Lecture 4: Optimal Nonlinear Income Taxation

Abdoulaye Ndiaye

NYU

- Famous A. Atkinson and J. Stiglitz, "The Design of Tax Structure: Direct vs. Indirect Taxation," Journal of Public Economics 6 (1976), 55–75. shows that

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

(i.e, commodity taxes not useful) under two assumptions on utility functions $u^h(x_1, \ldots, x_n, L)$

1. Weak separability between $x_1, \ldots, x_n)$ and L in utility
2. Homogeneity across individuals in the sub-utility of consumption $v(x_1, \ldots, x_n)$ [does not vary with h]

$$\text{(1) and (2): } u^h(x_1, \ldots, x_n, L) = U^h(v(x_1, \ldots, x_n), L)$$

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

- Gov can tax set non-linear tax on income: $R(wL)$ denotes after tax income
- Gov can also impose linear tax on consumption; cons prices $q = p + t$
- Gov revenue

$$G = \int_h [w^h L^h - R(w^h L^h) + (q - p) \cdot x^h] dh$$

- **Theorem:** Let (t, R) be any gov. policy with commodity taxes. Then there exists another government policy $(0, \tilde{R})$, with no commodity taxes, with the following properties

  1. All the agents in the economy have the same utility under $(0, \tilde{R})$ as under (t, R)
  2. All the agents supply the same amount of labor in the two allocations;
  3. Government revenue is higher under $(0, \tilde{R})$:

$$\tilde{G} = \int_h [w^h L^h - \tilde{R}(w^h L^h)] dh \geq G$$

Proof

- From pt of view of agent a gov policy is equivalent to a set $\mathcal{V}$ of $(\tilde{v}(Y), Y)$ where $\tilde{v}(Y)$ is the utility derived from consumption when before tax labor income is $Y = wL \in \mathbb{R}_+$:
$$\tilde{v}(Y) = \max_x \{v(x_1, \ldots, x_n) | (p+t) \cdot x = R(Y)\}$$
Indeed consumer h chooses labor supply by maximizing $U^h(v.L)$ for $(v, w^h L) \in \mathcal{V}$

- The new allocation, is obtained by keeping $\mathcal{V}$ unchanged. Since the agents have access to exactly the same menu $(\tilde{v}(Y), Y)$ as before, they choose the same labor supply. The menu can be supported with a more efficient choice of prices and incomes then initially.

- Define
$$\bar{z} = \arg\ \min_x \{p \cdot x | v(x) \geq \tilde{v}(Y)\}$$
and $\bar{R}(Y) = p \cdot \bar{x}$.

- From MWG this quantity is the expenditure function $\bar{R}(Y) = e(p, \tilde{v}(Y))$ and the maximum of $v(x)$ on the budget set $p \cdot x \leq \bar{R}(Y)$ is attained at $\bar{x}$.

- By definition $\bar{R}(Y)$ is smaller than $p \cdot x$ where x is the consumption under the reference allocation. Then
$$G = \int_h [w^h L^h - p \cdot x^h] dh \leq \int_h [w^h L^h - \bar{R}(w^h L^h)] dh = \bar{G}$$

Implications

- Interpretation: With separability and homogeneity, conditional on earnings Y, consumption choices x do not provide any information on ability
- Differentiated commodity taxes $(t_1, \ldots, t_n)$ create a tax distortion with no benefit $\Rightarrow$ Better to do all the redistribution with the individual income tax
- With only (weaker) linear income taxation (Diamond-Mirrlees), need v homothetic to obtain no commodity tax result [see previous exercice]
- Important implication for capital taxation. Two goods (consumption now and in the future) and labor. Separability assumption

$$u^h(c_1, c_2, l) = u(c_1, c_2) - \phi(l)$$

and budget constraint

$$c_1 + \frac{c_2}{1 + r(1 - \tau_K)} \leq w^h l^h - T(w^h l^h)$$

- it is optimal to set $\tau_k = 0$ when a nonlinear income tax is available

Non-linear Income Taxation in a Static Setting

- What determines optimal income tax?
- Specialize to 1 consumption good. Agents have preferences $u(c, l)$
- Suppose agents' type drawn from distribution $F(\theta), \theta \in \Theta \subset [0, \infty]$
- linear technology $y = \theta l$
- Feasibility
$$\int c(\theta) dF(\theta) = \int \theta l(\theta) dF(\theta)$$
- Neither θ nor l is observed. Then the government can only impose taxes on y. Let the tax on labor income be $T(y)$. Consumer of each type solves
$$\max u(c, y/\theta)$$
s.t
$$c = y - T(y)$$

Mirrlees Model results

- Optimal income tax trades-off redistribution and efficiency (as tax based on w only not feasible)
- $T(.) < 0$ at bottom (transfer) and $T(.) > 0$ further up (tax) [full integration of taxes/transfers]
- Mirrlees formulas complex, only a couple fairly general results:
- $0 \leq T'(.) \leq 1$, $T'(.) \geq 0$ is non-trivial (rules out EITC) [Seade '77]
- Marginal tax rate $T'(.)$ should be zero at the top (if skill distribution bounded) [Sadka '76-Seade '77]
- If everybody works and lowest $wl > 0$, $T'(.) = 0$ at bottom

Pareto Optimal Taxation

- What are the optimal taxes? There are more than one agent, so the social planner has to specify some social welfare function. The social planner solves the following problem for some weights $\alpha \geq 0$:

$$\max_T \int \alpha(\theta) u(c^*(\theta), y^*(\theta)/\theta) dF(\theta)$$

s.t.

$$(c^*(\theta), y^*(\theta)) = \arg \max_{\{c,y: c=y-T(y)\}} u(c, y/\theta)$$

and

$$\int c^*(\theta) dF(\theta) + g = \int y^*(\theta) dF(\theta)$$

$$g = \int T(y^*(\theta)) dF(\theta)$$

- The last condition is automatically satisfied because of the Walras law.

Two ways to characterize income taxes

1. Using mechanism design tools (analogue of "primal approach")
2. Maximization with respect to T directly (analogue of "dual approach")

- In this lecture, we focus on the first approach

Necessary conditions

Note that for any T from optimizing behavior of each individual (i.e. consumer i chooses (c_i, y_i) instead of (c_j, y_j)), the following conditions must be satisfied

$$u(c^*(\theta'), y^*(\theta')/\theta') \geq u(c^*(\theta''), y^*(\theta'')/\theta') \text{ for all } \theta', \theta''$$

Therefore these are necessary conditions on allocations.

Sufficient Conditions: Taxation Principle

- Are these conditions sufficient?
- Yes, since we impose no restriction on T we may choose such a T that $T(y) = y - c$ for $y = y^*(\theta)$ for some θ, and $T(y) = y$ for all other y.
- This is not the only tax function that implements the optimum, and we can find a continuum of such function.
- in particular, this tax function is not smooth.

Revelation Principle

What we could do is instead just ask them what their types are directly. Then the problem becomes

$$\max_{c,y} \int \alpha(\theta) u(c(\theta), y(\theta)/\theta) dF(\theta)$$

$$u(c(\theta'), y(\theta')/\theta') \geq u(c(\theta''), y(\theta'')/\theta') \text{ for all } \theta', \theta''$$

$$\int c(\theta) dF(\theta) + g = \int y(\theta) dF(\theta)$$

This is an application of the famous Revelation Principle

IC constraints

- Too many IC constraints:

$$u(c^{(}\theta'), y^{(}\theta')/\theta') \geq u(c^{(}\theta''), y^{(}\theta'')/\theta') \text{ for all } \theta', \theta''$$

- Would be nice not to worry about most of them

- Change of variables c, y to u, y

$$u(\theta) = u(c(\theta), y(\theta); \theta)$$

- Let's get the local IC in terms of u...

$$u(\theta) - u(c(\theta'), y(\theta'); \theta) \leq 0$$

  and with equality for $\theta' = \theta$. So $u(\theta) - u(c(\theta'), y(\theta'); \theta)$ is maximized at $\theta = \theta'$

- An Envelope Theorem [Milgrom and Segal] implies

$$u'(\theta) = u_\theta(c(\theta), y(\theta); \theta) = u_l(c(\theta), y(\theta)/\theta)(-\frac{y(\theta)}{\theta^2})$$

- Are these necessary conditions sufficient? Here yes with SCP u and $y(\theta)$ monotonically increasing

General Mirrlees Program

- The problem becomes

$$\max_{c,y,u} \int \alpha(\theta) u(c(\theta), y(\theta)/\theta) dF(\theta)$$

$$u'(\theta) = u_l(c(\theta), y(\theta)/\theta)(-\frac{y(\theta)}{\theta^2})$$

$$u(\theta) = u(c(\theta), y(\theta)/\theta)$$

$$\int c(\theta) dF(\theta) + g = \int y(\theta) dF(\theta)$$

- Consider isoelastic preferences $u(c, l) = u(c) - h(l) = \frac{c^{1-\sigma}}{1-\sigma} - \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon}$

- Suppose that $y'(\theta) \geq 0$ and verify ex-post

- Set up the Hamiltonian [see lecture notes]

- Marginal taxes obtained from FOC of individuals:

$$u'(c)[1 - \tau(\theta)] = \frac{h'(y/\theta)}{\theta}$$

Optimal Tax

- Optimal tax

$$\frac{\tau(\theta)}{1-\tau(\theta)} = A(\theta)B(\theta)C(\theta)$$

where

-
$$A = 1 + \frac{1}{\varepsilon}$$

-
$$B = \frac{1-F(\theta)}{\theta f(\theta)}$$

-
$$C(\theta) = \int_\theta^\infty \exp\left(\int_\theta^x \sigma \frac{\dot{c}(x')}{c(x)} dx'\right)(1-\lambda\alpha(x)u_c(X))\frac{f(x)}{1-F(\theta)} dx$$

-
$$\lambda = \int_\Theta \frac{f(x)}{u_c(x)} dx$$

Intuition

- Three key parameters:

  1. elasticity of labor supply ε
  2. distribution of types $F(\theta)$
  3. degree of redistribution α

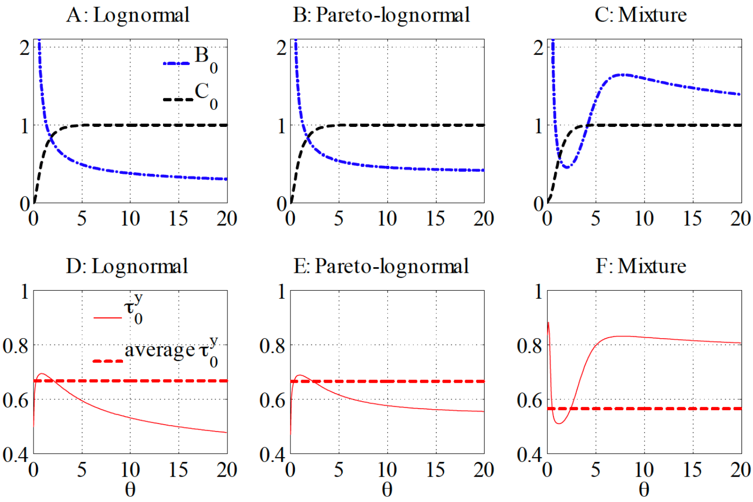
Intuition

- High labor elasticity $\Rightarrow$ taxes more distortionary $\Rightarrow$ low marginal taxes
- Marginal tax on type θ needed to redistribute to θ from $1 - F(\theta)$ more productive types
  1. large $1 - F(\theta) \Rightarrow$ high marginal taxes
  2. large $\theta f(\theta) \Rightarrow$ low marginal taxes
- tail ratio $(1 - F(\theta))/(\theta f(\theta))$ is key
- More redistribution $\Rightarrow$ larger $C(\theta) \Rightarrow$ higher marginal taxes

Understanding tail behavior Saez '01, GTT'18

- Suppose (!) marginal tax rates converge: $\lim_{\theta \longrightarrow \infty} \tau(\theta) = \bar{\tau} < 1$
- Once can show that $\lim_{\theta \longrightarrow \infty} C(\theta) = 1 + \frac{\sigma}{\sigma+\varepsilon}\frac{\bar{\tau}}{1-\bar{\tau}} = const > 0$
- 3 common class of distributions
  - lognormal
  - Pareto-lognormal
  - mixture of lognormals
- Right tail ($\theta \to \infty$)
  - lognormal and mixture: $\frac{\tau(\theta)}{1-\tau(\theta)} \sim \frac{1+1/\varepsilon}{\log \theta}$
  - Pareto-lognormal: $\frac{\tau(\theta)}{1-\tau(\theta)} \sim [\frac{a}{1+1/\varepsilon} - \frac{\sigma}{\sigma+1/\varepsilon}]^{-1}$
- Left tail ($\theta \to 0$): all three $\frac{\tau(\theta)}{1-\tau(\theta)} \sim -\frac{1}{\log \theta}$

Shape of Optimal Taxes

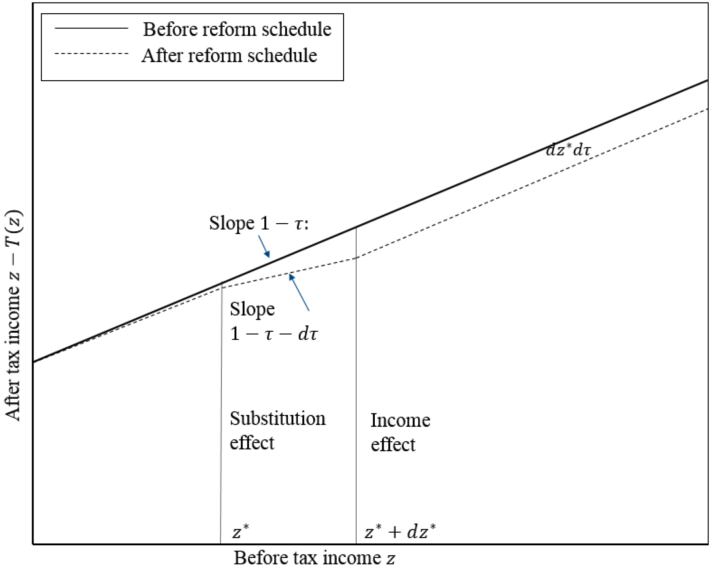


Another look at this problem

- Suppose we took some types in the region $[y, y + dy]$ and increase their marginal taxes by $d\tau$.
- Marginal taxes of everyone else are unchanged

1. Mechanical: how much tax revenues change for those whose marginal taxes were not affected
2. Behavioral: how much tax revenues change for those whose marginal taxes changed
3. Welfare: how much utility we lost from those who now pay higher taxes

Saez Reform


Before reform schedule
After reform schedule
$dz^* d\tau$
Slope $1-\tau$:
Slope
$1-\tau-d\tau$
Substitution
effect
Income
effect
z^*
z^*+dz^*
Before tax income z
After tax income $z-T(z)$

Mechanical Effect

- Everybody above θ (or $y(\theta)$) face the same marginal taxes but higher average taxes
- Average taxes increased by $d\tau dy$
- Assume no income effect, total revenues are

$$M = dyd\tau(1 - F(\theta))$$

Behavioral Effect

- Wages of consumers in $[\theta, \theta + d\theta]$ decreased from $\theta(1-\tau)$ to $\theta(1-\tau-d\tau)$, i.e. by $\frac{d\tau}{1-\tau}$
- if ε is elasticity of labor supply, by definition this change in marginal taxes affects labor by $l\varepsilon\frac{d\tau}{1-\tau}$
- from reduction in labor supply taxes on one individual drop by

$$\tau y \varepsilon \frac{d\tau}{1-\tau}$$

- What is the relationship between dy and $d\theta$?
  - there are $f(\theta)d\theta$ people in $[\theta, \theta + d\theta]$
  - direct calculation
    $$\frac{dy}{d\theta} = \frac{d(l\theta}{d\theta} = l(1+\varepsilon)$$
  - therefore
    $$dy = l(1+\varepsilon)d\theta$$
- If $\frac{f(\theta)dy}{l(1+\varepsilon)}$ people decrease taxes by $\tau y \varepsilon \frac{d\tau}{1-\tau}$, total behavioral effect is
  $$B = \frac{\tau}{1-\tau}\frac{\varepsilon}{1+\varepsilon}\theta f(\theta) d\tau dy$$

- As society, we value one extra dollar in hands of type θ as $\alpha(\theta)$
- So welfare loss from taking $dyd\tau$ from these types is

$$W = -dyd\tau \int_{\theta}^{\infty} \alpha(\theta')dF(\theta')$$

Optimal Taxes

- if T is optimal, then $M + B + W = 0$
- Direct substitution

$$y\frac{\tau}{1-\tau}\frac{\varepsilon}{1+\varepsilon}f(\theta) + \int_{\theta}^{\infty}(1-\alpha(\theta'))dF(\theta') = 0$$

- Equivalently

$$\frac{\tau}{1-\tau} = (1+1/\varepsilon)\frac{1-F(\theta)}{\theta f(\theta)}\int_{\theta}^{\infty}(1-\alpha)\frac{dF}{1-F(\theta)}$$

Bibliography

Atkinson, A. B., and Stiglitz, J. E. (1976). "The Design of Tax Structure: Direct versus Indirect Taxation." *Journal of Public Economics*, 6(1–2), 55–75.

Doligalski, P., Ndiaye, A., and Werquin, N. (2022). "Redistribution with Performance Pay." Federal Reserve Bank of Chicago Working Paper No. 2022–12.

Golosov, M., Troshkin, M., and Tsyvinski, A. (2016). "Redistribution and Social Insurance." *American Economic Review*, 106(2), 359–386.

Bibliography (cont.)

Kaplow, L. (2006). "On the Undesirability of Commodity Taxation Even When Income Taxation Is Not Optimal." *Journal of Public Economics*, 90(6–7), 1235–1250.

Laroque, G. (2005). "Indirect Taxation Is Superfluous under Separability and Taste Homogeneity: A Simple Proof." *Economics Letters*, 87(1), 141–144.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press.

Bibliography (cont.)

Milgrom, P., and Segal, I. (2002). "Envelope Theorems for Arbitrary Choice Sets." *Econometrica*, 70(2), 583–601.

Mirrlees, J. A. (1971). "An Exploration in the Theory of Optimum Income Taxation." *The Review of Economic Studies*, 38(2), 175–208.

Sadka, E. (1976). "On Income Distribution, Incentive Effects and Optimal Income Taxation." *The Review of Economic Studies*, 43(2), 261–267.

Saez, E. (2001). "Using Elasticities to Derive Optimal Income Tax Rates." *The Review of Economic Studies*, 68(1), 205–229.

Seade, J. K. (1977). "On the Shape of Optimal Tax Schedules." *Journal of Public Economics*, 7(2), 203–235.