Notes on Non-linear Taxation

Iván Werning

1 Income Taxation

1.1 Setup

- two goods
  (alternatively, other goods untaxed, perhaps because of Atkinson-Stiglitz case)

- Preference
$$U^i(c, Y) = U(c, Y, \theta^i)$$

- Technology
$$G + \int (c(\theta) - Y(\theta)) dF(\theta) \leq e$$

- F can be continuous or not (e.g. with finite types $G + \sum_i (c(\theta^i) - Y(\theta))\pi^i \leq e$)

- Income taxation: budget constraint is
$$B = \{(c, Y) | c \leq Y - T(Y)\}$$
  for some $T(Y)$.

- call $R(Y) \equiv Y - T(Y)$ the retention function

- Normative Criterion?

  1. Welfare function (Mirrlees, 1971)
  2. Pareto efficiency

1.2 Feasibility and Incentive compatibility

- agent behavior

$$\max_{c,Y} U^i(c,Y) \qquad \text{s.t.} c \leq Y - T(Y) = R(Y)$$

- Definition. An allocation and a tax function c, Y, T is feasible if: (i) RC holds; (ii) agents maximize $\{c(\theta), y(\theta)\}$ given $T(Y)$ [given $R(Y)$]

- An allocation $c(\theta), Y(\theta)$ is feasible if there exists a tax function that makes c, Y, T feasible.

- Note that resource constraint is the same as...

$$G - e \leq \int T(Y(\theta))dF(\theta)$$

  government budget constraint.

- Observation: if c, Y is feasible then

$$u(c(\theta), Y(\theta), \theta) \geq u(c(\theta'), Y(\theta'), \theta) \qquad \text{for all } \theta, \theta \in \Theta$$

  Incentive Compatibility Constraints (IC)

- Converse also true:

$$R(\tilde{Y}) \equiv \sup\{\tilde{c} | u(c(\theta), Y(\theta), \theta) \geq u(\tilde{c}, \tilde{Y}, \theta) \text{ for all } \theta\}$$

  IC holds $\Longrightarrow$ if agents faced with R then optimum is attainable (optimal by definition of R)

- note: $R(Y)$ continuous by theorem of the max.

- Taxation principle and revelation principle

- Marginal taxes: if $T'(\tilde{Y})$ exists and $\tilde{Y} = Y(\theta)$ for some θ then

$$T'(\tilde{Y}) = T'(Y(\theta)) = 1 - MRS(c(\theta), Y(\theta), \theta)$$

  where

$$MRS(c, Y, \theta) \equiv -\frac{U_Y(c, Y, \theta)}{U_c(c, Y, \theta)}$$

- Single crossing

$$MRS(c, Y, \theta) \qquad \text{is decreasing in } \theta$$

- Single crossing $\implies c(\theta)$ and $y(\theta)$ non-decreasing
- Finite types: Single crossing $\implies$ local IC are sufficient

$$\begin{aligned} u(c(\theta^i), y(\theta^i), \theta^i) &\geq u(c(\theta^{i-1}), y(\theta^{i-1}), \theta^i) \\ u(c(\theta^{i-1}), y(\theta^{i-1}), \theta^{i-1}) &\geq u(c(\theta^i), y(\theta^i), \theta^{i-1}) \end{aligned}$$

  [homework: show others are implied]

- note: local IC's imply monotonicity

1.3 Two types case

- Assume $\Theta = \{\theta_L, \theta_H\}$ with $\theta_L < \theta_H$
- result 1: pooling is inefficient
  ...only one IC binds
- result 2: for binding agent we have $MRS = 1$
  ... no taxation at the top.
- result 3: Pareto frontier has 3 regions

  1. First best
  2. IC for H binds
  3. IC for L binds

  and frontier bends backwards

- Program ($\bar{u}_H$ is parameter here)

$$\max u^L(c_L, y_L)$$

  subject to

$$u^H(c_H, y_H) \geq \bar{u}_H$$

  ICh, ICl and RC.

- First order conditions: derive same results

1.4 Laffer curve

- back to general case
- When is there a pareto improvement?
- Equivalently: when can we lower taxes and increase tax revenue?
- Given $T_0(Y)$ we get $Y_0(\theta)$ and we have

$$G - e \leq \int T_0(Y_0(\theta))dF(\theta)$$

- Is this Pareto efficient?
- suppose budget holds with equality.
- take T_1 prefered to T_0
- for feasibility we must have

$$G - e = \int T_0(Y_0(\theta))dF(\theta) \leq \int T_1(Y_1(\theta))dF(\theta)$$

  where $T_1(Y)$ generates $Y_1(\theta)$

- for improvement it must be that

$$T_1(Y_1(\theta)) \leq T_0(Y_1(\theta)) \qquad \text{for all } \theta$$

  and we can always make $T_1(\tilde{Y}) \leq T(\tilde{Y})$ at other points

- hence: (sophisticated) Laffer effect
- Result: there are such Laffer effects (we know from two type case)
- more general results: joint restrictions on...
  - tax schedule T
  - preference U
  - skill distribution F

  (note: allocation is implied)

- convert results: joint restrictions on...
  - tax schedule T
  - preference U
  - distribution of output $G(\tilde{Y})$ (where $G(Y(\theta)) = F(\theta)$)

  (note: skill distribution is implied)

1.5 IC with a continuum

- need to make IC simpler
  - necessary: local IC + monotonicity
  - also sufficient!

- informally

$$v(\theta) = \max_{\theta'} U(c(\theta'), y(\theta'), \theta) = U(c(\theta), y(\theta), \theta)$$

  first order condition...

$$U_c(c(\theta'), y(\theta'), \theta)c'(\theta') + U_y(c(\theta'), y(\theta'), \theta)y'(\theta') = 0$$

  or rearranging

$$\left[\frac{c'(\theta')}{y'(\theta')} - MRS(c(\theta'), y(\theta'), \theta)\right] U_c(c(\theta'), y(\theta'), \theta)y'(\theta') = 0$$

$$h(\theta, \theta')g(\theta, \theta') = 0$$

- we want this to hold for $\theta = \theta'$ (truth-telling)

- second order condition (informally)

$$h_{\theta'}(\theta, \theta)g(\theta, \theta) + h(\theta, \theta)g_{\theta'}(\theta, \theta) \leq 0$$

  Now, either g or h is zero so that we need to worry about the other term. In regions where $g = 0$, trivially satisfied. In other regions, with $y' > 0$ and $h = 0$ we need

$$h_{\theta'}(\theta, \theta) \leq 0$$

but we know that (since $h(\theta,\theta) = 0$ over this region):

$$h_{\theta'}(\theta,\theta) + h_\theta(\theta,\theta) = 0$$

and we know that $h_\theta(\theta,\theta) > 0$ by the single crossing condition! QED

- stronger result: not just local SOC but actually a max
- note that $g(\theta,\theta') > 0$
  - for $\theta' < \theta$ then we have
    $$h(\theta,\theta') > h(\theta',\theta') = 0$$
  - the reverse is true for $\theta' > \theta$
    $$h(\theta,\theta') < h(\theta',\theta') = 0$$
  - thus, $\theta' = \theta$ is optimal
- a better approach:
  - change of variables c, y to v,y
    $$v(\theta) = U(c(\theta), y(\theta), \theta)$$
    $$c(\theta) = e(v(\theta), y(\theta), \theta)$$
  - let's get the local IC in terms of v...
    $$v(\theta) - U(c(\theta'), y(\theta'), \theta) \leq 0$$
    and with equality for $\theta' = \theta$. So $v(\theta) - U(c(\theta'), y(\theta'), \theta)$ is maximized over θ at $\theta = \theta'$. The FOC must be
    $$v'(\theta) - U_\theta = 0$$
  - more generally, an Envelope theorem implies
    $$v'(\theta) = U_\theta(c(\theta), y(\theta), \theta)$$

or the integral version...

$$v(\theta) = \int^{\theta} U_\theta(c(\theta'), y(\theta'), \theta')d\theta'$$

[Milgrom and Segal]

- Result: v,y is incentive compatible (i.e. implies c, y that is IC) iff EC and monotonicity hold

- Duality: Pareto efficiency iff minimize resources subject to delivering $v(\theta)$ or more

2 Pareto Efficient Income Taxation

- Dual for Pareto efficiency:

$$\max_{y,v} \int (y(\theta) - e(v(\theta), y(\theta), \theta))f(\theta)d\theta$$

$$v'(\theta) = U_\theta(e(v(\theta), y(\theta), \theta), y(\theta), \theta)$$

$$v(\theta) \geq \bar{v}(\theta)$$

- where $\{\bar{v}(\theta)\}$ is some parameter

- Remarks:

  - from this we can compute $c(\theta) = e(v(\theta), y(\theta), \theta)$ and then with $c(\theta), y(\theta)$ find retention function $R(y)$ and tax function $T(y)$
  - in general we will have some multiplier $\zeta(\theta) = \lambda(\theta)f(\theta)/\eta$ on the last constraint (with $\lambda(\theta) = 0$ if the constraint is slack)
  - $\zeta(\theta) = \lambda(\theta)f(\theta)/\eta$ and $\bar{v}(\theta)$ related

- Solve

$$\max_{y,v}\{\int (y(\theta) - e(v(\theta), y(\theta), \theta))f(\theta)d\theta + \frac{1}{\eta}\int \lambda(\theta)v(\theta)f(\theta)d\theta\}$$

subject to

$$v'(\theta) = U_\theta(e(v(\theta), y(\theta), \theta), y(\theta), \theta)$$

- equivalently

$$\frac{1}{\eta}\max_{y,v}\{\eta\int(y(\theta)-e(v(\theta),y(\theta),\theta))f(\theta)d\theta+\int\lambda(\theta)v(\theta)f(\theta)d\theta\}$$

which is the Lagrangian that comes out of the problem with objective (Weflare function?)

$$\max_{y,v}\{\int\lambda(\theta)v(\theta)f(\theta)d\theta\}$$

$$v'(\theta)=U_\theta(e(v(\theta),y(\theta),\theta),y(\theta),\theta)$$

$$\int(y(\theta)-e(v(\theta),y(\theta),\theta))f(\theta)d\theta\leq e$$

for some e (related to η)

- Utilitarian case is then $\lambda(\theta)=1$
  - optimal control (v is state and y is control)
  - FOCs the same with $\lambda(\theta)=1$ (total multiplier $\zeta(\theta)=f(\theta)/\eta$)

- Form Lagrangian

$$L\equiv\int(y(\theta)-e(v(\theta),y(\theta),\theta))f(\theta)d\theta+\frac{1}{\eta}\int\lambda(\theta)v(\theta)f(\theta)d\theta$$
$$+\int\mu'(\theta)v(\theta)+\int\mu(\theta)U_\theta(e(v(\theta),y(\theta),\theta),y(\theta),\theta)$$

- defining

$$\hat{\mu}=\mu U_c$$

- FOCs:

$$-\hat{\mu}'(\theta)-\hat{\mu}(\theta)\frac{\partial MRS(c(\theta),y(\theta),\theta)}{\partial c}y'(\theta)+\zeta(\theta)U_c(c(\theta),y(\theta),\theta)=f(\theta)$$

$$\frac{\tau(\theta)}{1-\tau(\theta)}f(\theta)=-\hat{\mu}(\theta)\frac{\partial\log MRS(c(\theta),y(\theta),\theta)}{\partial\theta}$$

- Pareto efficiency: multiplier $\zeta(\theta)\geq0$ so check...

$$-\hat{\mu}'(\theta)-\hat{\mu}(\theta)\frac{\partial MRS(c(\theta),y(\theta),\theta)}{\partial c}y'(\theta)\leq f(\theta)$$

$$\frac{\tau(\theta)}{1-\tau(\theta)} f(\theta) = -\hat{\mu}(\theta) \frac{\partial \log MRS(c(\theta), y(\theta), \theta)}{\partial \theta}$$

- take tax system, utility as given...
  ... hence, take allocation, taxes and utility as given
- observe distribution of output: infer distribution of skills (more later)
- second equation gives $\hat{\mu}$ uniquely, first inequality is restriction
- note: anything goes: there exists an f such that condition is met given U and T
- given f and U: many T are inefficient, many efficient
- tax at top and bottom: $\tau(\bar{\theta}) \leq 0$ and $\tau(\underline{\theta}) \geq 0$

- Utilitarian: set $\zeta(\theta) = f(\theta)/\eta$ and solve ODEs:

$$-\hat{\mu}'(\theta) - \hat{\mu}(\theta) \frac{\partial MRS(c(\theta), y(\theta), \theta)}{\partial c} y'(\theta) + \frac{1}{\eta} f(\theta) U_c(c(\theta), y(\theta), \theta) = f(\theta)$$

$$\frac{\tau(\theta)}{1-\tau(\theta)} f(\theta) = -\hat{\mu}(\theta) \frac{\partial \log MRS(c(\theta), y(\theta), \theta)}{\partial \theta}$$

along with $\mu(\underline{\theta}) = \mu(\bar{\theta}) = 0$ [since state v is free at the boundaries; or note the special FOCs for them derived in recitation]

  - a bit more involved
  - can be done numerically
  - special cases: U quasi-linear in c (so that $\frac{\partial MRS(c(\theta), y(\theta), \theta)}{\partial c} = 0$)
  - check Diamond and Saez papers

- Saez identification: observe output distribution H...

$$F(\theta) = H(y(\theta))$$

$$f(\theta) = h(y(\theta)) y'(\theta)$$

- given U and T we can compute $y(\theta)$ and hence $y'(\theta)$...

$$y'(\theta) = \frac{-\frac{\partial \log MRS(c(\theta), y(\theta), \theta)}{\partial \theta}}{\frac{1}{\varepsilon_w^* Y} + \frac{T''(Y)}{1-T'(Y)}}$$

- ...nicer to solve for $y'(\theta)$ in terms of local conditions (some algebra later; see Recitation)

$$-\hat{\mu}'(\theta) - \hat{\mu}(\theta)\frac{\partial MRS(c(\theta), y(\theta), \theta)}{\partial c} \leq h(y(\theta))$$

$$\hat{\mu}(\theta) = \frac{T'(y)}{1 - T'(y)}\varepsilon^*_w(Y)Y\frac{h(y(\theta))}{1 + Y\varepsilon^*_w(Y)\frac{T''(Y)}{1-T'(Y)}}$$

- (utilitarian case: $-\hat{\mu}'(\theta) - \hat{\mu}(\theta)\frac{\partial MRS(c(\theta),y(\theta),\theta)}{\partial c} = (1 - \lambda U_c)h(y(\theta)))$

- Saez defines the "virtual density"...

$$h^*(Y) = \frac{h(y)}{1 + Y\varepsilon^*_w(Y)\frac{T''(Y)}{1-T'(Y)}} = \frac{h(y)}{\Phi(y)}$$

- after subsituting...

$$\frac{\tau}{1-\tau}\frac{\varepsilon^*_w}{\Phi}\left(-\frac{d\log\frac{\tau}{1-\tau}}{d\log y} - \frac{d\log h^*}{d\log Y} - 1 - \frac{d\log\varepsilon^*_w}{d\log Y} - \frac{\partial MRS}{\partial c}\frac{1}{y}\right) \leq 1$$

- think through role of each term

- intuition for inefficient tax: Laffer argument

- generalizes: tax at top and bottom: $\tau(\bar{\theta}) \leq 0$ and $\tau(\underline{\theta}) \geq 0$

$$\frac{d\log h^*}{d\log Y} = -\infty \qquad \text{or} \qquad \frac{d\log h^*}{d\log Y} = \infty$$

- Discussion:

  1. anything goes again: exists h^* given U and T
  2. linear tax optimal? Yes, depends on distribution.
  3. maximum level of asymptotic tax rate (many terms cancel)
  4. connection with Rawlsian optimum

- observable characteristics: Differential taxation?

  - Pareto efficient to ignore? Yes, in some cases...
    ... new condition is average of previous condition

- Pareto improvement to differentiate? Yes, in some cases...
  ... if previuos condition is violated for some group

3 Extensive Margin Model

- Diamond (1980): nonlinear taxation with extensive margin (no intensive margin).
- as before, preferences are
$$U(c, Y, \theta)$$
- but now
  - assume only two possible levels of Y for each θ
$$\{0, Y(\theta)\}$$
  - $Y(\theta)$ continuous and increasing
  - $\theta \in [0, \infty)$
  - $Y(0) = 0$
  - assume some measure N of agents simply cannot work
- for agent θ to prefer work we require
$$U(c(\theta), Y(\theta), \theta) \geq U(b, 0, 0)$$
- incentive constraint
  - like before, compares allocation intended for θ to others'
  - previously:
    * compared to all other bundles
    * binding were neighbours (with single crossing assumptions)
  - now: relevant binding constraint is always allocation meant for $\theta = 0$ i.e. $Y = 0$, so binding constraint skips neighbours, in this sense this is a violation of single crossing
- it might be optimal to make some agents that are capable of working not work, but we will assume instead that we make them all work

- Planning Problem:

$$\max\left\{NU(b,0,\theta)+\int U(c(\theta),Y(\theta),\theta)f(\theta)d\theta\right\}$$

s.t.

$$Nb+\int(c(\theta)-Y(\theta))f(\theta)d\theta\leq e$$
$$U(c(\theta),Y(\theta),\theta)\geq U(b,0,0)$$

- first order conditions:

$$U_b(b,0,0)=\lambda-\frac{1}{N}\int\mu(\theta)d\theta$$
$$U_c(c(\theta),Y(\theta),\theta)=\lambda+\mu(\theta)$$

and $\mu(\theta)\geq 0$

$$\mu(\theta)[U(c(\theta),Y(\theta),\theta)-U(b,0,0)]=0$$

as well as both constraints holding.

- combining the conditions gives:

$$U_c(c(\theta),Y(\theta),\theta)\leq U_b(b,0,0)+\frac{1}{N}\int\mu(\theta)d\theta<U_b(b,0,0)$$

as long as the work constraint binds for some agents

- with separable utility $u(c)-v(Y,\theta)$ this implies immediately that

$$c(\theta)>b$$

for all θ and that

$$\lim_{\theta\to 0}c(\theta)>b$$

- indeed with separable utility we must have $\mu(\theta)=0$ and $U(c(\theta),Y(\theta),\theta)>U(b,0,0)$ for low enough θ

- nice case has preferences independent of θ:

$$u(c,Y,\theta)=U(c,Y)$$

- then easy to see that defining the equalizing difference

$$U(c^e(Y), Y) = U(b, 0)$$

we have that optimal consumption is

$$c^*(Y) = \max\{c^e(Y), \bar{c}\}$$

where $\bar{c} \equiv (u')^{-1}(\lambda) > b$

- conclusions:
  - we get an upward discontinuity in $c(\theta)$
  - more general: symptomatic that marginal tax may be negative
  - possible odd result: consumption $c(\theta)$ may not be monotone (could be fixed with additional assumptions) (i.e. marginal taxes may be higher than 100%)

- comments:
  - overall:
    * with Utilitarian, less sharp restrictions on marginal taxes Mirrlees model, i.e. here they can be negative or higher than 100%.
    * but not clear if less implications for Pareto efficient.
  - Here: some people can work, others suffer infinite disutility (i.e. can't work). Diamond's paper has a more general joint distribution between skill and labor disutility. Optimum then needs to determine how many people work, at each skill level.
  - Saez combines intensive and extensive margin.

- Planning problem

$$\max\left\{\int\int_{n^*(\theta)} U(b,0,\theta)dnd\theta + \int\int^{n^*(\theta)} U(c(n), Y(n), \theta)f(\theta, n)dnd\theta\right\}$$

s.t.

$$Nb + \int (c(\theta) - Y(\theta))f(\theta)d\theta \leq e$$

$$U(c(\theta), Y(\theta), \theta) \geq U(b, 0, 0)$$