14.471: Fall 2012: Recitation 9: Pareto Efficient Income Taxation

Daan Struyven

November 9, 2012

Questions: When is an income tax system efficient? How to derive the test for optimal income taxation from "Werning, Iván., 2007. Pareto efficient income taxation"? What is the intuition for the different terms in the efficiency test?

Note: The NBER PE 2007 Meeting slides associated to Iván's 2007 paper

1 Definition of efficient allocation

Proposition. An allocation $\{c(\theta), y(\theta)\}$ is pareto efficient only if it solves the problem

$$\begin{aligned} \max_{y(\theta),v(\theta)} & \int_{\underline{\theta}}^{\overline{\theta}} (y(\theta) - c(\theta))\, dF(\theta) \\ \text{s.t.} \quad & U(c(\theta), y(\theta), \theta) \geq U(c(\theta'), y(\theta'), \theta) \quad \forall \theta, \theta' \\ & U(c(\theta), y(\theta), \theta) \geq \overline{v}(\theta) \quad \forall \theta \end{aligned}$$

Proof. Suppose not. Let $\{c^\star(\theta), y^\star(\theta)\}$ be a superior allocation. The star allocation is incentive compatible by construction of the problem. It is resource feasible because it increases the objective and the objective is the excess resources available. Furthermore, it is at least as good for all agents as the original allocation and has relaxed the resource constraint. Therefore, provided the value function is continuous in at least one $\overline{v}(\theta)$, it's possible to allocate the excess resources in a way that generates a pareto improvement.

We now work to simplify the problem. First, observe the that an equivalent way of writing the incentive constraints is $\theta = \arg\max_{\theta'} U(c(\theta'), y(\theta'), \theta)$. Define $v(\theta) \equiv \max_{\theta'} U(c(\theta'), y(\theta'), \theta)$. The envelope theorem then implies that $v'(\theta) = U_\theta(c(\theta), y(\theta), \theta)$.

The intuition for the dual is that we maximize net resources (we can redistribute them) subject to incentive and minimal utility level targets (and monotonicity).

As in lecture we change variables from (c, y) to (v, y). Define $c(\theta) = e(v(\theta), y(\theta), \theta)$. This function exists and is well-behaved with standard assumptions on U (since e is the inverse of U with respect to the first argument).

Lemma. Global IC is equivalent to local IC and $y(\theta)$ nondecreasing.

The modified problem using the change of variables and the local IC is

$$\begin{aligned} \max_{v(\theta),y(\theta)} & \int_{\underline{\theta}}^{\overline{\theta}} (y(\theta) - e(v(\theta), y(\theta), \theta))\, f(\theta)\, d\theta \\ \text{s.t.} \quad & v'(\theta) = U_\theta(e(v(\theta), y(\theta), \theta), y(\theta), \theta) \\ & v(\theta) \geq \overline{v}(\theta) \\ & y(\theta) \ \text{nondecreasing.} \end{aligned}$$

2 Helpful results from consumer's problem

We will derive first order conditions for this problem and manipulate them into an interpretable form. We start by deriving several results that will be of use in the manipulation. The formal equilibrium definition for this problem requires that consumers solve the associated problem:

$$\max_{c,y} U(c, y, \theta) \quad \text{s.t.} \quad c \leq y - T(y).$$

Substituting the budget constraint into the utility function and taking a first order condition in y,

$$U_c(y - T(y), y, \theta)(1 - T'(y)) + U_y(y - T(y), y, \theta) = 0.$$

Define $\tau(y) \equiv T'(y)$. Then the first order condition can be expressed simply as

$$1 - \tau(y) = -\frac{U_y}{U_c} \equiv MRS(c, y, \theta). \tag{1}$$

Define $w(y) = 1 - \tau(y)$, the after-tax marginal wage at income y. The first order condition can then also be written as

$$wU_c(y - T(y), y) + U_y(y - T(y), y) = 0.$$

Expressed as a function of w alone and differentiating with respect to w,

$$U_c + wU_{cc}(1 - T')\frac{dy}{dw} + wU_{cy}\frac{dy}{dw} + U_{yc}(1 - T')\frac{dy}{dw} + U_{yy}\frac{dy}{dw} = 0.$$

Solving for $\frac{dy}{dw}$, this implies

$$\frac{dy}{dw} = -\frac{U_c}{U_{cc}(1 - T')^2 + 2U_{cy}(1 - T') + U_{yy}}.$$

Therefore the local elasticity of earnings with respect to the after tax wage is given by

$$\begin{aligned}
\varepsilon_w^\star(y) &= \frac{\partial y}{\partial w}\frac{w}{y} \\
&= -\frac{U_c(1 - T')/y}{U_{cc}(1 - T')^2 + 2U_{cy}(1 - T') + U_{yy}} \\
&= -\frac{1/y}{\frac{U_{cc}}{U_c}(1 - T') + 2\frac{U_{cy}}{U_c} + \frac{U_{yy}}{U_c(1 - T')}}.
\end{aligned}$$

The form we will later desire is

$$-\frac{1}{\varepsilon_w^\star y} = \frac{U_{cc}}{U_c}(1 - T') + 2\frac{U_{cy}}{U_c} + \frac{U_{yy}}{U_c(1 - T')}. \tag{2}$$

We can also think of the consumer allocation as a function of θ. Note that $c(\theta) = y(\theta) - T(y(\theta))$, which implies $c' = y' - T'y'$. Then

$$c'/y' = 1 - T' = MRS = -U_y/U_c \tag{3}$$

using the results from the consumer problem shown in equation (1) above.

By definition, $v = U(e(v, y, \theta), y, \theta)$. Therefore, differentiating with respect to y,

$$\begin{aligned}
0 &= U_c e_y + U_y \\
\Rightarrow e_y &= -\frac{U_y}{U_c} = MRS,
\end{aligned} \tag{4}$$

and, differentiating with respect to v,

$$\begin{aligned} 1 &= U_c e_v \\ \Rightarrow e_v &= \frac{1}{U_c}. \end{aligned} \tag{5}$$

Next we do some algebra on the utility functions themselves. Observe that

$$\begin{aligned} \frac{\partial MRS}{\partial c} y' &= \frac{\partial}{\partial c}\left[-\frac{U_y\left(c, y, \theta\right)}{U_c\left(c, y, \theta\right)}\right] y' \\ &= -\frac{U_c U_{yc} - U_y U_{cc}}{U_c^2} y' \\ &= -\frac{U_{yc} - U_{cc}\frac{U_y}{U_c}}{U_c} y' \\ &= -\frac{U_{yc} + U_{cc}\frac{c'}{y'}}{U_c} y' \\ &= -\frac{U_{cy} y' + U_{cc} c'}{U_c}. \end{aligned} \tag{6}$$

and

$$\begin{aligned} -U_c \frac{\partial MRS}{\partial \theta} &= -U_c \frac{\partial}{\partial \theta}\left[-\frac{U_y\left(c, y, \theta\right)}{U_c\left(c, y, \theta\right)}\right] \\ &= U_c \frac{U_{y\theta} U_c - U_{c\theta} U_y}{U_c^2} \\ &= U_{\theta y} - U_{\theta c}\frac{U_y}{U_c} \\ &= U_{\theta y} + U_{\theta c} e_y. \end{aligned} \tag{7}$$

Returning to the first order condition for the consumer problem

$$U_c\left(y - T\left(y\right), y, \theta\right)\left(1 - T'\left(y\right)\right) + U_y\left(y - T\left(y\right), y, \theta\right) = 0.$$

and expressing it as a function of θ, we can totally differentiate the LHS and the RHS with respect to θ to obtain:

$$\begin{aligned} &\left[U_{cc}\left(y - T\left(y\right), y, \theta\right)\left(1 - T'\left(y\right)\right) y' + U_{cy} y'\left(\theta\right) + u_{c\theta}\right]\left(1 - T'\left(y\right)\right) \\ &\qquad - U_c\left(y - T\left(y\right), y, \theta\right) T''\left(y\right) y' + U_{yc}\left(1 - T'\left(y\right)\right) y' + U_{yy} y' + U_{y\theta} = 0. \end{aligned}$$

Simplify:

$$\begin{aligned} &U_{cc}\left(y - T\left(y\right), y, \theta\right)\left(1 - T'\left(y\right)\right)^2 y' + U_{cy} y'\left(\theta\right)\left(1 - T'\left(y\right)\right) \\ &\quad + U_{c\theta}\left(1 - T'\left(y\right)\right) - U_c\left(y - T\left(y\right), y, \theta\right) T''\left(y\right) y' + U_{yc}\left(1 - T'\left(y\right)\right) y' + U_{yy} y' + U_{y\theta} = 0. \end{aligned}$$

Solving for y' implies

$$y'\left(\theta\right) = -\frac{U_{c\theta}\left(1 - T'\right) + U_{y\theta}}{U_{cc}\left(1 - T'\right)^2 + 2U_{cy}\left(1 - T'\right) - U_c T'' + U_{yy}}. \tag{8}$$

Using equations (1), (4), and (7), in sequence

$$\begin{aligned} U_{c\theta}\left(1-T'\right)+U_{y\theta} &= -U_{c\theta}\frac{U_y}{U_c}+U_{y\theta} \\ &= U_{c\theta}e_y+U_{y\theta} \\ &= -U_c\frac{\partial MRS}{\partial\theta}. \end{aligned}$$

Since by equation (1) again $1-T'=MRS$,

$$-U_c\frac{\partial MRS}{\partial\theta}=-U_c\left(1-T'\right)\frac{\partial\log MRS}{\partial\theta}.$$

Therefore, equation (8) for $y'\left(\theta\right)$ can be rewritten

$$\begin{aligned} y'\left(\theta\right) &= -\frac{-U_c\left(1-T'\right)\frac{\partial\log MRS}{\partial\theta}}{U_{cc}\left(1-T'\right)^2+2U_{cy}\left(1-T'\right)-U_cT''+U_{yy}} \\ &= \frac{\frac{\partial\log MRS}{\partial\theta}}{\frac{U_{cc}}{U_c}\left(1-T'\right)+2\frac{U_{cy}}{U_c}-\frac{T''}{1-T'}+\frac{U_{yy}}{U_c(1-T')}}. \end{aligned}$$

Consulting equation (2) for $\frac{-1}{y\varepsilon_w^\star}$we see that

$$y'\left(\theta\right)=\frac{-\frac{\partial\log MRS}{\partial\theta}}{\frac{1}{y\varepsilon_w^\star}+\frac{T''}{1-T'}}. \tag{9}$$

3 Pareto Problem

We now tackle the pareto problem. We ignore the nondecreasing requirement and set up the Lagrangian

$$\begin{aligned} \mathcal{L}=\int_{\underline{\theta}}^{\overline{\theta}}\left[y\left(\theta\right)-e\left(v\left(\theta\right),y\left(\theta\right),\theta\right)\right]f\left(\theta\right)d\theta+\int_{\underline{\theta}}^{\overline{\theta}}\mu\left(\theta\right)\left(v'\left(\theta\right)-U_\theta\left(e\left(v\left(\theta\right),y\left(\theta\right),\theta\right),y\left(\theta\right),\theta\right)\right)d\theta \\ +\int_{\underline{\theta}}^{\overline{\theta}}\xi\left(\theta\right)\left(v\left(\theta\right)-\overline{v}\left(\theta\right)\right)d\theta. \end{aligned}$$

We intend to optimize with respect to $v\left(\theta\right)$, making it desirable to eliminate the $v'\left(\theta\right)$ from the Lagrangian. Integrating by parts,

$$\int_{\underline{\theta}}^{\overline{\theta}}\mu\left(\theta\right)v'\left(\theta\right)d\theta=\mu\left(\overline{\theta}\right)v\left(\overline{\theta}\right)-\mu\left(\underline{\theta}\right)v\left(\underline{\theta}\right)-\int_{\underline{\theta}}^{\overline{\theta}}\mu'\left(\theta\right)v\left(\theta\right)d\theta.$$

Substituting back into the original Lagrangian and replacing the multiplier $\xi(\theta)=\lambda(\theta)f(\theta)\frac{1}{\eta}$ on the minimal utility target constraints (to build 1 big integral density constraint)

$$\begin{aligned} \mathcal{L}=\int_{\underline{\theta}}^{\overline{\theta}}\left[y\left(\theta\right)-e\left(v\left(\theta\right),y\left(\theta\right),\theta\right)\right]f\left(\theta\right)d\theta+\mu\left(\overline{\theta}\right)v\left(\overline{\theta}\right)-\mu\left(\underline{\theta}\right)v\left(\underline{\theta}\right)-\int_{\underline{\theta}}^{\overline{\theta}}\mu'\left(\theta\right)v\left(\theta\right)d\theta \\ -\int_{\underline{\theta}}^{\overline{\theta}}\mu\left(\theta\right)U_\theta\left(e\left(v\left(\theta\right),y\left(\theta\right),\theta\right),y\left(\theta\right),\theta\right)d\theta+\frac{1}{\eta}\int_{\underline{\theta}}^{\overline{\theta}}\lambda\left(\theta\right)f\left(\theta\right)\left(v\left(\theta\right)-\overline{v}\left(\theta\right)\right)d\theta. \end{aligned}$$

Maximize wrt $y(\theta)$: The first order condition with respect to $y(\theta)$ for interior θ is

$$(1 - e_y) f - \mu [U_{\theta c} e_y + U_{\theta y}] = 0.$$

Combining equations (4) and (1) we see that $1 - e_y = \tau$ and from equation (7) directly $U_{\theta c} e_y + U_{\theta y} = -U_c \frac{\partial MRS}{\partial \theta}$. Substituting these relationships into the first order condition yields

$$\tau f = -\mu U_c \frac{\partial MRS}{\partial \theta}$$

or, equivalently,

$$\frac{\tau}{1-\tau} f = -\mu U_c \frac{\partial \log (MRS)}{\partial \theta}. \tag{10}$$

Observe from the first version of the equation that the sign of μ is the same as the sign of τ given the single crossing assumption ($\frac{\partial MRS}{\partial \theta} < 0$).

Maximize wrt $v(\theta)$: The first order condition for $v(\theta)$ for interior θ is

$$-e_v f - \mu' - \mu U_{\theta c} e_v + \xi = 0.$$

From equation (5), $e_v = U_c^{-1}$. Making this substitution and multiplying through by U_c,

$$-f - U_c \mu' - \mu U_{\theta c} + \xi = 0.$$

Since Lagrange multipliers must be weakly positive

$$-U_c \mu' - \mu U_{\theta c} \le f. \tag{11}$$

Define our new multiplier $\hat{\mu}(\theta) = U_c (c(\theta), y(\theta), \theta) \mu(\theta)$. (converting units of multiplier into goods/dollars)

$$\hat{\mu}' = U_c \mu' + \mu (U_{cc} c' + U_{cy} y' + U_{c\theta}).$$

Rearranging terms

$$\begin{aligned} -U_c \mu' - \mu U_{\theta c} &= -\hat{\mu}' + \mu (U_{cc} c' + U_{cy} y') \\ &= -\hat{\mu}' + \hat{\mu} \frac{U_{cc} c' + U_{cy} y'}{U_c}. \end{aligned}$$

Substituting into the first order condition for $v(\theta)$, equation (11),

$$-\hat{\mu}' + \hat{\mu} \frac{U_{cc} c' + U_{cy} y'}{U_c} \le f.$$

Using equation (6) to replace the fraction we obtain

$$-\hat{\mu}' - \hat{\mu} \frac{\partial MRS}{\partial c} y' \le f.$$

Combining the conditions based on FOC's Thus we have the system of equations (with arguments for clarity)

$$-\hat{\mu}'(\theta) - \hat{\mu}(\theta) \frac{\partial MRS (c(\theta), y(\theta), \theta)}{\partial c} y'(\theta) \le f(\theta) \tag{12}$$

$$\frac{\tau(\theta)}{1 - \tau(\theta)} f(\theta) = -\hat{\mu}(\theta) \frac{\partial \log MRS (c(\theta), y(\theta), \theta)}{\partial \theta}. \tag{13}$$

As discussed in class, we don't know $f(\theta)$ but since $y(\theta)$ is nondecreasing we do know the output distribution (Saez identification) $H(y(\theta)) = F(\theta)$ and therefore $h(y(\theta))\,y'(\theta) = f(\theta)$. This allows us to write the optimality condition that holds with equality, equation (13), as

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

Substituting from equation (9) for $y'(\theta)$,

$$\frac{\tau(\theta)}{1-\tau(\theta)} h(y(\theta))\,\frac{-\frac{\partial \log MRS}{\partial \theta}}{\frac{1}{y\varepsilon_w^\star} + \frac{T''}{1-T'}} = -\hat{\mu}(\theta)\,\frac{\partial \log MRS}{\partial \theta}.$$

Canceling the derivatives of $\log MRS$ and multiplying by -1,

$$\frac{\tau(\theta)}{1-\tau(\theta)} h(y(\theta))\,\frac{1}{\frac{1}{y\varepsilon_w^\star} + \frac{T''}{1-T'}} = \hat{\mu}(\theta).$$

Define the virtual density to be

$$h^\star(y) = \frac{h(y)}{1 + y\varepsilon_w^\star \frac{T''}{1-T'}}. \tag{14}$$

Then the condition becomes

$$\frac{\tau(\theta)}{1-\tau(\theta)} y\varepsilon_w^\star h^\star(y(\theta)) = \hat{\mu}(\theta).$$

Define $\hat{\hat{\mu}}(y) = \hat{\mu}\left(y^{-1}(y)\right)$ to change variables from θ to y. Then the condition is

$$\frac{T'(y)}{1-T'(y)} y\varepsilon_w^\star(y)\,h^\star(y) = \hat{\hat{\mu}}(y). \tag{15}$$

Written in logs

$$\log \hat{\hat{\mu}}(y) = \log\left(\frac{T'(y)}{1-T'(y)}\right) + \log(y) + \log(\varepsilon_w^\star(y)) + \log(h^\star(y)),$$

and therefore (differentiating with respect to $\log y$)

$$\frac{\hat{\hat{\mu}}'(y)}{\hat{\hat{\mu}}(y)} y = \frac{d\log\left(\frac{\tau}{1-\tau}\right)}{d\log y} + 1 + \frac{d\log(\varepsilon_w^\star(y))}{d\log y} + \frac{d\log(h^\star(y))}{d\log y}. \tag{16}$$

Returning to the optimality condition holding with inequality, equation (12), and dividing through by $y'(\theta)$,

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

The RHS is then seen to be $h(y(\theta))$. Using ' that the derivative of the inverse is the inverse of derivative'-rule, we can differentiate the definition of $\hat{\hat{\mu}}(\theta) = \hat{\mu}(Y^{-1}(y))$:

$$\hat{\hat{\mu}}'(y) = \frac{\hat{\mu}'\left(y^{-1}(y)\right)}{y'\left(y^{-1}(y)\right)}.$$

Thus the inequality condition can be rewritten

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

Substituting via the definition of the virtual density, equation (14),

$$-\hat{\hat{\mu}}'(y) - \hat{\hat{\mu}}(y)\,\frac{\partial MRS}{\partial c} \leq h^\star(y)\left(1 + y\varepsilon_w^\star \frac{T''}{1-T'}\right).$$

Dividing through by $\hat{\hat{\mu}}$ and multiplying by y

$$-\frac{\hat{\hat{\mu}}'(y)}{\hat{\hat{\mu}}}y - \frac{\partial MRS}{\partial c}y \leq \frac{h^\star(y)\left(1 + y\varepsilon_w^\star \frac{T''}{1-T'}\right)}{\hat{\hat{\mu}}}y.$$

Finally, using equation (15) for $\hat{\hat{\mu}}$and equation(16) for the first term,

$$-\frac{d\log\left(\frac{\tau}{1-\tau}\right)}{d\log y} - 1 - \frac{d\log\left(\varepsilon_w^\star(y)\right)}{d\log y} - \frac{d\log\left(h^\star(y)\right)}{d\log y} - \frac{\partial MRS}{\partial c}y \leq \frac{yh^\star(y)\left(1 + y\varepsilon_w^\star \frac{T''}{1-T'}\right)}{\frac{\tau(\theta)}{1-\tau(\theta)}y\varepsilon_w^\star h^\star(y(\theta))}.$$

This simplifies to

$$\frac{\tau(\theta)}{1-\tau(\theta)}\frac{\varepsilon_w^\star}{1 + y\varepsilon_w^\star\frac{T''}{1-T'}}\left(-\frac{d\log\left(\frac{\tau}{1-\tau}\right)}{d\log y} - 1 - \frac{d\log\left(\varepsilon_w^\star(y)\right)}{d\log y} - \frac{d\log\left(h^\star(y)\right)}{d\log y} - \frac{\partial MRS}{\partial c}y\right) \leq 1,$$

4 Intuition: Laffer argument:

- Intuition for each term/Laffer argument with following experiment:
  - Draw (Y, T) income-taxation system and draw the tax schedule
  - Reduce the tax locally at a point P
  - From the left of the point P, the "climbers" will work more, earn more and pay more taxes
  - From the left of the point P, the "fallers" will work less, earn less and pay fewer taxes
- High ϵ_w^* reduces $\tau(\theta)$:
  - trivial: larger behavioral response from a tax increase and higher probability to reduce tax income by increasing taxes
- A high convexity of the tax schedule increases taxes:
  - increases $\frac{d\log\frac{\tau(\theta)}{1-\tau(\theta)}}{d\log y}$
  - increases $\tau(\theta)$:
  - "tax cuts are costly because big tax revenue losses per fallers"
- A strong increase of -the elasticity of earnings with respect to the after tax wage- in income $\frac{d\log(\epsilon_w^*(y))}{d\log y}$ increases $\tau(\theta)$:
  - "tax cuts are costly because elastic red fallers will be attracted by lower tax rate, start to work less and hence reduce the tax base
- A fat tail/slow decline in the earnings density/low absolute value of the slope (i.e. $\frac{d\log(h^*(y))}{d\log y} > 0$ is small) leads to higher taxes $\tau(\theta)$:
  - "Tax cuts at the top are costly because many red fallers will start to work less and reduce the tax base" (fat tail: Piketty & Saez vs. lognormal tail: Mirrlees)
- A strong income effect on leisure/labor supply (i.e. $-\frac{\partial MRS}{\partial c} > 0$ is big) reduces taxes $\tau(\theta)$:
  - Strong income effects -> rich love leisure and thus need low taxes to incentivize them to work

14.471 Public Economics I
Fall 2012