Lecture 9: Optimal Taxation

Miguel Almunia

MSc in Economics - University of Warwick

Thursday, 6th March, 2014
Optimal Taxation

“Optimal” tax theory must combine lessons from excess burden and tax incidence:

- Optimal size of the pie? - **Efficiency**
- How is the pie distributed? - **Equity**

What is the best way to design taxes given equity and efficiency concerns?
Optimal Taxation

- **Efficiency perspective**: finance the govt through lump-sum taxation
  - Fixed amount per person regardless of individual characteristics
  - Does not induce behavioural responses, so no inefficiency

- **Equity perspective**: individual lump-sum taxes
  - Tax higher-ability individuals a larger lump sum

- **Problem**: we cannot observe individual abilities
  - Hence we tax outcomes, such as income or consumption
  - Creates distortions & inefficiency
Outline of Today’s Lecture

Traditional Optimal Taxation Theory:
- Ramsey (1927): Inverse elasticity rule for taxation of goods
- Mirrlees (1971): Optimal nonlinear income tax

Modern Optimal Taxation Theory:
- Saez (2001): optimal tax rate for top earners
  - Borrows some results from Diamond (1998)
- * Diamond & Saez (2011): survey of the literature
Traditional Approaches

- **Ramsey (1927):** linear tax systems
  \[ T(z) = t \cdot z \]
  - Rules out lump-sum taxes by assumption
  - Homogeneous agents

- **Merrlees (1971):** nonlinear tax systems
  \[ T(z) = f(z) \]
  - Permits lump sum taxes
  - Heterogeneity in agents' skills
Ramsey (1927): Tax Problem

Government levies taxes on goods in order to accomplish two goals:

1. Raise total revenue ($R$) of amount $E$
2. Minimize utility loss for agents in the economy

Representative individual

- No redistributive concerns
- As in efficiency analysis, assume that individual does not internalize the effect of $\tau_i$ on govt budget
Ramsey (1927): Tax Problem

United Kingdom, 1903-1930
Mathematician & Philosopher
Cambridge University
Ramsey (1927): Assumptions

1. No lump-sum or nonlinear taxes - Only proportional taxes
2. Cannot tax all commodities (e.g., leisure untaxed)
3. Production prices fixed (normalized to one):

\[ p_i = 1 \]

\[ \Rightarrow q_i = 1 + \tau_i \]

Note: full mathematical derivation in “extra slides” at the end
Ramsey (1927): Graphical Intuition
Taxation of elastic vs. inelastic goods
Ramsey (1927): Interpretation & Limitations

- **Intuition:** tax inelastic goods to minimize efficiency costs

- **Limitations:**
  - Does not take into account redistributive motives
  - Necessities usually more inelastic than luxuries
  - Thus, optimal Ramsey tax system is regressive
  - Restricted to linear tax systems
Application of Ramsey to Taxation of Savings

- Standard lifecycle model:
  \[
  \max \sum_t u_t (c_t) \quad \text{subj. to} \quad q_t c_t = W
  \]
  where \( q_t = (1 + \tau_t) p_t \)

- Consumption in each period treated like different commodities

- Let capital income tax be \( \theta \), levied on interest rate:
  \[
  q_t = \frac{1}{(1 + (1 - \theta) r)^t}
  \]
Application of Ramsey to Taxation of Savings

For any $\theta > 0$, implied optimal tax goes to infinity:

$$\frac{q_t}{p_t} = 1 + \tau_t = \left( \frac{1 + r}{1 + (1 - \theta) r} \right)^t$$

$$\lim_{t \to \infty} \tau_t = \infty$$

But Ramsey formula implies that optimal tax $\tau^*$ cannot be $\infty$ for any good.

Therefore, optimal capital income tax must be zero in long run (Judd 1985; Chamley 1986)
Zero Capital Taxation Result: Limitations

- The zero result within Ramsey model no longer holds when assumptions are relaxed, for example:
  - Allowing for progressive income taxation
  - Allowing for credit market imperfections
  - Finitely-lived agents with finite bequest elasticities
  - Allowing for myopic agents
Ramsey’s approach ignored equity-efficiency tradeoff

Miryylees (1971) incorporates behavioural responses:

1. Standard labour supply model
2. Individuals differ in ability $w$
3. Govt maximizes social welfare function (SWF) subj. to:
   - Budget constraint
   - Behavioural responses to taxation
Mirrlees (1971): Incorporating Behavioural Responses

Scotland, 1936
Professor of Economics
Cambridge University
1996 Nobel Prize Winner
Mirrlees (1971): Model

- Individual maximizes:
  \[ u(c, l) \text{ subj. to } c = wl - T(wl) \]

- Ability distribution given by density \( f(w) \)

- Government maximizes:
  \[
  \text{SWF} = \int G[u(c, l)] f(w) \, dw
  \]
  subj. to budget \( \int T(wl) f(w) \, dw \geq E \)
  subj. to indiv FOC \( w (1 - T'(\cdot)) u_c - u_l = 0 \)

where \( G(\cdot) \) is increasing and concave
Mirrlees (1971): Model

- Govt maximizes \textit{weighted} sum of utilities of ex-post consumption
- With equal weights and diminishing marginal utility, we would equalize everyone’s income
  - Utilitarianism leads to comunism!
- Is maximizing total ex-post utility the right objective function?
  - Deep debate dating back to Rawls, Nozick, Sen...
Mirrlees (1971): Results

- Mirrlees formulas are complicated, only a few general results:
  1. $T'(\cdot) \leq 1$: Obvious, because otherwise no one works
  2. $T'(\cdot) \geq 0$: Non-trivial. Rules out EITC (incentives for labour force participation with tax subsidies)
  3. $T'(\cdot) = 0$ at the bottom of the skill distribution (assuming everyone works)
  4. $T'(\cdot) = 0$ at the top of the skill distribution (if skill distribution is bounded)
MIRRELES (1971): Results

- Mirrlees model had huge impact in fields like contract theory
  - Models with asymmetric information
- But little impact on **practical tax policy**
- Recently, connected to empirical tax literature:
  - Diamond (AER, 1998), Saez (REStud, 2001)
  - Sufficient statistic formulas in terms of elasticities
Optimal Income Taxation: Sufficient Statistic Formulas

- Revenue-maximizing linear tax: Laffer curve
- Top income tax rate (Saez 2001)
- Overview of this literature: *Diamond and Saez (JEL, 2011)
Laffer Curve: Revenue Maximizing Rate

- Useful benchmark for optimal rate
- Let tax revenue be $R(\tau) = \tau \cdot z(t - \tau)$
- Notice: reported income $z$ is a function of net-of-tax rate $(1 - \tau)$
- $R(\tau)$ has an inverse U-shape:
  - No taxes: $R(\tau = 0) = 0$
  - Confiscatory taxes: $R(\tau = 1) = 0$
Laffer Curve: Revenue Maximizing Rate

- Revenue maximizing rate, $\tau^*$:

$$R'(\tau^*) = 0$$

$$z - \tau \frac{dz}{d(1 - \tau)} = 0$$

$$z \left[ \frac{(1 - \tau)}{z} \right] - \tau \epsilon \frac{dz}{d(1 - \tau)} \left[ \frac{(1 - \tau)}{z} \right] = 0$$

$$1 - \tau - \tau \epsilon = 0$$

$$\Rightarrow \tau^* = \frac{1}{1 + \epsilon}$$

- Strictly inefficient to have $\tau > \tau^*$ (Why?)
Laffer Curve: Revenue Maximizing Rate

Graphical representation
Derive optimal tax rate $\tau$ using perturbation argument

Assume there are no income effects: $\varepsilon^C = \varepsilon^U = \varepsilon$

- Diamond (1998) shows this is a key theoretical simplification

Assume that there are $N$ individuals above earnings $\bar{z}$

- Let $z^m (1 - \tau)$ be average income function of these individuals
Saez (2001): Optimal Income Tax Rate

Three effects of small $\Delta \tau > 0$ reform above $\bar{z}$:

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

$$\Delta M = N \cdot [z^m - \bar{z}] \Delta \tau$$

2. **Behavioural response**:

$$\Delta B = N \tau \Delta z^m = N \tau \left( - \Delta \tau \frac{\Delta z^m}{\Delta (1 - \tau)} \right)$$

$$= -N \frac{\tau}{1 - \tau} \bar{e} z^m \Delta \tau$$

3. **Welfare effect**: if govt values rich consumption at $\bar{g} \in (0, 1)$:

$$\Delta W = -\bar{g} \, dM$$
Saez (2001): Optimal Income Tax Rate

Source: Diamond and Saez (JEL, 2011)
Saez (2001): Optimal Income Tax Rate

- Optimal tax rate solves:

\[ \Delta M + \Delta W + \Delta B = 0 \]

- After some algebra:

\[ \frac{\tau^{*\text{top}}}{1 - \tau^{*\text{top}}} = \frac{(1 - \bar{g}) \left[ (z^m/\bar{z}) - 1 \right]}{\bar{\varepsilon}z^m/\bar{z}} \]

- Top tax rate \( \tau^{*\text{top}} \) is higher when:

  - \( \downarrow \bar{g} \): less weight on welfare of the rich
  - \( \downarrow \bar{\varepsilon} \): lower elasticity of taxable income
  - \( \uparrow z^m/\bar{z} \): higher income inequality
Saez (2001): Optimal Income Tax Rate

- With Pareto distribution, $z^m / \bar{z}$ converges to $\left(\frac{a}{a-1}\right)$, where $a$ is Pareto distn. parameter
- Simplified formula is then:

$$
\tau^{*\text{top}} = \frac{1 - \bar{g}}{1 - \bar{g} + \varepsilon a}
$$

- In the US, $z^m / \bar{z} \approx 3$ and quite stable over time
  - So Pareto distn. parameter is $\frac{z^m}{\bar{z}} = 3 = \frac{a}{a-1} \Rightarrow a = 1.5$
- We can estimate $\varepsilon$
- Society decides value of $\bar{g}$ (relative weight of rich on SWF)
US Income Distribution approximated by Pareto Distribution

Source: Diamond and Saez (JEL, 2011)
Lecture 9: Optimal Taxation

Zero Top Rate with Bounded Skill Distribution

- Suppose top earner make $z^T$, second makes $z^S$.
- Tax only raised on top earner, so $z^m = z^T$:
  \[
  \Delta M = N \Delta \tau \left[ z^m - \bar{z} \right] \to 0 < \Delta B = N \Delta \tau \cdot \bar{\epsilon} \cdot \frac{\tau}{1 - \tau} z^m
  \]

- Optimal $\tau = 0$ for top earner
  - Standard Mirrleesian result
- But the result applies **only** to the top earner
  - No practical relevance
Connection to Revenue Maximizing Tax Rate

- Revenue-maximizing top tax rate can be calculated by setting $\bar{g} = 0$
  - Utilitarian SWF: $\bar{g} = u_c (z^m) \rightarrow 0$ when $\bar{z} \rightarrow \infty$
  - Rawlsian SWF: $\bar{g} = 0$ for any $\bar{z} > \min(z)$
- If $\bar{g} = 0$, we obtain $\tau^{\text{top}} = \tau^{\text{max}} = \frac{1}{1+\varepsilon a}$
- Assuming $a = 1.5$:

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>0.2</th>
<th>0.5</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{g} = 0$</td>
<td>0.77</td>
<td>0.57</td>
<td>0.40</td>
</tr>
<tr>
<td>$\bar{g} = 0.5$</td>
<td>0.62</td>
<td>0.40</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Mathematical Derivation of Ramsey (1927)

Note: the following slides draw heavily on Raj Chetty’s lecture slides for his Public Economics course at Harvard University, available at www.rajchetty.com
Ramsey (1927): Individual’s problem

- Individual maximizes utility over goods (leisure untaxed):

$$\max_{\{x_1, \ldots, x_N, l\}} u(x_1, \ldots, x_N, l)$$

subj. to

$$q_1 x_1 + \ldots + q_N x_N = wl + Y$$

- Lagrangian:

$$\mathcal{L} = u(x_1, \ldots, x_N, l) + \alpha (wl + Y - q_1 x_1 - \ldots - q_N x_N)$$

- Solve and obtain indirect utility $V(q, Y)$
Ramsey (1927): Government’s problem

- Then government maximizes welfare of representative individual:

\[
\max \ V (q, Y)
\]

subject to the tax revenue requirement:

\[
\sum_{i=1}^{N} \tau_i x_i (q, Y) \geq E
\]

- Government’s Lagrangian:

\[
\mathcal{L} = V (q, Y) + \lambda \left[ E - \sum_{i=1}^{N} \tau_i x_i (q, Y) \right]
\]
Ramsey (1927): Government’s problem

\[ \mathcal{L} = V(q, Y) + \lambda \left[ E - \sum_{i=1}^{N} \tau_i x_i (q, Y) \right] \]

\[ \Rightarrow \frac{\partial \mathcal{L}}{\partial q_i} = \frac{\partial V}{\partial q_i} + \lambda \left[ x_i \text{ Mech. Effect} + \sum_j \tau_j \frac{\partial x_j}{\partial q_i} \right] \]

Using Roy’s Identity \( \left( \frac{\partial V}{\partial q_i} = -\alpha x_i \right) \):

\[ (\lambda - \alpha) x_i + \lambda \sum \tau_j \frac{\partial x_j}{\partial q_i} = 0 \]
Ramsey (1927): Perturbation Argument
Equivalent but more intuitive than original Ramsey analysis

- Effect of tax increase \((\tau_i \text{ to } \tau_i + d\tau_i)\) on social welfare:
  - Marginal effect on govt revenue:
    \[
    dR = x_i d\tau_i + \sum_j \tau_j dx_j
    \]
    mech. effect behav. response
  - Marginal effect on private surplus:
    \[
    dU = \frac{\partial V}{\partial q_i} d\tau_i = -\alpha x_i d\tau_i
    \]
  - Optimum characterized by balancing the two marginal effects:
    \[
    dU + \lambda dR = 0
    \]
Ramsey (1927): Equilibrium

After some algebra, arrive at the following expression for the optimal tax rates, $\tau_j^*$:

$$\sum_j \tau_j \frac{\partial x_j}{\partial q_i} = -\frac{x_i}{\lambda} (\lambda - \alpha)$$

- $N$ equations in $N$ unknowns
- Notice the relevance of cross-price elasticities
- $\lambda = \text{Lagrange multiplier in the govt’s problem}$
- $\alpha = \text{Lagrange multiplier in the individual’s problem (mgl utility of money)}$
Ramsey (1927): Compensated Elasticities

- Starting from the perturbation argument formula:

\[ dU + \lambda dR = 0 \]

- Rewrite in terms of Hicksian elasticities and use Slutsky equation:

\[ \frac{\partial x_j}{\partial q_i} = \frac{\partial h_j}{\partial q_i} - x_i \frac{\partial x_j}{\partial Y} \]

- Combining the two equations:

\[ (\lambda - \alpha) x_i + \lambda \sum_j \tau_j \left( \frac{\partial h_j}{\partial q_i} - x_i \frac{\partial x_j}{\partial Y} \right) = 0 \]

\[ \Rightarrow \frac{1}{x_i} \sum_j \tau_j \frac{\partial h_j}{\partial q_i} = -\frac{\theta}{\lambda} \]
Ramsey (1927): Compensated Elasticities

- $\theta$ is independent of $i$ and measures the value for the govt of introducing a $1$ lump sum tax:

$$\theta = \lambda - \alpha - \lambda \frac{\partial \sum_j \tau_j x_j}{\partial Y}$$

- Three effects of introducing a $1$ lump sum tax:
  1. Direct value of $\lambda$ for the govt
  2. Loss in welfare of $\alpha$ for the individual
  3. Behavioural effect $\rightarrow$ Loss in tax revenue: $\frac{\partial \sum_j \tau_j x_j}{\partial Y}$
Ramsey (1927): Index of Discouragement

\[ \frac{1}{x_i} \sum_j \tau_j \frac{\partial h_j}{\partial q_i} = \frac{\theta}{\lambda} \]

- Suppose revenue requirement \( E \) is small, so all taxes are small
- Then tax \( \tau_j \) on good \( j \) reduces consumption of good \( i \) (holding utility constant) by approx

\[ dh_i = \tau_j \frac{\partial h_i}{\partial q_j} \]

- Numerator of LHS (top equation): total reduction in cons of good \( i \)
Dividing by $x_i$ yields % reduction in consumption of each good $i$ – “index of discouragement” of the tax system on good $i$

Ramsey tax formula says the indices of discouragement must be equal across goods at the optimum
Ramsey (1927): Inverse Elasticity Rule

- Introducing elasticities, we can write Ramsey formula as:

\[ \sum_{j=1}^{N} \frac{\tau_j}{1 + \tau_j} \varepsilon_{ij} = \frac{\theta}{\lambda} \]

- \( \theta = \lambda - \alpha - \lambda \frac{\partial \sum j \tau_j x_j}{\partial Y} \) is the value for the govt of introducing a $1 lump sum tax

- Consider special case where \( \varepsilon_{ij} = 0 \) for all \( i \neq j \)

  - Slutsky matrix is diagonal
  - Obtain classic **inverse elasticity rule**:

\[ \frac{\tau_i}{1 + \tau_i} = \frac{1}{\varepsilon_{ii}} \frac{\theta}{\lambda} \]