{T(z)}{z - T(z)} = \frac{1}{\eta_m(z)} \cdot (1 - g(z))$$

What determines the elasticity $\eta_m(z)$?

MIGRATION EFFECTS IN THE STANDARD MODEL

$\eta_m(z)$ depends on size of jurisdiction: large for cities, zero worldwide \Rightarrow
(1) Redistribution easier in large jurisdictions, (2) Tax coordination across countries increases ability to redistribute (big issue currently in EU), (3) visa system, cost of migration, ...

Top revenue maximizing tax rate formula (Brewer-Saez-Shepard '10):

$$\tau = \frac{1}{1 + a \cdot e + \bar{\eta}^m}$$

where $\bar{\eta}_m$ is the elasticity of top earners to disposable income.

EXTENSION 2: RENT SEEKING EFFECTS

Pay may not be equal to the marginal economic product for top income earners. Why? Overpaid or underpaid?

Piketty, Saez, and Stantcheva (2014) "A Tale of Three Elasticities."

Actual output is y , but individual only receives share η of actual output. To increase either productive effort or rent-seeking, effort is required.

$$u^i(c, \eta, y) = c - h_i(y) - k_i(\eta)$$

Define bargained earnings: $b = (\eta - 1)y$.

Average bargaining is $E(b)$, extracted equally from everyone else (good assumption?) Means $E(b)$ can be perfectly canceled by $-T(0)$.

RENT SEEKING ELASTICITIES

Given tax, individual maximizes:

$$u^i(c, y, \eta) = \eta \cdot y - T(\eta \cdot y) - h_i(y) - k_i(\eta)$$

What will y_i and η_i depend on?

Average reported income, productive income and bargained earnings in the top bracket:

$$z(1 - \tau), \quad y(1 - \tau), \quad \eta(1 - \tau)$$

Total compensation elasticity e : $e = \frac{1-\tau}{z} \frac{dz}{d(1-\tau)}$ (what is it driven by?)

Real labor supply elasticity e_y : $e_y = \frac{1-\tau}{y} \frac{dy}{d(1-\tau)} \geq 0$.

Thus the bargaining elasticity component $e_b = \frac{db}{d(1-\tau)} \frac{1-\tau}{z} = s \cdot e$ with

$$s = \frac{db/d(1-\tau)}{dz/d(1-\tau)}$$

s and e_b positive if $\eta > 1$.

OPTIMAL TAX RATE WITH RENT SEEKING

Suppose rent-seeking only at the top, $E(b) = qb(1 - \tau)$ where q fraction of top earners.

Government maximizes tax revenues from top bracket earners:

$$T = \tau[y(1 - \tau) + b(1 - \tau) - z^*]q - E(b)$$

Why does $E(b)$ enter?

$$\tau^* = \frac{1 + a \cdot e_b}{1 + a \cdot e} = 1 - \frac{a(y/z)e_y}{1 + a \cdot e}$$

How does τ^* change with e , e_y , and e_b ? When is $\tau^* = 1$ optimal?

Trickle up vs trickle down: what happens to τ^* when top earners are overpaid? Underpaid?

How would you measure e_b (even b itself?)

OPTIMAL NON-LINEAR TAX WITH INCOME EFFECTS

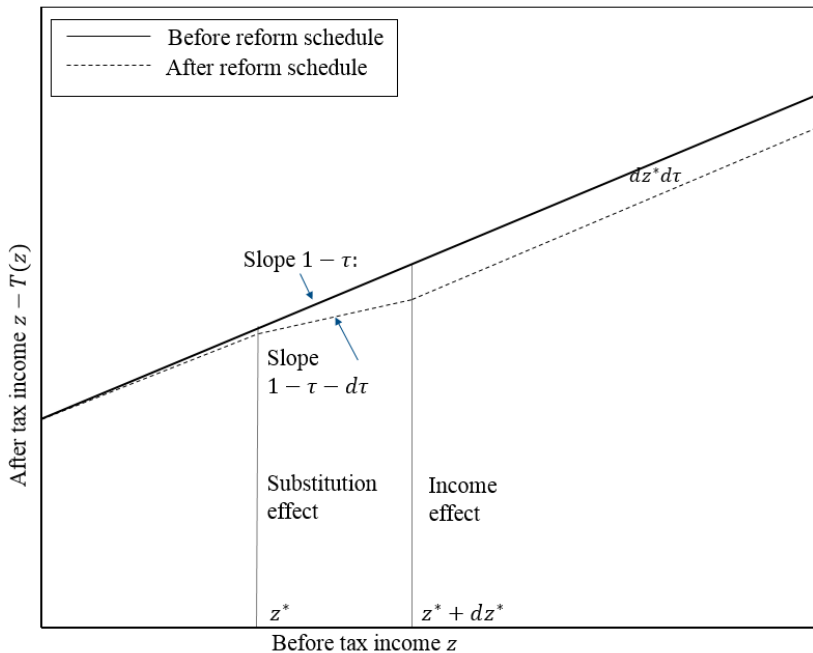
Consider effect of small reform where marginal tax rates increased by $d\tau$ in $[z^*, z^* + dz^*]$.

What are the effects on tax receipts?

Mechanical effect net of welfare loss, M:

Every tax payer with income z above z^* pays additional $d\tau dz^*$, valued at $(1 - g(z))d\tau dz^*$.

$$M = d\tau dz^* \int_{z^*}^{\infty} (1 - g(z))h(z)dz$$



BEHAVIORAL EFFECT PART 1: SUBSTITUTION

In $[z^*, z^* + dz^*]$, income changes by dz .

Marginal tax rate changes *directly* by $d\tau$, but also additionally *indirectly* by $dT'(z) = T''(z)dz$. Why? When is this not the case?

$$dz = -\varepsilon_{(z)}^c z^* \frac{d\tau + dT'(z)}{1 - T'(z)} \Rightarrow dz = -\varepsilon_{(z)}^c z^* \frac{d\tau}{1 - T'(z) + \varepsilon_{(z)}^c z^* T''(z)}$$

Define the *virtual density*: density that would occur at z if tax schedule replaced by linearized tax schedule. What is the linearized schedule (τ, R) such that income is $(1 - \tau)z + R$?

$$\frac{h^*(z)}{1 - T'(z)} = \frac{h(z)}{1 - T'(z) + \varepsilon_{(z)}^c z^* T''(z)}$$

BEHAVIORAL EFFECT PART 1: SUBSTITUTION

Overall elasticity/substitution effect is then:

$$E = -\varepsilon_{(z)}^c z^* \frac{T'(z)}{1 - T'(z)} h^*(z^*) d\tau dz^*$$

Can derive expression without taking into account endogenous (indirect) change in marginal tax rates if use the virtual density instead of true one.

BEHAVIORAL EFFECT PART 2: INCOME EFFECT

Taxpayers with income above z^* pay $-dR = d\tau dz^*$ additional taxes. Their change in income is:

$$dz = -\varepsilon_{(z)}^c z \frac{T'' dz}{1 - T'} - \eta \frac{d\tau dz^*}{1 - T'(z)} \Rightarrow dz = -\eta \frac{d\tau dz^*}{1 - T'(z) + z \varepsilon_{(z)}^c T''(z)}$$

Why?

Total income effect response:

$$I = d\tau dz^* \int_{z^*}^{\infty} -\eta_{(z)} \frac{T'(z)}{1 - T'(z)} h^*(z) dz$$

At the optimum: $M + E + I = 0$.

PUTTING THE EFFECTS TOGETHER

$$\frac{T'(z)}{1 - T'(z)} = \frac{1}{\varepsilon_{(z)}^c} \left(\frac{1 - H(z^*)}{z^* h^*(z^*)} \right) \\ \times \left[\int_{z^*}^{\infty} (1 - g(z)) \frac{h(z)}{1 - H(z^*)} dz + \int_{z^*}^{\infty} -\eta \frac{T'(z)}{1 - T'(z)} \frac{h^*(z)}{1 - H(z^*)} dz \right]$$

First-order differential equation. See Saez (2001) Appendix for solution (is standard).

Change of variable from z to n ?

Recall with a linear tax: $\frac{\dot{z}_n}{z_n} = \frac{1 + \varepsilon_{(z_n)}^u}{n}$.

What happens with nonlinear tax? See Saez (2001) Appendix for derivation.

$$\frac{\dot{z}_n}{z_n} = \frac{1 + \varepsilon_{(z_n)}^u}{n} - \dot{z}_n \frac{T''(z_n)}{1 - T'(z_n)} \varepsilon_{z(n)}^c$$

ATKINSON-STIGLITZ THEOREM

Famous Atkinson-Stiglitz JpubE' 76 shows that

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

(i.e. commodity taxes not useful) under two assumptions on utility functions $u^h(c_1, \dots, c_K, z)$

- 1) Weak separability between (c_1, \dots, c_K) and z in utility
- 2) Homogeneity across individuals in the sub-utility of consumption $v(c_1, \dots, c_K)$ [does not vary with h]

$$(1) \text{ and } (2): \quad u^h(c_1, \dots, c_K, z) = U^h(v(c_1, \dots, c_K), z)$$

Original proof was based on optimum conditions, new straightforward proof by Laroque EL '05, and Kaplow JpubE '06.

ATKINSON-STIGLITZ THEOREM PROOF

Let $V(y, p + t) = \max_c v(c_1, \dots, c_K)$ st $(p + t) \cdot c \leq y$ be the indirect utility of consumption c [common to all individuals]

Start with $(T(\cdot), t)$. Let $c(t)$ be consumer choice.

Replace $(T(\cdot), t)$ with $(\bar{T}(\cdot), t = 0)$ where $\bar{T}(z)$ such that $V(z - T(z), p + t) = V(z - \bar{T}(z), p) \Rightarrow$ Utility $U^h(V, z)$ and labor supply choices z unchanged for all individuals.

Attaining $V(z - \bar{T}(z), p)$ at price p costs at least $z - \bar{T}(z)$

Consumer also attains $V(z - \bar{T}(z), p) = V(z - T(z), p + t)$ when choosing $c(t) \Rightarrow z - \bar{T}(z) \leq p \cdot c(t) = z - T(z) - t \cdot c(t)$

$\Rightarrow \bar{T}(z) \geq T(z) + t \cdot c(t)$: the government collects more taxes with $(\bar{T}(\cdot), t = 0)$

ATKINSON-STIGLITZ INTUITION

With separability and homogeneity, conditional on earnings z , consumption choices $c = (c_1, \dots, c_K)$ do not provide any information on ability

⇒ Differentiated commodity taxes t_1, \dots, t_K create a tax distortion with no benefit ⇒ Better to do all the redistribution with the individual income tax

Note: With weaker linear income taxation tool (Diamond-Mirrlees AER '71, Diamond JpubE '75), need $v(c_1, \dots, c_K)$ homothetic (linear Engel curves, Deaton EMA '81) to obtain no commodity tax result

[Unless Engel curves are linear, commodity taxation can be useful to “non-linearize” the tax system]

Generalization of Atkinson-Stiglitz to Heterogeneous Tastes – Saez (2002)

Can we generalize AS to case with heterogeneous consumption preferences?

Individuals indexed by h , utility $U(c, z)$ with $c = (c_1, \dots, c_K)$.

Nonlinear income tax $T(z)$.

Pre-tax prices: p , post-tax price: $q = p + t$.

Budget constraint $q \cdot c \leq z - T(z)$.

Demands: $c^h(q, R, z)$, labor supply $z^h(q, T)$, indirect utility $v^h(q, R, z)$.

Suppose that $T(z)$ is optimally chosen at zero commodity taxation $p = q$ to max

$$W = \sum_h \alpha^h v^h(p, z^h - T(z^h), z^h) \quad \text{s.t.} \quad \sum_h T(z^h) \geq E$$

Marginal social welfare weight $g^h = \alpha^h v_R^h / \lambda$.

Can commodity taxation improve welfare?

Imagine dt_1 .

Mechanical revenue effect: $dM_1 = \sum_h c_1^h dt_1 = C_1 dt_1$.

Welfare effect (envelope theorem): $dU_1 = - \sum_h g^h c_1^h dt_1$.

Behavioral labor supply response: $dB_1 = - \sum_h T'(z^h) dz_{t_1}^h$ with $dz_{t_1}^h = dt_1 \frac{\partial z^h}{\partial q_1}$.

Why no behavioral response on revenue from changes in consumption?

If no commodity tax introduction can increase welfare, need

$$dW = dM_1 + dU_1 + dB_1 = 0.$$

Can commodity taxation improve welfare? (II)

Find a small income tax reform that “mimics” commodity tax change:

$$dT(z) = C_1(z)dt_1.$$

This reform has zero first order welfare impact, why?

Mechanical Revenue effect:

$$dM_T = \sum_h dT(z^h) = \sum_h C_1(z^h)dt_1 = C_1 dt_1 = dM_1 \text{ (why?)}$$

$$\text{Welfare effect: } dW_T = - \sum_h g^h C_1(z^h)dt_1$$

$$\text{Behavioral effect: } dB_T = - \sum_h T'(z^h)dz_T^h.$$

Subtract one from other (using that $dW_T = 0$).

Can commodity taxation improve welfare? (III)

$$\frac{dW}{dt_1} = \underbrace{-\sum_h g^h [c_1^h - C_1(z^h)]}_{\text{Pure Welfare Effect}} + \underbrace{\sum_h T'(z^h) \left[\frac{dz_T^h}{dT(z^h)} \cdot \frac{dT(z^h)}{dt_1} - \frac{dz_1^h}{dt_1} \right]}_{\text{Behavioral Effect}}$$

- Pure welfare effect is zero if conditional on z , g^h and c^h are uncorrelated.
- What does this mean? Is this reasonable? (recall weights are "generalized" since α^h depends on h directly).
- Young/old? medical expenses?
- Always satisfied in "standard AS" assumptions.

Behavioral Effects under Commodity and Income Taxation

Since $T'(z^h) \geq 0$ (remember), increasing $dt_1 > 0$ more efficient than equivalent income tax increase if labor supply increase from commodity tax change larger than that of income tax change.

When is this the case? Can show that:

$$E[dz_{t_1}^h] = -dt_1 \left(\mathbb{E} \left[\frac{z_c^h}{1 + T''(z)z_c^h} \frac{dc_1^h}{dz} \right] + \mathbb{E} \left[\frac{z_R^h}{1 + T''(z)z_c^h} c_1^h \right] \right)$$

$$E[dz_T^h] = -dt_1 \left(\mathbb{E} \left[\frac{z_c^h}{1 + T''(z)z_c^h} \right] \frac{dC_1(z)}{dz} + \mathbb{E} \left[\frac{z_R^h}{1 + T''(z)z_c^h} \right] C_1(z) \right)$$

Need $E[dz_{t_1}^h] = E[dz_T^h]$ for no commodity taxation.

Assumptions needed for behavioral effects to be the same under Commodity and Income Taxation

Assumption 2: Conditional on z , behavioral responses z_c^h and z_R^h independent of consumption patterns c_1^h and $\frac{dc_1^h}{dz}$.

Do you think this holds?

Assumption 3: For any income level, $E\left(\frac{dc_1^h}{dz} \mid z^h = z\right) = \frac{dC_1(z)}{dz}$.

This is the *key* assumption. What does it say? Why is this not mechanically true?

$\frac{dC_1(z)}{dz} = \lim_{dz \rightarrow 0} \frac{E(c_1^h \mid z^h = z + dz) - E(c_1^h \mid z^h = z)}{dz}$ is cross-sectional variation in consumption of good 1 when income changes.

What is $E\left(\frac{dc_1^h}{dz} \mid z^h = z\right)$?

Assumptions needed for behavioral effects to be the same under Commodity and Income Taxation

Imagine 2 groups:

Group A: $z^h = z$. consume $C_1(z)$ on average.

Group B: $z^h = z - dz$. Consumes on average $dc_1 = \frac{dC_1(z)}{dz}$ less of good 1.

Group A': Individuals from group A who are *forced* to reduce their income to $z - dz$. Reduce their consumption relative to group A by

$$dc'_1 = E \left(\frac{dc_1^h}{dz} \mid z^h = z \right) dz.$$

Assumption 3 says: Group B = Group A' for consumption of good 1.

Always true in AS since consumption only depends on after tax income with separability $C_1(z) = c_1^h(z)$.

Why would Group A' not have same consumption of good 1 as Group B?

WHEN ATKINSON-STIGLITZ ASSUMPTIONS FAIL

Thought experiment we just did was: force high earners to work less and earn only as much as low earners: if high earners consume more of good k than low earners, taxing good 1 is desirable.

1) High earners are “different” (since if left to choose, choose to work more. If they have a relatively higher/lower taste for good 1 (independently of income), tax more/less good 1 . [indirect tagging] Cigarettes? Fancy wine? How would you see this empirically?

2) High earners now have more leisure. If Good 1 positively related to leisure (consumption of 1 increases when leisure increases keeping after-tax income constant), tax it! [tax on holiday trips, subsidy on work related expenses such as child care]

In general Atkinson-Stiglitz assumption is a good starting place for most goods \Rightarrow Zero-rating on some goods under VAT for redistribution is inefficient and administratively burdensome [Mirrlees review]

REFERENCES (for lectures 2 and 3)

Akerlof, G. "The Economics of Tagging as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning", *American Economic Review*, Vol. 68, 1978, 8-19. (web)

Atkinson, A.B. and J. Stiglitz "The design of tax structure: Direct versus indirect taxation", *Journal of Public Economics*, Vol. 6, 1976, 55-75. (web)

Besley, T. and S. Coate "Workfare versus Welfare: Incentives Arguments for Work Requirements in Poverty-Alleviation Programs", *American Economic Review*, Vol. 82, 1992, 249-261. (web)

Boskin, M. and E. Sheshinski "Optimal tax treatment of the family: Married couples", *Journal of Public Economics*, Vol. 20, 1983, 281-297 (web)

Brewer, M., E. Saez, and A. Shephard "Means Testing and Tax Rates on Earnings", in *The Mirrlees Review: Reforming the Tax System for the 21st Century*, Oxford University Press, 2010. (web)

Cremer, H., F. Gahvari, and N. Ladooux "Externalities and optimal taxation", Journal of Public Economics, Vol. 70, 1998, 343-364. (web)

Deaton, A. "Optimal Taxes and the Structure of Preferences", Econometrica, Vol. 49, 1981, 1245-1260 (web)

Diamond, P. "A many-person Ramsey tax rule", Journal of Public Economics, Vol.4, 1975, 335-342. (web)

Diamond, P. "Income Taxation with Fixed Hours of Work" Journal of Public Economics, Vol. 13, 1980, 101-110. (web)

Diamond, P. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", American Economic Review, Vol. 88, 1998, 83-95. (web) Skim this for historical reasons.

Diamond, P. and E. Saez "From Basic Research to Policy Recommendations: The Case for a Progressive Tax", Journal of Economic Perspectives, 25(4), 2011, 165-190. (web)

Edgeworth, F. "The Pure Theory of Taxation", The Economic Journal, Vol. 7, 1897, 550-571. (web)

Kaplow, L. "On the undesirability of commodity taxation even when income taxation is not optimal", *Journal of Public Economics*, Vol. 90, 2006, 1235-1250. (web)

Kaplow, L. *The Theory of Taxation and Public Economics*. Princeton University Press, 2008.

Kaplow, L. and S. Shavell "Any Non-welfarist Method of Policy Assessment Violates the Pareto Principle," *Journal of Political Economy*, 109(2), (April 2001), 281-286 (web)

Kleven, H., C. Kreiner and E. Saez "The Optimal Income Taxation of Couples", *Econometrica*, Vol. 77, 2009, 537-560. (web)

Laroque, G. "Indirect Taxation is Superfluous under Separability and Taste Homogeneity: A Simple Proof", *Economic Letters*, Vol. 87, 2005, 141-144. (web)

Lehmann, E., L. Simula, A. Trannoy "Tax Me if You Can! Optimal Nonlinear Income Tax between Competing Governments," *Quarterly Journal of Economics* 129(4), 2014, 1995-2030. (web)

Mankiw, G. and M. Weinzierl "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution", *AEJ: Economic Policy*, Vol. 2, 2010, 155-176. (web)

Mirrlees, J. "An Exploration in the Theory of Optimal Income Taxation", *Review of Economic Studies*, Vol. 38, 1971, 175-208. (web) Read it for historical reasons, it is not very easy to follow.

Mirrlees, J. "Migration and Optimal Income Taxes", *Journal of Public Economics*, Vol. 18, 1982, 319-341. (web)

Nichols, A. and R. Zeckhauser "Targeting Transfers Through Restrictions on Recipients", *American Economic Review*, Vol. 72, 1982, 372-377. (web)

Piketty, Thomas and Emmanuel Saez "Optimal Labor Income Taxation," *Handbook of Public Economics*, Volume 5, Amsterdam: Elsevier-North Holland, 2013. (web)

Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities", *American Economic Journal: Economic Policy*, 6(1), 2014, 230-271 (web)

Sadka, E. "On Income Distribution, Incentives Effects and Optimal Income Taxation", *Review of Economic Studies*, Vol. 43, 1976, 261-268. (web)

Saez, E. "Using Elasticities to Derive Optimal Income Tax Rates", *Review of Economics Studies*, Vol. 68, 2001, 205-229. (web)

Saez, E. "The Desirability of Commodity Taxation under Non-linear Income Taxation and Heterogeneous Tastes", *Journal of Public Economics*, Vol. 83, 2002, 217-230. (web)

Sandmo, A. "Optimal Taxation in the Presence of Externalities", *The Swedish Journal of Economics*, Vol. 77, 1975, 86-98. (web)

Seade, Jesus K. "On the shape of optimal tax schedules." *Journal of public Economics* 7.2 (1977): 203-235. (web)

Stiglitz, J. "Self-selection and Pareto Efficient Taxation", *Journal of Public Economics*, Vol. 17, 1982, 213-240. (web)