I. Static Optimal Taxation

How should government set taxes?

Suppose that the government needs to raise a given amount of revenue to finance public programs and that it cares about social welfare.

The *Second Welfare Theorem* tells us that any Pareto efficient allocation can be achieved through lump-sum taxation.

Thus, raising revenues with lump sum taxes seems like a good idea.

However, lump sum taxes are a bit impractical.

This is because the government will not typically want everyone to pay exactly the same tax.

For example, under most notions of social welfare, those citizens with greater income generating ability should pay more.

But observing things like the ability to generate income is difficult.

In reality, therefore, taxes are a function of observables like earnings, value of property, consumption of goods.

But such taxes influence the economic decisions of citizens, which leads to distortions.

The question is then how best to raise revenue given these distortions.

This is a non-trivial problem and leads to the literature on optimal taxation.

The literature on static optimal taxation can be divided into three parts:

(i) optimal commodity taxation - this deals with the optimal linear taxation of the consumption of goods and services;

(ii) optimal income taxation - this deals with the optimal non-linear taxation of income;

(iii) optimal mixed taxation - this combines linear taxes on consumption goods with non-linear income taxation.

Optimal Commodity Taxation

The optimal commodity taxation literature tries to shed light on how tax rates should differ across different goods and services.

Before getting into the details, it is worthwhile reviewing what insights partial equilibrium analysis provides.

A standard partial equilibrium analysis of the problem would suggest that the deadweight loss from taxation will be lowest on goods which are supplied or demanded inelastically.

However, this analysis ignores equity concerns and interactions between commodities.

To capture these things we need to set up a more sophisticated model.

We will set up the simplest version of such a model and derive a version of the *Many Person Ramsey Tax Rule* - which is a formula that characterizes the optimal taxes.

We will then explain the *Production Efficiency Theorem* which is the second major result of optimal commodity taxation.

1. Many Person Ramsey Tax Rules

Model

The model is a general equilibrium model with a very simple linear production technology.

There are I consumers, indexed by $i = 1, ..., I$

There are n consumer goods, indexed by $j = 1, ..., n$

Denote labor by l

Consumer i's utility function is $u_i(x^i, l^i)$, where $x^i = (x_1^i, ..., x_n^i)$

Each good j is produced from labor with a linear technology, $x_j = l_j$

Assume competitive production so that the producer price of good j is $p_j = w$, where w is the wage rate.

Without loss of generality, let $w = 1$, which means that $p_j = 1$ for all j.

The government needs to hire T units of labor. Thus it needs T units of tax revenue (T is exogenous).

To raise the revenue, the government imposes linear taxes on goods $j = 1, ..., n$. These are $(t_1, ..., t_n)$.

The taxes raise consumer prices to $(1 + t_1, ..., 1 + t_n)$.

Note: There is no loss of generality in not considering a linear tax on labor income. A tax on labor income is equivalent to a uniform tax on all commodities.

Let $q = (1 + t_1, ..., 1 + t_n)$ denote the post-tax price vector.

Consumer i's problem given the price vector q and the wage w is

$$\max_{(x^i, l^i)} \{u_i(x^i, l^i) : q \cdot x^i \leq wl^i + R^i\}$$

where R^i denotes consumer i's non-labor income (which will be zero in the model).

Consumer i's indirect utility function is denoted $V_i(q, w, R^i)$ and his demand function is denoted $x^i(q, w, R^i)$.

Since $w = 1$ and $R^i = 0$, we will just write $V_i(q)$ and $x^i(q)$ in what follows.

The government's objective is to maximize the social welfare function $W(V_1(q), ..., V_I(q))$ where $\partial W/\partial V_i > 0$ for all i

The government's problem is:

$$\max_q W(V_1(q), ..., V_I(q)),$$

$$s.t. \sum_i^I \sum_j^n (q_j - 1) x_j^i(q) = T.$$

The Lagrangian for the problem is

$$L = W + \lambda [\sum_i^I \sum_j^n (q_j - 1) x_j^i(q) - T]$$

The first order conditions are

$$\sum_{i}^{I} \frac{\partial W}{\partial V_i} \frac{\partial V_i}{\partial q_k} = -\lambda \sum_{i}^{I} (x_k^i + \sum_{j}^{n} t_j \frac{\partial x_j^i}{\partial q_k}) \text{ for } k = 1, ..., n$$

By *Roy's Identity*,

$$\frac{\partial V_i}{\partial q_k} = -\alpha_i x_k^i$$

where α_i is the marginal utility of income of consumer i; that is,

$$\alpha_i = \frac{\partial V_i}{\partial R_i}$$

Let

$$\beta_i = \frac{\partial W}{\partial V_i} \alpha_i$$

this is interpreted as the "social marginal utility of income" of consumer i.

Using these definitions:

$$\sum_{i}^{I} \beta_i x_k^i = \lambda \sum_{i}^{I} (x_k^i + \sum_{j}^{n} t_j \frac{\partial x_j^i}{\partial q_k}) \text{ for } k = 1, ..., n$$

Recall the *Slutsky equation*:

$$\frac{\partial x_j^i}{\partial q_k} = s_{jk}^i - x_k^i \frac{\partial x_j^i}{\partial R_i}$$

where $s_{jk}^i = \frac{\partial x_j^i}{\partial q_k}|_{u=u_i}$ (the derivative of the *Hicksian* or *compensated* demand function).

We can now write:

$$\sum_{j=1}^{n} t_j \sum_{i=1}^{I} s_{jk}^i = \frac{\sum_{i} \beta_i x_k^i}{\lambda} - \sum_{i}^{I} x_k^i + \sum_{i}^{I} x_k^i \sum_{j}^{n} t_j \frac{\partial x_j^i}{\partial R_i}$$

Define

$$
\begin{aligned}
b_i &= \frac{\beta_i}{\lambda} + \sum_j^n t_j \frac{\partial x_j^i}{\partial R_i} \\
&= \frac{SMU \text{ of } i\text{'s income}}{\text{marginal value of revenue}} \\
&\quad + \text{Tax revenue consequences} \\
&= \text{Net } SMU \text{ of } i\text{'s income}
\end{aligned}
$$

Let $x_k = \sum_i^I x_k^i$ be the aggregate demand for good k.

Using the symmetry of the Slutsky matrix, we get

$$
\sum_j^n t_j \sum_i^I s_{kj}^i = -x_k\left(1 - \sum_i^I b_i \frac{x_k^i}{x_k}\right)
$$

Now denote

$$
\bar{b} = \left(\sum_i b_i\right)/I
$$

and let

$$\begin{aligned}\theta_k &= \sum_i (\frac{b_i}{\bar{b}} - 1)(\frac{I x_k^i}{x_k} - 1)/I \\ &= cov(\frac{b_i}{\bar{b}}, \frac{I x_k^i}{x_k})\end{aligned}$$

Dividing through the first order condition by x_k we have that

$$-\frac{\sum_{j=1}^n t_j \sum_{i=1}^I s_{kj}^i}{x_k} = 1 - \sum_{i=1}^I b_i \frac{x_k^i}{x_k}$$

Now note that

$$\begin{aligned}\sum_{i=1}^I b_i \frac{x_k^i}{x_k} &= \frac{1}{x_k} \sum_{i=1}^I b_i x_k^i \\ &= \frac{I}{x_k} \left(\left[\frac{1}{I} \sum_{i=1}^I b_i x_k^i - \bar{b} \frac{x_k}{I} \right] + \bar{b} \frac{x_k}{I} \right)\end{aligned}$$

$$
\begin{aligned}
&= \frac{I}{x_k}\left(cov\left(b_i, x_k^i\right) + \bar{b}\frac{x_k}{I}\right) \\
&= (\frac{I}{x_k})(\frac{\bar{b}x_k}{I})\left(cov\left(\frac{b_i}{\bar{b}}, \frac{Ix_k^i}{x_k}\right) + 1\right) \\
&= \bar{b}\left(cov\left(\frac{b_i}{\bar{b}}, \frac{Ix_k^i}{x_k}\right) + 1\right)
\end{aligned}
$$

For the third step, note that in general

$$
cov\left(x_i, z_i\right) = E(x_i z_i) - E(x_i)E(z_i).
$$

Then we get

$$
-\frac{\sum_j t_j \sum_i s_{kj}^i}{x_k} = 1 - \bar{b} - \bar{b}\theta_k
$$

which tells us

Discouragement index of good k

= Distributive factor of good k

This condition is the *Many Person Ramsey Tax Rule* and was first derived by Diamond in a 1975 paper in *Journal of Public Economics*.

The discouragement index of good k is to be interpreted as the relative reduction in compensated demand for good k resulting from the tax system.

When all goods have the same θ_k, then the tax rule says that the discouragement index should be constant across goods.

When goods have different θ_k's, then goods with high θ_k should be discouraged the least.

The formula is insightful because it separates out distributional and efficiency effects.

The Case of Identical Consumers

Assume that all consumers are identical in the sense that they have a common utility function $u(x, l)$ and identical endowments of labor.

In this case, $b_i = \bar{b}$ for all i. This means that $\theta_k = 0$ for all k and hence that

$$-\frac{\sum_j t_j \sum_i s^i_{kj}}{x_k} = 1 - \bar{b}$$

Thus, in an optimal system, the discouragement index should be constant across goods.

This formula is known as the *Ramsey Rule* because it was first derived by Frank Ramsey in 1927. Ramsey was the first to pose and solve the optimal commodity taxation problem.

Two goods

Now consider the two good case $n = 2$. Then, it can be shown that

$$t_1 - t_2 = -\frac{1-\bar{b}}{D} x_1 x_2 [\varepsilon_{1l} - \varepsilon_{2l}]$$

where D is the determinant of the Slutsky sub-matrix for the two goods

$$D = s_{11}s_{22} - s_{12}^2 > 0$$

and ε_{il} is the compensated elasticity of demand for good i with respect to a change in the wage rate; that is

$$\begin{aligned} \varepsilon_{il} &= s_{il}/x_i \\ &= \frac{w \left. \frac{\partial x_i}{\partial w} \right|_u}{x_i}. \end{aligned}$$

We conclude that if the cross elasticities are the same, then the taxes are the same.

On the other hand, if $\varepsilon_{1l} < \varepsilon_{2l}$, then $t_1 > t_2$ (since $1 > \bar{b}$).

If $\varepsilon_{1l} < \varepsilon_{2l}$ then a wage increase leads to a greater increase in the compensated demand for good 2 than good 1.

Thus, good 1 must be taxed at a higher rate than good 2 if when the consumer works more he increases his consumption of good 2 more than good 1.

This result was obtained by Corlett and Hague (1953 *REStud*).

The interpretation is that when preferences are not seperable between consumption goods and leisure, the government should deviate from uniform taxation by taxing more heavily goods that are complementary with leisure (skis) than the goods that are complementary to labor (child care).

Uniform taxation

More generally, Deaton (1981 *Econometrica*) presents conditions for uniform taxation to be optimal - the requirement is that all goods are *quasi-separable* from leisure.

This is a complex condition - see Deaton's paper for definition.

It turns out that a utility function is quasi-separable if and only if its expenditure function can be written as

$$e(u, q, w) = e^*(u, b(u, q), w)$$

for some real valued function $b(u, q)$.

This is a stringent condition which is not likely to be satisfied in reality.

Thus, the bottom line is that uniform taxation of commodities is highly unlikely to be optimal which is an important finding in its own right.

Inverse Elasticity Rule

A really clean result emerges only when there are no cross-price effects ($s^i_{kj} = 0$ for all $k \neq j$).

Then

$$\frac{t_k}{1+t_k} = \frac{1-b}{\varepsilon_k}$$

where $\varepsilon_k = -s_{kk} q_k / x_k$.

To see this, note that

$$-\frac{\sum_j t_j \sum_i s^i_{kj}}{x_k} = -\frac{t_k s_{kk}}{\widetilde{x}_k}$$

where $\widetilde{x}_k$ is the individual demand for good k.

Thus, recalling that $q_k = 1 + t_k$, we have that

$$-\frac{t_k q_k s_{kk}}{\widetilde{x}_k (1+t_k)} = 1 - b$$

which implies the result.

This condition is called *Inverse Elasticity Rule*. It implies tax rates should be inversely proportional to demand elasticities.

This is because, as our partial equilibrium suggested, when demand for a good is elastic, the deadweight loss from taxing the good is greater, so better not to tax on this good.

This condition is satisfied, for example, when

$$u(x,l) = \sum_j \varphi_j(x_j) - l.$$

2. Production Efficiency Theorem

Diamond and Mirrlees (1971) consider the optimal commodity tax problem for a more general production technology.

They showed that if production exhibits constant returns to scale or production exhibits decreasing returns to scale and profits are taxed at 100%, then the optimal tax system always maintains the economy on the boundary of its production possibilities frontier.

This is known as the *Production Efficiency Theorem*.

They also provided optimal tax formulas for this more general environment, which look very similar to those we just derived.

The Production Efficiency Theorem means that the marginal rate of technical substitution between any two given inputs is the same in all productive units that use them.

This means that the use of non-uniform taxes on productive factors is non-optimal.

It also implies that intermediate goods (e.g., steel) should not be taxed.

This result was regarded as very surprising at the time, since the prior literature on the *Theory of the Second Best* had suggested that anything might be optimal if the first best was not achievable (see Lipsey and Lancaster 1956).

Model

To understand the result, consider an economy with I consumers, indexed by $i = 1, ..., I$

There are n commodities, indexed by $k = 1, ..., n$

Let ω^i denote consumer i's endowment of the commodities.

Let z^i denote consumer i's consumption vector and let $x^i = z^i - \omega^i$ denote his net consumption vector.

Let $u_i(z^i)$ denote consumer i's utility function and let q denote the vector of consumer prices.

There are J firms indexed by $j = 1, ..., J$

Firm j's production possibility set is Y_j and assume it is convex.

Let $Y = \sum_j Y_j$ be the aggregate production possibility set.

Let p denote the vector of producer prices. The vector of commodity taxes is given by $t = q - p$

Let the government's requirements of the n commodities be denoted by the vector g

Since there are either no profits or government taxes all the profits away, consumer i's problem is

$$\max_{z^i} u_i(z^i) \ s.t. \ qz^i \leq q\omega^i$$

This is equivalent to

$$\max_{x^i} u_i(x^i + \omega^i) \ s.t. \ qx^i \leq 0$$

Let $x^i(q)$ denote the solution to this problem (i.e., the consumer's demand function) and assume that it is well-defined and continuous.

Let $v_i(q)$ denote consumer i's indirect utility; i.e.,

$$v_i(q) = u_i(x^i(q) + \omega^i)$$

Let the government's social welfare function be

$$W(v_1(q), ..., v_I(q))$$

and assume that $\partial W / \partial v_i > 0$ for all i.

Analysis

To understand the result, suppose first that the government can choose both the consumer price vector q and all the firms' production vectors $(y_j)_{j=1}^J$

Then the government's problem is

$$\begin{array}{c} \max_{\{q,(y_j)_{j=1}^J\}} W(v_1(q), ..., v_I(q)) \\ s.t. \quad \sum_i x^i(q) + g \leq \sum_j y_j \end{array}$$

Proposition *If* $\{q^*, (y_j^*)_{j=1}^J\}$ *solves this problem, then*

$$y^* = \sum_j y_j^* \in bdyY.$$

Sketch of Proof: Suppose to the contrary that $y^* \in intY$. Choose a consumption good k for which all consumers have a positive net demand (e.g., ice cream). Then, for all consumers $i = 1, ..., I$

$$\partial v_i(q^*)/\partial q_k < 0$$

Since

$$\sum_i x^i(q^*) + g \in intY$$

the planner can reduce q_k slightly below q_k^* and

$$\sum_i x^i(q_{-k}^*, q_k) + g$$

will remain in the interior of Y. Accordingly, the planner can choose the price vector (q_{-k}^*, q_k) and production vectors $(y_j)_{j=1}^J$ such that

$$\sum_i x^i(q_{-k}^*, q_k) + g = \sum_j y_j$$

and make each consumer better off - a contradiction. ■

We now want to generalize this to the case in which the government simply levies a vector of commodity taxes t (which are paid by consumers), taxes any profits at 100% and uses these revenues to purchase g from producers.

It turns out that the government can achieve the same utility allocation for consumers as it could in the earlier problem, which implies that the aggregate production vector must be on the boundary of the production possibility frontier.

To understand this, let $\{q^*, (y_j^*)_{j=1}^J\}$ be the solution to the earlier problem.

Since $y^* = \sum_j y_j^* \in bdyY$ there must exist some producer price vector p^* such that

$$y^* \in \arg\max_{y \in Y} p^* \cdot y$$

or, equivalently, for all firms j

$$y_j^* \in \arg\max_{y_j \in Y_j} p^* \cdot y_j$$

Moreover, since

$$q^* \cdot \sum_i x^i(q^*) = 0$$

and

$$\sum_i x^i(q^*) + g = \sum_j y_j^*,$$

we have that

$$\begin{aligned} 0 &= q^* \cdot \sum_i x^i(q^*) \\ &= q^* \cdot \sum_i x^i(q^*) - p^* \cdot \sum_j y_j^* + p^* \cdot \sum_j y_j^* \\ &= (q^* - p^*) \cdot \sum_i x^i(q^*) - p^* \cdot g + p^* \cdot \sum_j y_j^* \end{aligned}$$

which implies that

$$p^* \cdot g = (q^* - p^*) \cdot \sum_i x^i(q^*) + p^* \cdot \sum_j y_j^*.$$

Thus, if the government were to announce the tax vector $t^* = q^* - p^*$, tax profits at 100% and purchase the vector g from producers with the tax revenues, then there would exist a competitive equilibrium with consumer and producer prices q^* and p^* in which consumer i consumes $x^i(q^*)$, firm j produces y_j^* and the government's budget is balanced!

Thus, t^* is the optimal tax vector and the resulting aggregate production bundle is on the boundary of Y which implies production efficiency.

Moreover, the optimal tax vector satisfies Ramsey Rule like formulas.

Notice that the Production Efficiency Theorem does not necessarily hold when there are decreasing returns and the government cannot tax profits at 100%.

Nor does it necessarily hold when there are increasing returns.

Implications

The Production Efficiency Theorem implies that Intermediate Goods should not be taxed.

Intermediate goods are goods that are neither direct inputs nor outputs for individual consumption.

Consider an economy with two goods x and y and labor l.

Good y is produced from labor with the technology $y = l_y$ where l_y is the labor used in y production.

Good x is produced from labor and good y with the CRS technology $x = f(y, l_x)$.

Consumer i has utility function $u_i(x^i, l^i)$ implying that good y is an intermediate good.

Let $w = 1$, so that $p_y = 1$ and suppose the government levies taxes $t_x > 0$ and $t_y > 0$.

Then firms in the x industry solve

$$\max p_x f(y, l_x) - l_x - (1 + t_y)y$$

This means that

$$p_x \frac{\partial f(y, l_x)}{\partial y} = (1 + t_y)$$

and

$$p_x \frac{\partial f(y, l_x)}{\partial l_x} = 1$$

Aggregate production will be inefficient - we can reduce l_x by Δl and increase l_y and y by Δl, leading to a change in output of

$$\Delta x = \left(\frac{\partial f}{\partial y} - \frac{\partial f}{\partial l_x} \right) \Delta l > 0.$$

Thus, the Production Efficiency Theorem implies that it must be the case in an optimal system that $t_y = 0$.

The Production Efficiency Theorem also implies that non-uniform taxes on productive factors are not optimal.

By productive factors, I mean inputs that consumers are endowed with but do not consume directly (e.g., natural resources).

Consider an economy with two consumption goods x_1 and x_2 and two factors z_1 and z_2.

Consumers are endowed with the two factors but obtain utility only from the consumption goods so that $u_i(x_1^i, x_2^i)$.

Let the aggregate endowments of z_1 and z_2 be $\overline{z}_1$ and $\overline{z}_2$.

Good x_i is produced from the two factors with the CRS technology $x_i = f_i(z_1^i, z_2^i)$.

Let $p_{z_1} = 1$ and suppose the government levies taxes t_{x_1} and t_{x_2}.

Moreover, suppose the government levies a tax $t^1_{z_2}$ on the use of z_2 in industry 1 but not industry 2.

Then firms in the x_1 industry solve

$$\max p_{x_1} f_1(z^1_1, z^1_2) - z^1_1 - (p_{z2} + t^1_{z_2}) z^1_2$$

and firms in the x_2 industry solve

$$\max p_{x2} f_2(z^2_1, z^2_2) - z^2_1 - p_{z2} z^2_2$$

This means that

$$\frac{\partial f_1(z^1_1, z^1_2)/\partial z^1_1}{\partial f_1(z^1_1, z^1_2)/\partial z^1_2} \neq \frac{\partial f_2(z^2_1, z^2_2)/\partial z^2_1}{\partial f_2(z^2_1, z^2_2)/\partial z^2_2}$$

Aggregate production will be inefficient - we can rearrange inputs between the industries and increase both x_1 and x_2.

Thus, by the Production Efficiency Theorem, in an optimal system $t^1_{z_2} = 0$.

Notice, however, that if the government levies a tax t_{z_2} on the use of z_2 in both industry 1 and industry 2, this would not create a production inefficiency.

Optimal Income Taxation

As noted last lecture, the main difference between commodity (or indirect) taxation and income (or direct) taxation, is that the government can use non-linear tax schedules when taxing income.

Designing optimal non-linear tax schedules is obviously more complicated than designing linear tax schedules.

Mirrlees (1971) proposed a simple framework for thinking about the problem of optimal non-linear taxation.

The key insight is to assume that the government can observe citizens' incomes but not their income generating abilities.

The problem of optimal taxation can then be modeled as a mechanism design problem in which citizens announce their abilities to the government and receive a consumption-income bundle in exchange.

This schedule of consumption-income bundles implicitly defines the optimal non-linear income tax schedule.

The Mirrlees Model

There are a continuum of individuals.

There are two goods - consumption and leisure.

Individuals get utility from consumption x and work l according to the utility function $x - \varphi(l)$ where φ is increasing, strictly convex, and twice continuously differentiable.

Individuals are endowed with $\bar{l}$ units of time in each period. Assume that $\varphi'(0) = 0$ and that $\lim_{l \to \bar{l}} \varphi'(l) = \infty$.

Individuals differ in their income generating abilities.

An individual with income generating ability a earns income $y = al$ if he works an amount l.

There are a continuum of ability levels $[\underline{a}, \overline{a}]$.

Let $F(a)$ denote the fraction of individuals with ability less than or equal to a.

The economy also has a government. This government spends an amount G. While this spending does not directly impact individuals' utilities, the government must raise the revenue necessary to finance it.

The First Best

An *allocation* in this economy is described by $\{x(a), y(a)\}$.

To be *feasible* an allocation must satisfy the aggregate resource constraint

$$\int x(a)dF(a) + G \leq \int y(a)dF(a).$$

The utility of an individual with ability a under such an allocation is

$$U(a) = x(a) - \varphi(y(a)/a).$$

Assume that government seeks to maximize the social welfare function

$$W = \int \Psi(U(a))dF(a)$$

where $\Psi(\cdot)$ is increasing and concave.

The first best problem is

$$\begin{array}{c} \max_{\{x(\cdot),y(\cdot)\}} \int \Psi(U(a))dF(a) \\ s.t. \;\; \int x(a)dF(a) + G \leq \int y(a)dF(a). \;\; (R) \end{array}$$

The Lagrangian for this problem can be written as:

$$L = \int [\Psi(U(a)) + \mu(y(a) - x(a))]dF(a) - \mu G$$

Maximizing this "pointwise" with respect to $(x(a), y(a))$ yields the first order conditions

$$\Psi'(U(a)) = \mu$$

and

$$\Psi'(U(a))\frac{\varphi'(y(a)/a)}{a} = \mu.$$

Thus, for all a

$$a = \varphi'(y(a)/a)$$

and for all a and a'

$$U(a) = U(a').$$

The first condition says that individuals must work up until the point at which their marginal disutility of work equals their marginal product.

This means that higher ability types should work more.

The second condition says that all individuals should have the same utility level.

First Best Taxes

If the government can observe individuals' abilities, it can implement the first best allocation with a very simple tax system.

Individuals of ability $\overline{a}$ are required to pay a tax $T(\overline{a})$ and individuals of ability $a < \overline{a}$ are required to pay a tax $T(a)$ such that

$$y(a) - T(a) - \varphi(y(a)/a) = y(\overline{a}) - T(\overline{a}) - \varphi(y(\overline{a})/\overline{a}).$$

where $y(a)$ is the first best earnings level.

This implies that for all a

$$T(a) = T(\overline{a}) - [y(\overline{a}) - \varphi(y(\overline{a})/\overline{a}) - (y(a) - \varphi(y(a)/a))]$$

Observe that i) all individuals have the same utility level and ii) lower ability individuals pay lower taxes.

The tax $T(\overline{a})$ is set to satisfy the government's budget constraint; i.e.,

$$\int \left(T(\overline{a}) - \left[\begin{array}{c} y(\overline{a}) - \varphi(y(\overline{a})/\overline{a}) \\ -(y(a) - \varphi(y(a)/a)) \end{array} \right] \right) dF(a) = G$$

Given this system, individuals of ability a will solve

$$\max\{y - T(a) - \varphi(y/a) : y/a \in [0, \bar{l}]\}$$

Clearly, individuals of ability a will choose to earn $y(a)$ - the first best earnings level.

The Second Best

The first best allocation will not be implementable when the government is unable to observe individuals' income generating abilities and hence unable to impose ability-specific lump sum taxes.

Suppose for example that the government announced that all those who earned $y(a)$ would pay tax $T(a)$.

Then, since for all $a > \underline{a}$,

$$y(a) - T(a) - \varphi(y(a)/a) < y(\underline{a}) - T(\underline{a}) - \varphi(y(\underline{a})/a),$$

all inividuals would choose to earn $y(\underline{a})$ and the government's budget would not balance.

To account for this problem, we need to ensure that it is always in individuals' interests to claim the bundles intended for them.

Formally, we require that the allocation satisfy the following set of incentive constraints: for all a

$$x(a) - \varphi(y(a)/a) \geq x(a') - \varphi(y(a')/a) \text{ for all } a'$$

The second best problem is

$$\begin{gathered} \max_{\{x(\cdot),y(\cdot)\}} \int \Psi(U(a))dF(a) \\ s.t. \;\; \int x(a)dF(a) + G \leq \int y(a)dF(a) \quad (R) \\ \forall a \;\; x(a) - \varphi(y(a)/a) \geq x(a') - \varphi(y(a')/a) \;\; \forall a' \end{gathered}$$

The strict mechanism design interpretation is that the government asks people to report their abilities with the understanding that those reporting ability a get the bundle $(x(a), y(a))$.

Obviously, this is not the way the tax system works - the way it works is that people choose how much income to earn y and then pay taxes $T(y)$.

But the solution to the mechanism design problem implicitly defines an income tax system in the sense that individuals who earn $y(a)$ pay taxes $T(y(a)) = y(a) - x(a)$.

Sketch of Solution Procedure

The key to solving the second best problem is to simplify the incentive constraints.

Define the function

$$v(a'; a) = x(a') - \varphi(y(a')/a).$$

The interpretation is that this is type a's utility when he claims to be type a'.

Assuming that the allocation is differentiable, incentive compatibility implies that

$$\frac{\partial v(a; a)}{\partial a'} = 0$$

and

$$\frac{\partial^2 v(a; a)}{\partial a'^2} \leq 0.$$

Differentiating the first order condition, we get

$$\frac{\partial^2 v(a;a)}{\partial a'^2} + \frac{\partial^2 v(a;a)}{\partial a' \partial a} = 0.$$

Thus, the second order condition is equivalent to

$$\frac{\partial^2 v(a;a)}{\partial a' \partial a} \geq 0.$$

Now we have that

$$\frac{\partial v(a';a)}{\partial a'} = x'(a') - \varphi'(\frac{y(a')}{a})\frac{y'(a')}{a}$$

and

$$\frac{\partial^2 v(a;a)}{\partial a' \partial a} = y'(a)[\frac{\varphi'(y(a)/a)}{a^2} + \frac{\varphi''(y(a)/a)y(a)}{a^3}].$$

Since the term in the square brackets is positive, the second order condition is satisfied if and only if $y'(a) \geq 0$.

We conclude that *necessary conditions* for the allocation $\{x(\cdot), y(\cdot)\}$ to satisfy the incentive compatability constraints are that

$$y'(a) \geq 0$$

and

$$x'(a) - \varphi'(\frac{y(a)}{a})\frac{y'(a)}{a} = 0$$

It can also be shown that these are *sufficient conditions* (see, for example, Laffont (1989) *Economics of Uncertainty and Information*)

Thus, we can recast the problem as

$$\begin{gathered} \max_{\{x(\cdot),y(\cdot)\}} \int \Psi(U(a))dF(a) \\ s.t. \ \int x(a)dF(a) + G \leq \int y(a)dF(a). \quad (R) \\ \forall a \ \ x'(a) - \varphi'(y(a)/a)\frac{y'(a)}{a} = 0 \ \ \& \ \ y'(a) \geq 0 \end{gathered}$$

To further simplify, note that

$$x(a) = U(a) + \varphi(\frac{y(a)}{a})$$

and using the first order condition, we get

$$\begin{aligned} U'(a) &= x'(a) - \varphi'(\frac{y(a)}{a})[\frac{y'(a)a - y(a)}{a^2}] \\ &= \varphi'(\frac{y(a)}{a})\frac{y(a)}{a^2}. \end{aligned}$$

Thus, the problem can be restated as

$$\begin{gathered} \max_{\{x(\cdot),y(\cdot)\}} \int \Psi(U(a))dF(a) \\ s.t.\ \int [y(a) - U(a) - \varphi(\tfrac{y(a)}{a})]dF(a) \geq G \quad (R) \\ \forall a \;\; U'(a) = \varphi'(\tfrac{y(a)}{a})\tfrac{y(a)}{a^2} \;\; \& \;\; y'(a) \geq 0. \end{gathered}$$

This is a problem that can be solved via the techniques of optimal control theory - $U(a)$ is the state variable and $y(a)$ is the control variable.

Typically, the constraint that $y'(a) \geq 0$ is ignored initially and then conditions are found that guarantee that it is satisfied.

Let us now explain how to solve this optimal control problem. We will follow Diamond (1988) AER.

A textbook on how to solve optimal control problems is *Dynamic Optimization* by Kamien and Schwartz.

It is actually easier to let $l(a) = \frac{y(a)}{a}$ be the control variable, rather than $y(a)$.

With $U(a)$ being the state variable and $l(a)$ the control variable, the Hamiltonian for the problem is

$$\mathcal{H} = \{\Psi(U(a)) - \lambda[U(a) + \varphi(l(a)) - a \cdot l(a)]\} f(a) + \mu(a)\left[\varphi'(l(a))\frac{l(a)}{a}\right].$$

The first-order conditions are given by

$$\mu'(a) = -\frac{\partial \mathcal{H}}{\partial U(a)}$$

and

$$\frac{\partial \mathcal{H}}{\partial l(a)} = 0.$$

The first condition implies that

$$\mu'(a) = -\left\{\Psi'(U(a)) - \lambda\right\} f(a), \qquad (1)$$

and the second implies that

$$\begin{aligned} & \lambda\left[\varphi'(l(a)) - a\right] f(a) \\ = \; & \frac{1}{a}\mu(a)\left[\varphi''(l(a))\, l(a) + \varphi'(l(a))\right]. \end{aligned}$$

Rewrite this as

$$\begin{aligned} & \lambda\left[\frac{\varphi'(l(a))}{a} - 1\right] a f(a) \\ = \; & \frac{\varphi'(l(a))}{a}\left[\left(\frac{\varphi''(l(a))\, l(a)}{\varphi'(l(a))} + 1\right)\right]\mu(a). \qquad (2) \end{aligned}$$

Setting $\mu(\overline{a})$ equal to zero we can integrate $\mu'(a)$ in (1) to get

$$\mu(a) = \int_a^{\overline{a}} \left\{\Psi'(U(a)) - \lambda\right\} f(a)\, da.$$

Note also that for any individual of type a, we have that

$$1 - T'(y(a)) = \frac{\varphi'(l(a))}{a}.$$

Also let

$$\varepsilon(a) = \frac{a^2(1 - T'(y(a)))}{al(a)\varphi''(l(a))}$$

This is the elasticity of labor supply of an individual of type a; that is,

$$\varepsilon(a) = \frac{w}{l}\frac{dl}{dw}$$

where w is the net marginal wage which equals $a(1 - T'(y(a)))$.

Using these, we can rewrite (2) as

$$\begin{aligned} -\lambda T'(a)\,af(a) &= \big(1 - T'(a)\big)\left[\frac{1}{\varepsilon(a)} + 1\right] \\ &\quad \times \int_a^{\overline{a}} \big\{\Psi'(U(a)) - \lambda\big\}\, f(a)\,da \end{aligned}$$

or

$$\frac{T'(a)}{1-T'(a)} = \left[\frac{1}{\varepsilon(a)}+1\right]\frac{1-F(a)}{af(a)}\frac{1}{1-F(a)}$$
$$\times\frac{1}{\lambda}\int_{a}^{\overline{a}}\left\{\lambda-\Psi'(U(a))\right\}f(a)\,da.$$

Diamond argues (in the paragraph preceding Eqn 11 in his paper) that with quasi-linear preferences the Lagrange multiplier on the government budget constraint equals the average SMU,

$$\lambda = \int_{\underline{a}}^{\overline{a}}\Psi'(U(a))\,f(a)\;da.$$

Let

$$D(a) = \frac{1}{1-F(a)}\int_{a}^{\overline{a}}\Psi'(U(\alpha))f(\alpha)d\alpha$$

$D(a)$ is the average value of $\Psi'(U(\alpha))$ on the interval $[a,\overline{a}]$. Note that it is decreasing in a.

Then the previous equation can be rewritten as

$$\frac{T'(y(a))}{1-T'(y(a))} = [1 + \frac{1}{\varepsilon(a)}][\frac{1-F(a)}{af(a)}][1 - \frac{D(a)}{D(\underline{a})}].$$

This is not a closed form solution for the optimal marginal tax rates.

Note that $D(a)$ depends upon the optimal utility levels which are endogenous.

Nonetheless, the formula is useful and can be understood intuitively.

To understand the condition intuitively, consider the following perturbation:

Raise the marginal tax rate for individuals with income $y(a)$ (say, in a small interval $y(a)$ to $y(a) + \delta$ where δ is infinitessimal) leaving all other marginal tax rates unchanged.

To see the effects, note first that for any individual of type a', we have that

$$1 - T'(y(a')) = \frac{\varphi'(y(a')/a')}{a'}$$

and hence the labor supply distortion of a type a' just depends upon the marginal tax rate at his income level $y(a')$. The higher the marginal tax rate, the higher the distortion.

There are two effects of the perturbation. First, individuals with income levels in the treated interval will have their labor supply further distorted.

Second, since

$$T(y) = T(y(a)) + \int_{y(a)}^{y} T'(z)dz,$$

individuals with income levels above $y(a) + \delta$ will pay higher taxes, but they will face no additional distortion because their marginal tax rates are the same.

The first effect is a cost and the second effect is a benefit because the social welfare function values redistribution and because individuals with income levels above are $y(a) + \delta$ an above average slice of the population.

Roughly speaking, the higher is the benefit from the second effect relative to the cost of the first effect, the higher should be the marginal tax rate at $y(a)$.

The formula for the optimal marginal tax rates reflects this logic.

The higher the ratio $(1 - F(a))/af(a)$ the higher is benefit relative to cost because the distorted group is smaller relative to the group who pays more taxes.

The smaller the labor supply elasticity the lower is the cost, because the lower is the distortion.

The higher is $1 - \frac{D(a)}{D(\underline{a})}$, the higher is the benefit because the average value of $\Psi'(U(\alpha))$ on the interval $[a, \overline{a}]$ is lower.

What does the first order condition imply?

(i) Since the RHS is non-negative, $T'(y(a)) \in [0, 1)$

(ii) $T'(y(\underline{a})) = 0$ and $T'(y(\overline{a})) = 0$

The latter results assume that there are some upper and lower bounds to the ability distribution, as opposed to the support just being $[0, \infty)$

If some individual have zero ability (which seems plausible), they will not work and the zero marginal tax rate result does not hold.

The "no distortion at the top" result is a local result and does not imply that marginal rates near the top of the income distribution should be zero or near zero.

To get any further one has to simulate.

Simulations

There is a long tradition of simulating the optimal tax schedule, starting with the original Mirrlees paper.

One needs to specify the social welfare function, the utility function, and the ability distribution.

See the handout for Mirrlees results.

Many different shapes emerge and it seems hard to get general results.

Answers are particularly sensitive to the specification of the distribution of abilities, but this is not observable.

It is striking however that marginal tax rates are not increasing.

Diamond (1988) AER finds conditions under which marginal rates are increasing for sufficiently high incomes.

He assumes that

$$\varphi(l) = Al^{1+1/\varepsilon}$$

This implies that

$$\begin{aligned}\varepsilon(a) &= \frac{a^2(1-T'(y(a)))}{y(a)\varphi''(\frac{y(a)}{a})} \\ &= \frac{a\varphi'(\frac{y(a)}{a})}{y(a)\varphi''(\frac{y(a)}{a})} = \varepsilon\end{aligned}$$

He also assumes that above some ability level a_0 the distribution of productivities obeys a *Pareto distribution* with a density function

$$f(a) = \frac{B}{a^{1+\varsigma}}.$$

This implies that

$$\frac{1-F(a)}{af(a)}$$

is constant above a_0 (Diamond assumes no upper bound on the ability distribution).

Thus the shape of the marginal tax rates above $y(a_0)$ just depends on the third term $[1 - \frac{D(a)}{D(\underline{a})}]$ which is increasing.

Accordingly, marginal tax rates are increasing for incomes above $y(a_0)$.

But this conclusion does rely on the quasi-linear preferences (see the comment on Diamond by Dahan and Strawczynski 2000)

Saez (2001) *REStud* represents the state of the art in simulations.

He calibrates the exogenous ability distribution $F(a)$ which given the chosen utility function and the actual U.S. tax schedule $T(y)$ yields the actual empirically observed U.S. income distribution.

His analysis assumes the same utility function as Diamond and that $\Psi(U) = \log U$

His analysis suggests a U-shaped pattern of marginal rates to be optimal (see handout - note that $\zeta = \varepsilon$).

Optimal Mixed Taxation

Suppose that the government can employ both non-linear income taxation and linear commodity taxes.

What can be said about the optimal mixed tax system?

There is an important result that has been established about this case.

Lets consider a model which combines the optimal commodity tax and optimal income tax models.

There are I consumers, indexed by $i = 1, ..., I$.

There are n consumer goods, indexed by $j = 1, ..., n$.

Denote labor by l.

Consumers are all endowed with $\bar{l}$ units of time but differ in their productivities.

Consumer i produces $a_i l$ efficiency units of labor if he works an amount l.

Consumers have an identical utility function $u(x, l)$, where $x = (x_1, ..., x_n)$.

Each good j is produced from labor with a linear technology, $x_j = \frac{l_j}{\alpha_j}$.

Assume competitive production so that the producer price of good j is $p_j = \alpha_j w$, where w is the wage rate.

Without loss of generality, let $w = 1$.

To raise the revenue, the government imposes linear taxes on goods $j = 1, ..., n$. These are $t = (t_1, ..., t_n)$.

The government also employs a non-linear income tax schedule $T(y)$ where y denotes earnings.

Letting $R(y) = y - T(y)$ denote post-tax income, we can summarize a tax system by $(t, R(y))$.

Government revenue under the tax system $(t, R(y))$ is

$$G = \sum_i \left\{ t \cdot x^i + a_i l^i - R(a_i l^i) \right\}$$

Assumption: *The common utility function satisfies*

$$u(x, l) = U(v(x), l)$$

where the function $v(\cdot)$ is continuous and exhibits non-satiation.

This says that consumption goods and labor are *weakly separable*.

It means that the consumer's marginal disutility of labor is the same at any pair of consumption bundles that give rise to the same level of consumption utility v.

Proposition *Let* $(t^o, R^o(\cdot))$ *be any tax system such that for all* i *the utility attained by consumer* i

$$u_i = \max_{(x^i, l^i)} \{u(x^i, l^i) : (p + t^o) \cdot x^i \leq R^o(a_i l^i)\},$$

is well-defined. Then there exists another tax system $(0, R(\cdot))$ *which (i) provides all consumers the same utility; (ii) induces the same labor supply from all consumers; and (iii) provides the government with at least as much revenue.*

Thus, in this model, under the Assumption, commodity taxes are redundant when non-linear income taxes are available.

This result was first proved by Atkinson and Stiglitz *JPubE* (1976).

This particular formulation is taken from Laroque *EcLett* (2005) as is the following proof.

Proof: From the point of view of consumers, any tax system $(t, R(\cdot))$ is equivalent to a set $\mathcal{V} = (\varphi(y), y)$ where $\varphi(y)$ is the utility derived from consumption when pre tax income is y; that is

$$\varphi(y) = \max_x \{v(x) : (p + t) \cdot x \leq R(y)\},$$

and $y \geq 0$.

Indeed, consumer i chooses his labor supply by maximizing $U(\varphi(a_i l), l)$.

Let consumption utility under the tax system $(t^o, R^o(\cdot))$ be denoted by

$$\varphi^o(y) = \max_x \{v(x) : (p + t^o) \cdot x \leq R^o(y)\}.$$

Denote consumer i's labor supply under the tax system $(t^o, R^o(\cdot))$ by l_o^i.

Define

$$\overline{x}(y) = \arg\min\{\, p \cdot x \,|\, v(x) \geq \varphi^o(y)\},$$

and let

$$\overline{R}(y) = p \cdot \overline{x}(y).$$

Thus $\overline{x}(y)$ is the cheapest way of reaching consumption utility $\varphi^o(y)$ given the prices p and $\overline{R}(y)$ is the cost of that cheapest bundle.

Given the assumed properties of v it is the case that

$$\overline{x}(y) = \arg\max_x\{v(x) : p \cdot x \leq \overline{R}(y)\}.$$

Now consider the tax system $(0, \overline{R}(\cdot))$.

This keeps the set $\mathcal{V}$ exactly the same as under $(t^o, R^o(\cdot))$ since, for all y

$$\max_x\{v(x) : p \cdot x \leq \overline{R}(y)\} = \varphi^o(y).$$

Since consumer i has the same choices as before, he chooses the same labor supply l_o^i.

Consumer i chooses the consumption bundle $\overline{x}(a_i l_o^i)$ and obtains the same utility as under the original tax system $(t^o, T^o(\cdot))$.

It remains to show that the government obtains at least as much revenue.

Let x_o^i denote i's consumption under the original tax system $(t^o, T^o(\cdot))$.

Then, since $v(x_o^i) = \varphi^o(a_i l_o^i)$, we have that

$$p \cdot x_o^i \geq p \cdot \overline{x}(a_i l_o^i) = \overline{R}(a_i l_o^i)$$

This is because $\overline{x}(a_i l_o^i)$ is the cheapest way of reaching $\varphi^o(a_i l_o^i)$.

Thus,

$$\sum_i \left\{t^o \cdot x_o^i + a_i l_o^i - R^o(a_i l_o^i)\right\} = \sum_i \left\{a_i l_o^i - p \cdot x_o^i\right\}$$
$$\leq \sum_i \left\{a_i l_o^i - \overline{R}(a_i l_o^i)\right\}.$$

■

Observe that when $p \cdot x_o^i > p \cdot \overline{x}(a_i l_o^i)$ which will be the case with tax distortions, government revenue is strictly higher without the commodity taxes.

This permits a Pareto improvement when commodity taxes are eliminated.

The weak separability assumption is viewed by many as a good starting point in discussions of optimal tax policy and thus this result is seen as fundamental.

II. Public Goods

A *public good* is a good for which use of a unit by one consumer does not preclude its use by others.

This property is known as *non-rivalness in consumption*.

Classic examples are lighthouses, radio broadcasts, national defense, and air quality.

Public goods can be *excludable* or *non-excludable*.

Once a non-excludable public good has been provided to one consumer, it is impossible to prevent others from consuming it.

Lighthouses and national defense are non-excludable; Pay-per-view TV broadcasts are excludable.

Public goods are sometimes defined to be goods that are both non-rival in consumption and non-excludable.

Many goods lie between the extremes of a public good and a private good in that they can be shared, but eventually additional consumers impose negative externalities on others.

This type of good is referred to as a *public good with congestion* or an *impure public good.*

Public goods are interesting to public economists because they will be under-provided by the market.

The neo-classical theory of public goods developed by Samuelson (1954) explains why.

The Neo-classical Theory of Public Goods

A Model

Consider a community consisting of n consumers indexed by $i = 1, ..., n$

There are two goods - a numeraire private good z and a public good x

Each consumer i has quasi-linear utility

$$u_i = z_i + \varphi_i(x_i)$$

where $\varphi_i(\cdot)$ is increasing, strictly concave and satisfies $\varphi_i(0) = 0$.

z_i is consumer i's consumption of the private good and x_i his consumption of the public good.

Each consumer i has some endowment of the numeraire (or income) y_i.

The cost (in terms of units of the numeraire) of providing a unit of public good is c.

An *allocation* for this community consists of a description of what each consumer is consuming $(z_i, x_i)_{i=1}^n$

An allocation $(z_i, x_i)_{i=1}^n$ is *feasible* if (i) for all i, $x_i \in [0, x]$ and $z_i \geq 0$ and (ii)

$$\sum_i z_i \leq \sum_i y_i - cx.$$

The interpretation is that x is the aggregate amount of public good produced.

Constraint (i) says that no consumer can consume more public good than the total amount provided.

Note the difference in this feasibility constraint from the one for a private good.

If the public good is non-excludable then it must be the case that for all i, $x_i = x$.

Efficiency

An allocation $(z_i, x_i)_{i=1}^n$ is *Pareto efficient* if (i) it is feasible and (ii) there exists no alternative feasible allocation which Pareto dominates it.

Proposition 1 *An allocation* $(z_i^e, x_i^e)_{i=1}^n$ *such that* $z_i^e > 0$ *for all* i *is Pareto Efficient if and only if (i) for all* i, $x_i^e = x^e$; *(ii)*

$$\sum_i \varphi_i'(x^e) \leq c \quad (= \text{ if } x^e > 0),$$

and (iii) $\sum_i z_i^e = \sum_i y_i - cx^e$.

You will be asked to prove this proposition in the next problem set.

Condition (i) says that all consumers must get to consume all the public good provided.

This will necessarily be satisfied if the public good is non-excludable.

Condition (ii) says that the optimal level of the public good is such that the sum of marginal benefits must equal the marginal cost.

This is known as the *Samuelson Rule*.

Condition (iii) says that all the economy's resources should be used for consumption.

In this model the efficient level of the public good x^e is independent of the allocation of the numeraire among consumers.

This comes from the quasi-linear preferences and is not a general feature.

In general, it does not make sense to talk about *the* efficient level of the public good.

Market Provision of Public Goods

(i) Non-excludable public goods

Suppose the public good is provided via the market mechanism, with no collective action by the consumers.

Competition would ensure that the price of a unit of public good were c.

How much would each consumer choose to buy?

The amount demanded by a consumer would depend upon what he expected others to demand.

Thus, we have a strategic problem as opposed to a decision-theoretic problem.

Let ω_i denote the amount of public good purchased by consumer i.

A vector of public good purchases $(\omega_i^*)_{i=1}^n$ is an *equilibrium* if

$$\omega_i^* = \arg \max_{\omega_i \in [0, y_i/c]} y_i - c\omega_i + \varphi_i(\sum_{j \neq i} \omega_j^* + \omega_i)$$

An allocation $(z_i^*, x_i^*)_{i=1}^n$ is a *market equilibrium* if there exists an equilibrium vector of public good purchases $(\omega_i^*)_{i=1}^n$ such that for all i (i) $x_i^* = \sum_j \omega_j^*$ and (ii) $z_i^* = y_i - c\omega_i^*$.

Proposition 2: *Suppose that the efficient level of the public x^e is positive. Then if $(z_i^*, x_i^*)_{i=1}^n$ is a market equilibrium it is not efficient.*

Proof: A market equilibrium satisfies conditions (i) and (iii) of Proposition 1, but not condition (ii).

In a market equilibrium, for all consumers i

$$\varphi_i'(\sum_j \omega_j^*) \leq c \ \ (= \text{ if } \omega_i^* > 0).$$

This follows from the first order condition for the consumer's problem.

It follows that if $x^e > 0$, the market must under-provide the public good.

If $\omega_i^* = 0$ for all i this is immediate.

If $\omega_i^* > 0$ for some i, then we must have that

$$\sum_i \varphi_i'(\sum_j \omega_j^*) > c$$

which implies that $\sum_j \omega_j^* < x^e$. ■

The problem with the market is *free-riding* - everybody free rides on everybody else's provision.

(ii) Excludable Public Goods

When public goods are excludable, market provision is more promising, because free riders can be excluded.

However, it is not obvious how to think about market provision since a public goods producer can sell the same unit of the good to multiple consumers.

An idealized notion of how a market for excludable public goods might work is the *Lindahl equilibrium*

An allocation $(z_i^*, x_i^*)_{i=1}^n$ is a *Lindahl equilibrium* if there exists an equilibrium vector of personalized public good prices $(p_i^*)_{i=1}^n$ such that (i) for all i (z_i^*, x_i^*) solves the problem

$$\begin{array}{cc} \max & z_i + \varphi_i(x_i) \\ s.t. & z_i + p_i^* x_i = y_i \end{array} ;$$

(ii) for all i and j, $x_i^* = x_j^*$; and (iii) $\sum_i p_i^* = c$.

The logic of the Lindahl equilibrium is that each consumer faces a personalized price which induces him to choose the same level of public good.

Firms producing the public good receive the sum of these personalized prices for each unit they produce - hence condition (iii) of the definition.

Proposition 3: *If* $(z_i^*, x_i^*)_{i=1}^n$ *is a Lindahl equilibrium, it is Pareto efficient.*

Proof: For all consumers i

$$\varphi_i'(x_i^*) = p_i^*,$$

and

$$x_i^* = x^*.$$

Moreover,

$$\sum_i p_i^* = c.$$

Accordingly,

$$\sum_i \varphi_i'(x^*) = c.$$

Thus all three conditions from Proposition 1 are satisfied. ■

The Lindahl equilibrium is unrealistic in the sense that everybody faces a personalized price, which makes them demand the same level of the public good.

If there is a common price, inefficiency will result.

To illustrate, change the model to make the public good discrete; that is, assume $x \in \{0, 1\}$

Assume that $\sum_i \varphi_i(1) \geq c$, so that $x^e = 1$.

Recall efficiency requires not only that the good be provided but also that all consumers consume it.

The Lindahl personalized public good prices $(p_i^*)_{i=1}^n$ would be such that for all i, $p_i^* \in [0, \varphi_i(1)]$ and $\sum_i p_i^* = c$

But suppose each consumer's price must be the same.

For all p, let

$$D(p) = \#\{i : \varphi_i(1) \geq p\}$$

Suppose that for all p

$$pD(p) < c.$$

Then the market would not provide the good.

If $pD(p) \geq c$ for some p, then the market would provide the good at the price such that $pD(p) = c$ (assuming such a price exists!)

But individuals for whom $\varphi_i(1) < p$ would be inefficiently excluded.

Conclusion

The bottom line is that market mechanisms are unlikely to provide public goods efficiently.

This suggests a case for government provision.

In principle, the government can provide the efficient level, financing it with taxation, and then allow all consumers to consume it.

Post Neo-classical Public Goods Research

1. Mechanism Design

A large body of work has studied the design of "mechanisms" that the government can use to determine public good provision.

The motivation for this literature is as follows:

(i) market mechanisms will fail to provide public goods efficiently, suggesting the government should do it;

(ii) to provide public goods efficiently, the government needs to know individuals' willingness to pay for public goods;

(iii) individuals need to be given appropriate incentives to reveal their true willingness to pay.

To illustrate, consider the following parameterized version of our public goods model.

Each consumer i has quasi-linear utility given by

$$u_i = z_i + \theta_i \ln x_i$$

where $\theta_i > 0$.

By the *Samuelson Rule*, the efficient level of the public good is

$$x^* = \frac{\sum_i \theta_i}{c}$$

and hence the government needs to know the θ_i in order to figure out the efficient level.

Suppose that a naive government were to ask each individual what his θ_i were and then provide the Samuelson level, financing it by a uniform "head tax".

Would an individual have an incentive to truthfully reveal his willingness to pay?

It is easy to see that truth telling is not a dominant strategy.

Assuming that all other individuals are telling the truth, citizen i's problem would be to choose a report r_i to maximize

$$y_i - \frac{c}{n}\left(\frac{\sum_{j\neq i}\theta_j + r_i}{c}\right) + \theta_i \ln\left(\frac{\sum_{j\neq i}\theta_j + r_i}{c}\right).$$

The solution to this problem is

$$r_i = n\theta_i - \sum_{j\neq i}\theta_j \neq \theta_i$$

The question then is: is it possible to find rules for the provision and financing of the public good that would both lead individuals to report their true valuations and enable the government to achieve efficiency?

The rules for provision and financing represent the *mechanism* and can be represented as functions $x(r_1, ..., r_n)$ and $\{t_i(r_1, ..., r_n)\}_{i=1}^{n}$.

When considering the problem of mechanism design, it is necessary to specify how individuals will behave in submitting their reports - this is because submitting reports is a strategic choice.

There are two main approaches: require that truthful reporting be a *Dominant Strategy equilibrium* or a *Bayesian Nash equilibrium*.

The three main results in the literature are as follows:

Result 1 *There exists a mechanism that with quasi-linear preferences enables the government to provide the efficient level of the public good when true reporting is a dominant strategy equilibrium.*

This is known as the *Pivot Mechanism* or the *Vickrey-Clarke-Groves Mechanism*.

Result 2 *There is no mechanism that enables the government to provide both the efficient level of the public good and balance its budget when true reporting is a dominant strategy equilibrium.*

This means that it is impossible to achieve full efficiency; i.e., both conditions (ii) and (iii) of Proposition 1.

Result 3 *There exists a mechanism that with quasi-linear preferences enables the government to provide the efficient level of the public good and balance its budget when true reporting forms a Bayesian Nash equilibrium.*

There is still a large amount of on-going research on mechanism design, but it has yet to yield practically useful results in the public goods context (in contrast to the theory of auctions).

More information on this literature can be found in Mas-Colell, Green, and Whinston.

2. The Political Economy of Public Good Provision

A large amount of work has focused on the political determination of public good provision.

The motivation is that public goods are typically provided by government and we want to understand the forces that shape their provision.

A major focus of this research is to understand how political decision-making distorts provision away from the normative ideal.

There are many different approaches and we will cover only the standard one.

This assumes that there is some predetermined method of financing (e.g., head tax, income tax) and that given this the political process will select the *majority preferred* level of the public good.

The justification for this assumption is not completely clear, but the basic idea is that only a majority preferred level would be stable in the sense of not being vulnerable to being voted out by a majority of citizens.

To illustrate the approach, suppose that the public good is to be financed by a head tax.

Then using the parameterized utility function from before, consumer i's preferences over public good levels are

$$v_i(x) = \theta_i \ln x - \frac{c}{n}x$$

The ideal level of the public good from the viewpoint of consumer i is $x_i^* = n\theta_i / c$.

A level of public good x^* is *majority preferred* if for any $x \neq x^*$ a majority of consumers prefer x^* to x.

There is a convenient chracterization of the majority preferred public good level - if consumers' preferences over policy are *single-peaked*, it is the level preferred by the *median voter*.

Consumer m is a median voter if less than half the population prefer a lower level of the public good and less than half prefer a higher level.

In political economy models of public good provision, it is therefore often simply assumed directly that the level preferred by the median voter will be selected.

Note that this level will equal the Samuelson level if and only if

$$\theta_m = \frac{\sum_i \theta_i}{n}.$$

The condition is that the median voter is also the *mean voter*.

There is no reason to suppose that this will be the case.

Large differences between the median voter's preferred level and the Samuelson level arise when there are significant differences in the *intensity* of preferences between those on either side of the median.

3. Optimal Provision of Public Goods when Taxation is Distortionary

The *Samuelson Rule* assumes that the government can finance public good provision with non-distortionary "lump-sum" taxation.

In reality, this is unlikely to be the case for reasons that we have already discussed.

When public goods must be financed by distortionary taxation, the Samuelson Rule needs to be modified.

Much attention has been devoted to understanding how.

Results depend on what tax instruments are available and the underlying economic environment.

The simplest case is that in which individuals are identical and the government finances public good provision with a linear income tax.

To analyze this case, extend our public good provision model to have three goods - a public good x; consumption z; and labor l.

Each citizen's utility is

$$u_i = z_i + \varphi(x) - \frac{l^{(1+\frac{1}{\varepsilon})}}{1+\varepsilon},$$

where $\varepsilon > 0$.

Individuals have no endowments of income but all can work at the exogenous wage rate w.

Given the wage rate w each citizen will work an amount $l^*(w) = (\varepsilon w)^\varepsilon$, so that ε is the elasticity of labor supply.

The associated indirect utility function is given by

$$v(w, x) = \frac{\varepsilon^\varepsilon w^{\varepsilon+1}}{\varepsilon+1} + \varphi(x).$$

The government raises funds with a proportional tax on labor income at rate t.

The revenues raised with tax rate t are given by

$$R(t) = ntwl^*(w(1-t)) = nt(1-t)^\varepsilon \varepsilon^\varepsilon w^{\varepsilon+1}.$$

Such a revenue function is known as a *Laffer Curve.*

The function is hump-shaped and maximized at the tax rate $t = 1/(1+\varepsilon)$.

If the tax rate exceeds $1/(1+\varepsilon)$, the economy is said to be "on the wrong side of the Laffer Curve".

The levels of the public good x that are feasible are:

$$x \in [0, \frac{R(\frac{1}{1+\varepsilon})}{c}].$$

For any x in this range, there exists a unique tax rate $t(x)$ in the range $[0, 1/(1+\varepsilon)]$ which raises enough revenue to finance it; i.e., which satisfies $R(t) = cx$.

The tax rate function $t(x)$ is increasing in x, with derivative

$$\frac{dt(x)}{dx} = \frac{c}{R'(t)} = \frac{c}{(1-t)^{\varepsilon-1}[1-t(1+\varepsilon)]n\varepsilon^{\varepsilon}w^{\varepsilon+1}}$$

Given that individuals are identical, the Pareto efficient public good level solves the problem

$$\max_{x\in[0,R(\frac{1}{1+\varepsilon})/c]} nv(w(1-t(x)),x).$$

The first order condition is that

$$\begin{aligned} n\varphi'(x^*) &= n\varepsilon^{\varepsilon}[w(1-t(x^*))]^{\varepsilon}w\frac{dt(x^*)}{dx} \\ &= (\frac{1-t(x^*)}{1-t(x^*)(1+\varepsilon)})\cdot c \end{aligned}$$

The expression in brackets is known as the *marginal cost of public funds.*

It represents the social cost of raising an additional \$1 of tax revenue and exceeds 1.

The optimal level of public goods is below that prescribed by the Samuelson Rule.

The intuitive reason is that in addition to the direct costs of providing the public good (i.e., c) there is a distortionary cost created by the financing.

That such considerations would reduce the optimal level of public goods was first noted by Pigou in the 1940s.

Nonetheless, this under-provision result can be reversed in more complicated models.

In a general optimal commodity tax model with public goods, it could be that the public good is complementary with highly taxed consumer goods.

In this case government revenues will be increased when public good provision is expanded (see Atkinson and Stern (1974)).

Alternatively, in a world with heterogeneous consumers, the public good could be valued more by consumers with higher social weight (i.e., the poor).

In models with non-linear income taxes, the Samuelson Rule remains valid if individuals have identical utility functions and if consumer goods and the public good are weakly separable from labor.

Thus, in a model with three goods - a public good x; consumption z; and labor l - if we can write utility as

$$u(z, x, l) = U(v(z, x), l)$$

for some real valued function $v(z, x)$, increasing in both arguments, the Samuelson Rule applies.

In the next problem set, I will ask you to verify this in a two type Mirrlees model extended to include public goods.

To extend the model, we will assume that consumers have utility functions

$$z + B(x) - \varphi(l)$$

and that the cost of the public good is c per unit.

Boadway and Keen (1993) provide a general treatment of the two type Mirrlees model with public goods.

4. Private Provision of Public Goods

4.1 Voluntary Contributions

A large body of work studies private voluntary contributions to public goods, with leading applications being to charitable giving and contributions to public radio.

The starting point is the basic model of private purchases of a non-excludable public good that we studied last time; although with utility functions that are general

$$u_i = u_i(z_i, x)$$

The assumptions imposed on the utility function are the usual ones plus the requirement that both goods are normal.

The model of voluntary contributions is as follows: let ω_i denote the amount contributed by citizen i.

Then a vector of contributions $(\omega_i^*)_{i=1}^n$ is an *equilibrium* if

$$\omega_i^* = \arg \max_{\omega_i \in [0, y_i]} u_i(y_i - \omega_i, \sum_{j \neq i} \omega_j^* + \omega_i)$$

The public good is measured in dollar terms, so effectively $c = 1$.

This model turns out to have some very strong implications - see Andreoni's survey article "Philanthropy" for a good discussion.

The first implication concerns *crowd out*.

Let $(\omega_i^*)_{i=1}^n$ be an equilibrium without government and suppose the government were to now provide some public good, so that the total amount provided was $\sum_j \omega_j + g$ where g is the publicly provided level.

Suppose this was financed by taxing each individual cg/n.

Then, if $cg/n \leq \omega_i^*$ for each citizen i, the total amount of the public good provided would remain unchanged.

To see why, note that if

$$\omega_i^* = \arg \max_{\omega_i \in [0, y_i]} u_i(y_i - \omega_i, \sum_{j \neq i} \omega_j^* + \omega_i)$$

then, provided that $\omega_i^* \geq cg/n$, we have that

$$\omega_i^* - \frac{cg}{n} =$$
$$\arg \max_{\omega_i \in [0, y_i]} u_i(y_i - \frac{cg}{n} - \omega_i, \sum_{j \neq i} (\omega_j^* - \frac{cg}{n}) + \omega_i + cg).$$

Thus there is 100% crowd out of private provision!

This notion of crowd out is much beloved by conservative economists.

In reality, of course, many people do not contribute anything so that $cg/n > \omega_i^*$ for many citizens.

In this case, public provision will raise total provision.

However, it can be shown (see Andreoni's survey paper), that if the number of private contributors remains (numerically) large, the increase in total provision must be small, so there is still almost 100% crowd out.

The second implication concerns what happens when we increase the number of citizens; i.e., let n become large.

Then the proportion of citizens contributing shrinks to zero and the average per person contribution shrinks to zero.

Both implications are inconsistent with what we know about charitable giving and contributions to public radio.

Large proportions of citizens contribute to charities like the United Way or Red Cross.

Kingma *JPE* (1989) studies how public radio contributions respond to different government contributions.

He found that for every $1 increase in government funding, private contributions fell by 13.5cents - very small crowd out.

Brunner *Public Choice* (1998) studies how contributions to public radio depend on the size of the listening audience.

He finds that the proportion of contributors decreases in the size of the listening audience, but that there is no effect on the average contribution per contributor.

This model of voluntary contributions has also been subject to intense scrutiny in *laboratory experiments*.

A simple example of the type of experiment used is as follows:

5 undergrads are placed in a room to play 10 rounds of a simple game.

In each round, the students are given \$1.

They can keep it or place it in a public fund.

All dollars placed in the public fund are doubled and divided equally among the players.

The voluntary contribution model predicts that no player would ever contribute to the public fund, but in the lab 30-70% of players contribute.

Moreover, even though contribution rates tend to decline over the rounds of play, they rarely go to zero.

Alternative Models

The poor performance of the standard voluntary contribution model has lead researchers to develop alternatives.

Andreoni *JPE* (1989) proposes the *Warm Glow Model*, whereby citizens' utility is allowed to depend directly on the contributions they make.

Thus,

$$u_i = u_i(z_i, x, \omega_i)$$

This model delivers sensible predictions, but begs the question as to why individuals care about their specific contributions.

One option which has been suggested here is that individuals contribute to signal something about themselves - e.g., their wealth.

However, many contributions are unobserved.

Another option is that individuals experience pride when they give or feel guilt if they do not.

Without knowing why people care about their own contributions, it is difficult to know how to treat the utility from giving in policy analyses.

For example, consider evaluating a policy that replaces private contributions to a public good with tax-financed contributions without changing any individual's contribution or the total amount provided.

Are people worse off or better off? This is unclear.

Further Developments

Interest in voluntary contributions to public goods has led into more detailed empirical study of individuals' charitable giving and also the fund-raising behavior of non-profits.

This interesting research is discussed in Andreoni's survey paper.

Also, some very interesting field experiments are being performed with charitable giving - for example, Karlan and List *AER* (2007).

4.2 Provision via Advertising

A second (much smaller) strand of literature on the private provision of public goods looks at provision via advertising.

Think of broadcasting, which is a leading example of a public good.

Historically and still today in many countries it is a non-excludable public good, in the sense that if you have the hardware (radio or tv) you can pick up the signals for free.

Nonetheless, in most developed countries, the bulk of broadcasting is supplied by the market.

Market provision is possible because broadcasters put advertisements on their broadcasts and sell advertising slots to producers wishing to advertise their products.

The internet, newspapers and yellow page directories are the same.

Such advertising means that broadcasting performs a dual role in the economy - providing programming to consumers and allowing producers to contact viewers.

It is natural to wonder about the efficiency of market provision of public goods like broadcasting and there are lots of interesting regulatory issues.

For example, excessive advertising; amount, quality, and variety of programming; ownership structure; desirability of pricing.

Anderson and Coate (2005) initiate an analysis of these issues.

A-C model a broadcasting system in which programs are broadcast over the air and viewers/listeners can costlessly access programming.

We consider only the one channel version - see the paper for the multi-channel version.

Anderson-Coate Model

There is one channel which can broadcast one program.

The program can carry advertisements.

Each advertisement takes a fixed amount of time and thus ads reduce the substantive content of a program.

The cost of producing the program is independent of the number of ads and equals K.

There are N potential viewers, each characterized by a taste parameter $\lambda \in [0, 1]$.

A type λ viewer obtains a *net viewing benefit* $\beta - \gamma a - \tau\lambda$ from watching the program with a advertisements where $\tau > \beta > 0$ and $\gamma > 0$.

Not watching the program yields a zero benefit.

γ measures the *nuisance cost* of ads and τ measures the extent to which higher types dislike the program.

Viewers' tastes are uniformly distributed, so that the fraction of viewers with taste parameter less than λ is just λ.

Ads are placed by producers of new goods and inform viewers of the nature and prices of these goods.

There are m producers of new goods, each of which produces at most one good.

New goods are produced at a constant cost per unit, wlog set equal to zero.

Each new good is characterized by some type $\sigma \in [0, \overline{\sigma}]$ where $\overline{\sigma} \leq 1$.

A viewer has willingness to pay $\omega > 0$ with probability σ for a new good of type σ and willingness to pay 0 with probability $1 - \sigma$.

The fraction of producers with new goods of type less than σ is $F(\sigma)$.

Since a consumer will pay ω or 0, each new producer will advertise a price of ω.

Thus, a new producer with a good of type σ is willing to pay $\sigma\omega$ to contact a viewer.

Accordingly, if an advertisement reaches V viewers and costs P, the number of firms wishing to advertise is $a_d(P,V) = m \cdot [1 - F(P/V\omega)]$.

Let $P(a,V)$ denote the corresponding inverse demand curve.

Since viewers get no surplus from new goods, they get no *informational benefit* from watching a program with advertisements.

Viewers therefore choose to watch the program only if their net viewing benefit is positive.

Optimal provision

Think of the program as a discrete public good which costs K to provide and may be consumed by two types of agents - viewers and advertisers.

An advertiser "consuming" a program, places its advertisement on that program.

The optimality problem is to decide if the public good should be provided and who should consume it.

Following the *Samuelson Rule*, provision will be desirable if the sum of benefits exceeds cost.

Suppose that the program has a advertisements and hence is "consumed" by a new producers.

Then, viewers for whom $\lambda \leq \frac{\beta - \gamma a}{\tau}$ will watch and obtain a benefit $\beta - \gamma a - \tau\lambda$.

If the a advertisements are allocated to the new producers who value them the most, the benefits generated by the program are

$$B(a) = N\int_0^{\frac{\beta-\gamma a}{\tau}} (\beta - \gamma a - \tau\lambda)d\lambda + \int_0^a P(\alpha, N(\frac{\beta-\gamma a}{\tau}))d\alpha.$$

The optimal level of advertising equates marginal social benefit and cost.

The marginal social benefit is the increase in advertiser benefits created by an additional advertisement which is

$$P(a, N(\frac{\beta-\gamma a}{\tau})) - N\frac{\gamma}{\tau}\int_0^a \frac{\partial P}{\partial V}d\alpha.$$

The marginal social cost is the reduction in viewer benefits created by an additional advertisement, which is

$$N(\frac{\beta-\gamma a}{\tau})\gamma.$$

The situation can be illustrated graphically.

Provision is desirable if the operating cost K is less than the maximized benefits $B(a^o)$.

Market provision

Imagine the channel is controlled by a profit-maximizing broadcaster.

The broadcaster chooses whether to broadcast the program and, if so, the level of advertising.

If it runs a advertisements, its program will be watched by viewers for whom $\lambda \leq \frac{\beta - \gamma a}{\tau}$.

To sell a advertisements it must set a price $P(a, N(\frac{\beta - \gamma a}{\tau}))$ so its revenues will be

$$\pi(a) = P(a, N(\frac{\beta - \gamma a}{\tau}))a.$$

The revenue maximizing advertising level, a^*, is the level at which marginal revenue is zero; i.e.,

$$P(a^*, N(\tfrac{\beta - \gamma a^*}{\tau})) + \frac{\partial P(a^*, N(\frac{\beta - \gamma a^*}{\tau}))}{\partial a} a^*$$
$$- N \frac{\gamma}{\tau} \frac{\partial P(a^*, N(\frac{\beta - \gamma a^*}{\tau}))}{\partial V} a^* = 0.$$

The broadcaster will provide the program if $\pi(a^*)$ exceeds K.

Optimal vs Market Provision

Conditional on the broadcaster providing the program, will it have too few or too many advertisements?

Proposition 1 $a^* < a^o$ *for* γ *low enough and* $a^* > a^o$ *for* γ *high enough.*

This proposition can be established diagrammatically.

The most striking finding is the possibility that the program has too few ads.

The reason is that the broadcaster has a monopoly in delivering its audience to advertisers.

The broadcaster holds down ads in order to keep up the prices it receives.

This logic applies even with competitive broadcasters, if each broadcaster has a monopoly in delivering its audience to advertisers.

Turning to programming, the question is whether the monopoly will make an efficient decision with respect to the broadcast of the program.

Proposition 2 *There exists a range of operating costs for which the broadcast should optimally be provided but the broadcaster will not provide it.*

To establish this, we need only show that

$$B(a^o) > \pi(a^*).$$

But this follows from the fact that

$$\begin{gathered}\pi(a^*) = P(a^*, N(\tfrac{\beta-\gamma a^*}{\tau}))a^* \\ < \int_0^{a^*} P(\alpha, N(\tfrac{\beta-\gamma a^*}{\tau}))d\alpha + N\int_0^{\frac{\beta-\gamma a^*}{\tau}} (\beta - \gamma a^* - \tau\lambda)d\lambda \\ \leq B(a^o)\end{gathered}$$

Intuitively, the broadcaster's profits are bounded by the advertiser surplus, but total surplus includes viewer and advertiser surplus.

The market may produce something close to the optimum for a range of parameter values; that is, the program could be provided and the advertising level could be approximately optimal.

Thus the market does not *necessarily* provide broadcasting inefficiently.

5. Durable Public Goods

In practice, many important public goods are durable, lasting for many years and depreciating relatively slowly.

Important examples are public infrastructure, defense capability, and environmental quality.

For a durable public good, the level in period t is given by

$$g_t = (1 - \delta)g_{t-1} + I_t,$$

where δ is the depreciation rate and I_t is investment in period t.

A distinction is made between *irreversible* and *reversible* public goods: for the former, investment must be non-negative and for the latter disinvestment is possible (i.e., I_t can be negative).

Recent years have seen increasing interest in this type of public good.

While it is relatively straightforward to derive the rules for the optimal provision of such goods (next problem set?), understanding private and political provision is much more challenging.

This is because the durable nature of these goods creates a dynamic linkage across decision-making periods.

Battaglini, Nunnari, and Palfrey (2012) analyze the private provision of durable public goods in a model which is the natural dynamic extension of the standard static model.

There are n infinitely-lived consumers indexed by $i = 1, ..., n$

There are two goods - a numeraire private good z and a public good x

Each consumer i has per-period utility given by the quasi-linear function

$$u_i = z_i + \varphi(x)$$

where $\varphi(\cdot)$ is increasing, strictly concave, and satisfies $\varphi(0) = 0$ and the Inada conditions.

In each period, each consumer obtains an income y.

The cost (in terms of units of the numeraire) of providing a unit of public good is 1.

Consumers have discount rates $\beta \in (0, 1)$.

In each period, given the current level of the public good x, each consumer chooses how much investment to privately provide ω_i.

Next period's public good level is then given by

$$x' = (1 - \delta) \left(x + \sum_i \omega_i \right).$$

The authors focus on symmetric Markov equilibria which are described by a common private investment strategy $\omega(x)$ and a value function $V(x)$.

In the irreversible case, $\omega(x)$ and $V(x)$ must satisfy the conditions that

$$\omega(x) =$$

$$\arg \max_{\omega \in [0,y]} \left\{ \begin{array}{c} y - \omega + \varphi(x + (n-1)\omega(x) + \omega) \\ +\beta V((1-\delta)[x + (n-1)\omega(x) + \omega]) \end{array} \right\},$$

and

$$V(x) = y - \omega(x) + \varphi(x + n\omega(x)) + \beta V((1-\delta)\left[x + n\omega(x)\right]).$$

Battaglini et al show that the consumers' contributions are gradual and aggregate investment is inefficiently slow.

However, with sufficiently patient consumers ($\beta \to 1$) and a sufficiently durable public good ($\delta \to 0$) the equilibrium level of the public good converges to the efficient level.

III. Externalities

An *externality* arises whenever the utility or production possibility of an agent depends *directly* on the actions of another agent.

The caveat of "directly" is important because everybody's actions impact everybody else through the price mechanism.

Price related effects are known as *pecuniary externalities* and these have no implications for government intervention.

Externalities can go from producers to consumers, consumers to consumers, and producers to producers, and can be negative or positive.

Externalities are interesting to public finance economists because they create market failure.

Government intervention to regulate externalities is very widespread, most notably in the area of the environment.

1. Simple Models of Externalities

Many different models of externalites can be constructed - producer-producer, producer-consumer, etc.

Consumer-consumer Externality

Consider a community consisting of 2 consumers indexed by $i = 1, 2$

Suppose that there are two goods - a numeraire good z and another good x (think cigarettes)

Consumer 1 has utility function

$$u_1 = z_1 + \varphi(x_1)$$

where $\varphi(\cdot)$ is increasing, strictly concave and satisfies $\varphi(0) = 0$.

Here z_1 is 1's consumption of numeraire and x_1 his consumption of the other good.

Consumer 2 has utility function

$$u_2 = z_2 - \xi(x_1)$$

where $\xi(\cdot)$ is increasing, strictly convex and satisfies $\xi(0) = 0$.

Thus, consumer 2 gets no utility from his own consumption of good x, and negative utility from 1's consumption (for example, 2 is a non-smoker and 1 is a smoker).

Each consumer i has some endowment of the numeraire (or income) y_i.

The cost (in terms of units of the numeraire) of providing a unit of good x is c.

An *allocation* for this community consists of a description of what each consumer is consuming $(z_i, x_i)_{i=1}^2$

An allocation $(z_i, x_i)_{i=1}^2$ is *feasible* if (i) for all i $x_i \geq 0$ and $z_i \geq 0$ and (ii)

$$\sum_i z_i + c \sum_i x_i \leq \sum_i y_i.$$

Efficiency

An allocation $(z_i, x_i)_{i=1}^2$ is *Pareto efficient* if (i) it is feasible and (ii) there exists no alternative feasible allocation which Pareto dominates it.

Proposition 1: *An allocation $(z_i^e, x_i^e)_{i=1}^2$ such that $z_i^e > 0$ for all i is Pareto Efficient if and only if (i)*

$$\varphi'(x_1^e) - \xi'(x_1^e) \leq c \;\; (= \text{ if } x_1^e > 0),$$

(ii) $x_2^e = 0$; and (iii) $\sum_i z_i^e + c \sum_i x_i^e = \sum_i y_i$.

Condition (i) says that the level of 1's consumption of good x is such that social marginal benefit equals social marginal cost.

Social marginal benefit includes both 1's consumption benefit and 2's negative disutility from 1's activities.

Market Failure

If good x is provided via the market, competition would ensure that its price was c.

An allocation $(z_i^*, x_i^*)_{i=1}^2$ is a *market equilibrium* if

$$(z_1^*, x_1^*) = \arg\max\{z_1 + \varphi(x_1) : y_1 \geq cx_1 + z_1\}$$

and

$$(z_2^*, x_2^*) = \arg\max\{z_2 - \xi(x_1^*) : y_2 \geq cx_2 + z_2\}.$$

Proposition 2: *Suppose that $\varphi'(0) > c$. Then if $(z_i^*, x_i^*)_{i=1}^2$ is a market equilibrium it is not efficient.*

While conditions (ii) and (iii) are satisfied, (i) is not.

In the market equilibrium, we have that

$$\varphi'(x_1^*) = c$$

which implies that $x_1^* > x_1^e$.

Government Intervention

The government can restore efficiency by either imposing a quota or a tax.

The quota could be imposed either on Mr 1's consumption of good x or on firms' production of good x.

The tax could either be imposed on Mr 1's consumption or on firms' production. The optimal tax is $\xi'(x_1^e)$.

This type of tax is known as a *Pigouvian corrective tax*.

Any distributional consequences of the tax could be dealt with by lump sum redistribution.

Producer-consumer Externality

There are n identical consumers.

There are two goods - a numeraire good z and another good x.

Each consumer has utility function

$$u = z + \varphi(x) - D(E)$$

where $\varphi(\cdot)$ is increasing, strictly concave, and satisfies $\varphi(0) = 0$, and $D(\cdot)$ is increasing, strictly convex, and satisfies $D(0) = 0$.

The variable E is the level of environmental pollution and D is the "damage function".

Consumers have an endowment of the numeraire good y.

Good x is produced from good z with the linear technology $x = z/c$ by competitive firms

Each unit of x produced emits γ units of pollution.

Analysis

To simplify, we focus on efficient allocations in which all consumers receive the same allocation.

Letting Z and X denote aggregate amounts of good z and good x, the efficiency problem is

$$\max \tfrac{Z}{n} + \varphi(\tfrac{X}{n}) - D(\gamma X)$$
$$s.t.\ Z + cX = ny$$

The efficient level of X satisfies the first order condition

$$\varphi'(\frac{X^e}{n}) = c + nD'(\gamma X^e)\gamma$$

The equilibrium level of X satisfies the first order condition

$$\varphi'(\frac{X^*}{n}) = c$$

The equilibrium level is too high and the government can restore efficiency by either imposing a quota or a tax.

The optimal tax is $nD'(\gamma X^e)\gamma$.

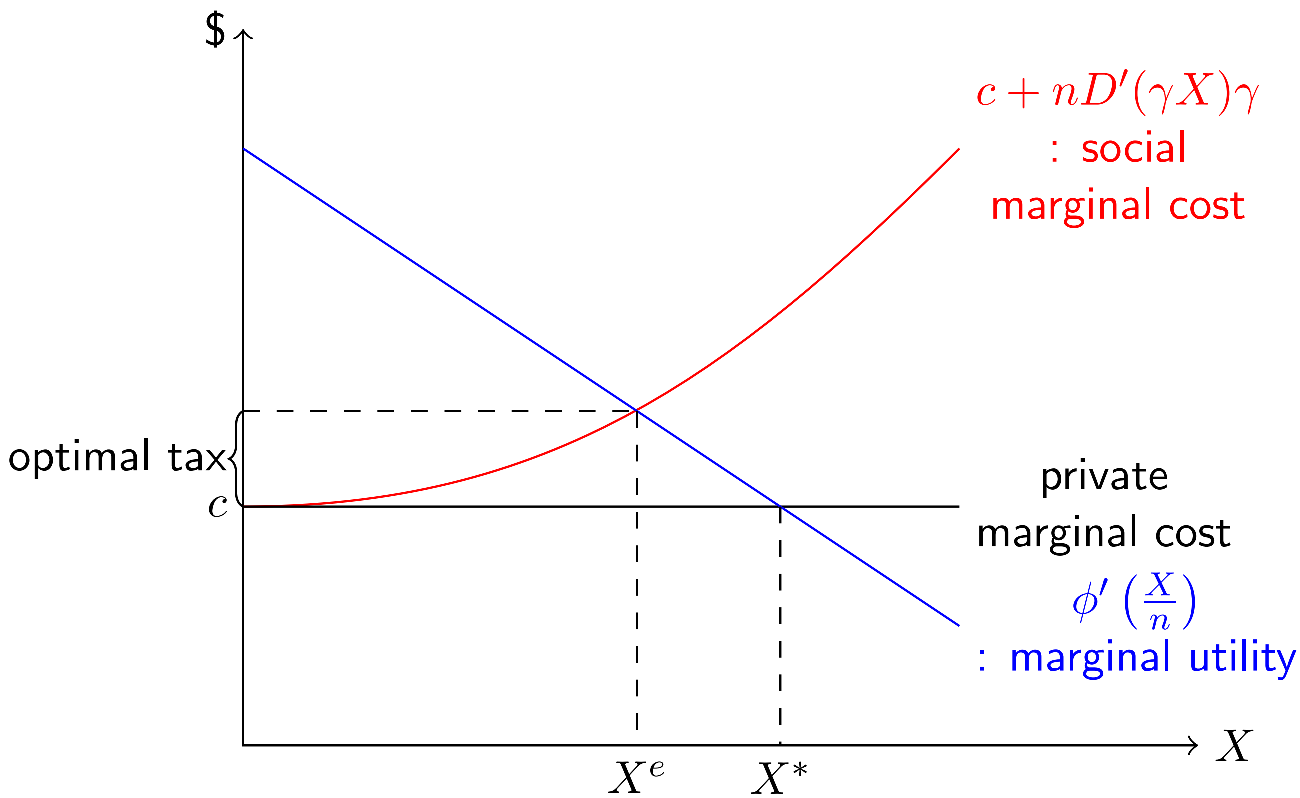

Figure 1: Producer-Consumer Externality

2. Coase Theorem

Coase (1960) *JLE* pointed out that externality problems arise from two factors: (i) lack of clearly specified property rights and (ii) transactions costs.

To illustrate, consider the consumer-consumer externality model and suppose we assign Mr 1 the right to consume as much x as he likes and that there are no transactions costs.

Then Mr 2, recognizing Mr 1's right would offer him a transfer in exchange for him cutting back his x consumption.

Mr 1 would agree to the deal provided it made him better off.

Assuming a one round bargaining procedure with Mr 2 making a take it or leave it offer, Mr 2 would make an offer (T, x_1) to solve

$$\max y_2 - T - \xi(x_1)$$
$$y_1 + T - cx_1 + \varphi(x_1) \geq y_1 - cx_1^* + \varphi(x_1^*)$$

Clearly,

$$T = cx_1 - \varphi(x_1) - cx_1^* + \varphi(x_1^*)$$

and thus, the level of x_1 that 1 would choose solves

$$\max y_2 - cx_1 + \varphi(x_1) - \xi(x_1) - cx_1^* + \varphi(x_1^*)$$

which means that

$$\varphi'(x_1) - \xi'(x_1) \leq c \quad (= \text{ if } x_1 > 0).$$

implying that $x_1 = x_1^e$!

Alternatively, we could assign Mr 2 the right not to be infringed by Mr 1's consumption.

Then Mr 1, recognizing Mr 2's right would offer him a transfer in exchange for allowing him to consume x.

Mr 2 would agree to the deal provided it made him better off.

Assuming a one round bargaining procedure with Mr 1 making a take it or leave it offer, Mr 1 would make an offer (T, x_1) to solve

$$\max y_1 - T - cx_1 + \varphi(x_1)$$
$$y_2 + T - \xi(x_1) \geq y_2$$

The same conclusion emerges - the negotiated level of x_1 would be efficient. This yields:

Coase Theorem: *If property rights are clearly specified and there are no transactions costs, bargaining will lead to an efficient outcome no matter how the rights are allocated.*

The allocation of rights have distributional consequences, but no efficiency consequences.

The allocation of rights is crucial because otherwise the participants would be fighting over who should be paying who.

What are the transactions costs that Coase had in mind?

Most obviously, if there are a large number of citizens impacted by the externality, there will be significant transactions costs to bring everybody to the table.

Less obviously, if participants have to pay an ex ante cost in order for an agreement to be reached (such as showing up to a meeting) this can lead to inefficiencies arising for strategic reasons - see Anderlini and Felli "Costly Bargaining and Renegotiation" *Econometrica*, 2001.

In this case an agreement need not be reached even if the surplus created from such an agreement would exceed the negotiation costs.

Imperfect information is also a problem.

Affected parties are unlikely to have perfect information about each others' benefits and costs.

With bilateral asymmetric information, bargaining will not lead to an efficient outcome (Myerson and Satterthwaite, *Journal of Economic Theory* (1983)).

To illustrate these difficulties, return to our example, but assume that x_1 is a discrete choice; that is, $x_1 \in \{0, \overline{x}_1\}$.

Further assume that

$$\varphi(x_1) = \varphi(x_1; \theta)$$

and

$$\xi(x_1) = \xi(x_1; \eta)$$

where θ and η are random variables with ranges $[\underline{\theta}, \overline{\theta}]$ and $[\underline{\eta}, \overline{\eta}]$.

Let

$$b(\theta) = \varphi(\overline{x}_1; \theta) - c\overline{x}_1$$

and

$$s(\eta) = \xi(\overline{x}_1; \eta)$$

Let $G(b)$ and $F(s)$ be the CDFs of the variables b and s induced by the random variables θ and η.

Assume that Mr 2 has the right not to be infringed by Mr 1's consumption, so that without bargaining the outcome will be $x_1 = 0$.

This outcome will be inefficient whenever

$$b(\theta) > s(\eta)$$

Now consider the bargaining problem, where Mr 1 offers Mr 2 a transfer T in exchange for letting him consume $\overline{x}_1$.

Mr 1 knows b but does not know s.

Mr 2 will agree if and only if $T \geq s$. The probability of this is $F(T)$.

Thus, given b, Mr 1 will choose T to solve

$$\max_{T} F(T)(b - T)$$

The objective function takes on value 0 at $T = b$ and is strictly positive at $T \in (s(\underline{\eta}), b)$.

Therefore the solution to the problem T_b^* is such that $T_b^* \in (s(\underline{\eta}), b)$ if $s(\underline{\eta}) < b$.

This implies inefficiency since Mr 2 will reject the offer whenever

$$b > s > T_b^*.$$

3. Price vs. Quantity Regulations

In our simple models of externalities, government intervention in either the form of a tax or a quota could restore efficiency.

In more realistic models, there is a difference between price and quantity regulation.

This was pointed out in by Weitzman *REStud* (1974).

This paper is a classic in the theory of externalities and in environmental economics more generally.

Weitzman's Model

The government is concerned with pollution reduction

Let q denote the amount of pollution reduction

$C(q)$ is the cost of pollution reduction with $C' > 0$ and $C'' > 0$

$B(q)$ is the benefit of pollution reduction with $B' > 0$ and $B'' < 0$

It is assumed that $B'(0) > C'(0)$, so that some amount of pollution reduction is socially desirable

The optimal level of pollution reduction is

$$q^* = \arg\max\{B(q) - C(q)\}$$

which implies that $B'(q^*) = C'(q^*)$.

The government can achieve q^* in two ways

1. *quantity regulation*–order firms or industry to do q^* units of reduction

2. *price regulation*–pay firms a price $p^* = B'(q^*)$ for each unit of reduction undertaken (this price could be avoiding a tax)

Firms would then choose

$$q = \arg\max\{p^* q - C(q)\}$$

Notice that this framework is quite general - we can convert our simple models of externalities from last time into the framework.

Consumer-consumer Example

Define $q = x_1^* - x_1$ where x_1^* is the market equilibrium level of x_1.

Also define the benefit and cost functions as following:

$$
\begin{aligned}
B(q) &= \xi(x_1^*) - \xi(x_1^* - q) \\
C(q) &= \{\phi(x_1^*) - cx_1^*\} - \{\phi(x_1^* - q) - c(x_1^* - q)\} \\
&= \phi(x_1^*) - \phi(x_1^* - q) - cq
\end{aligned}
$$

Mr 1 given the tax t would choose

$$
q = \arg\max\{tq - [\phi(x_1^*) - \phi(x_1^* - q) - cq]\}
$$

$$
\text{(FOC) } t + c = \phi'(x_1^* - q)
$$

Thus,

$$
\begin{aligned}
B'(q) &= \xi'(x_1^* - q) \\
C'(q) &= \phi'(x_1^* - q) - c
\end{aligned}
$$

Uncertainty

Suppose there is uncertainty in the costs and benefits of pollution reduction, so that benefits are $B(q,\theta)$ and costs are $C(q,\eta)$, where θ and η are realizations of independent random variables.

If the government could observe the realizations of θ and η then it could achieve the first best reduction in pollution by either a quantity regulation

$$q^*(\theta,\eta) = \arg\max\{B(q,\theta) - C(q,\eta)\}$$

or a price regulation

$$p^*(\theta,\eta) = \partial B(q^*(\theta,\eta),\theta)/\partial q$$

In reality, the government is unlikely to be able to condition policy on the realizations of these shocks.

More realistically, the government must commit in advance to a required level of pollution reduction q or a price p.

In this case, the optimal quantity of pollution reduction is

$$q^* = \arg\max E[B(q,\theta) - C(q,\eta)]$$
$$\text{(FOC) } E[\partial B(q^*,\theta)/\partial q] = E[\partial C(q^*,\eta)/\partial q]$$

How about the optimal price regulation?

Given the price regulation p, firms will choose

$$\widehat{q}(p,\eta) = \arg\max\{pq - C(q,\eta)\}$$
$$\text{(FOC) } p = \partial C(\widehat{q},\eta)/\partial q$$

Thus, the optimal price regulation is

$$p^* = \arg\max E[B(\widehat{q}(p,\eta),\theta) - C(\widehat{q}(p,\eta),\eta)]$$

In general it is not going to be the case that $\partial B(q^*, \theta)/\partial q = \partial C(q^*, \eta)/\partial q$ nor that $\partial B(\widehat{q}, \theta)/\partial q = \partial C(\widehat{q}, \eta)/\partial q$.

Thus, we are in the world of second best and the question is which policy instrument is better.

Define the advantage of price over quantity regulation as

$$\Delta = E[\{B(\widehat{q}(p^*, \eta), \theta) - C(\widehat{q}(p^*, \eta), \eta)\} \\ -\{B(q^*, \theta) - C(q^*, \eta)\}].$$

Assume the following functional forms for benefits and costs in a neighborhood of q^* (these are justified as second order Taylor approximations around the point q^*)

$$\begin{aligned} B(q, \theta) =& b(\theta) + (B' + \beta(\theta))(q - q^*) + \frac{B''}{2}(q - q^*)^2 \\ C(q, \eta) =& a(\eta) + (C' + \alpha(\eta))(q - q^*) + \frac{C''}{2}(q - q^*)^2 \\ E[\beta(\theta)] =& E[\alpha(\eta)] = 0 \end{aligned}$$

Observe that

$$
\begin{aligned}
\partial B(q,\theta)/\partial q &= B' + \beta(\theta) + B''(q - q^*) \\
\partial^2 B(q,\theta)/\partial q^2 &= B'' \\
\partial C(q,\eta)/\partial q &= C' + \alpha(\eta) + C''(q - q^*) \\
\partial^2 C(q,\eta)/\partial q^2 &= C''.
\end{aligned}
$$

This implies

$$
\begin{aligned}
E[\partial B(q^*,\theta)/\partial q] &= B' \\
E[\partial C(q^*,\eta)/\partial q] &= C' \\
E[\partial^2 B(q^*,\theta)/\partial q^2] &= B'' \\
E[\partial^2 C(q^*,\eta)/\partial q^2] &= C''
\end{aligned}
$$

Also let

$$\begin{aligned}\sigma^2 &= E[\alpha^2(\eta)] \\ &= E[\{\partial C(q^*,\eta)/\partial q - E[\partial C(q^*,\eta)/\partial q]\}^2] \\ &= \text{variance of MC at } q^*\end{aligned}$$

Weitzman's Theorem:

$$\Delta = \frac{\sigma^2}{2(C'')^2}[B'' + C'']$$

The Theorem implies that $\Delta > 0$ (i.e. price regulation is better than quantity regulation) if and only if $C'' > |B''|$.

To understand the result, consider the special case in which θ is constant; i.e., there is no uncertainty in benefits.

Suppose η can take on three values $\eta_1 < \eta_2 < \eta_3$ such that

$$\partial C(q,\eta_1)/\partial q > \partial C(q,\eta_2)/\partial q > \partial C(q,\eta_3)/\partial q$$

Suppose that the government sets policies assuming that η_2 will arise

Case 1: If $\partial B/\partial q$ is flatter, i.e. $|B''| < |C''|$, then we can see diagrammatically that the deadweight cost of quantity regulation is greater than that of price regulation when η_1 or η_3 is realized.

Intuitively, when $\partial B/\partial q$ is flatter, quantity ranges over a wider interval than price, thus making the outcome harder to control with quantity regulation. See Figure 2.

Case 2: If $\partial B/\partial q$ is steeper, i.e. $|B''| > |C''|$, then the deadweight cost of quantity regulation is smaller than that of price regulation.

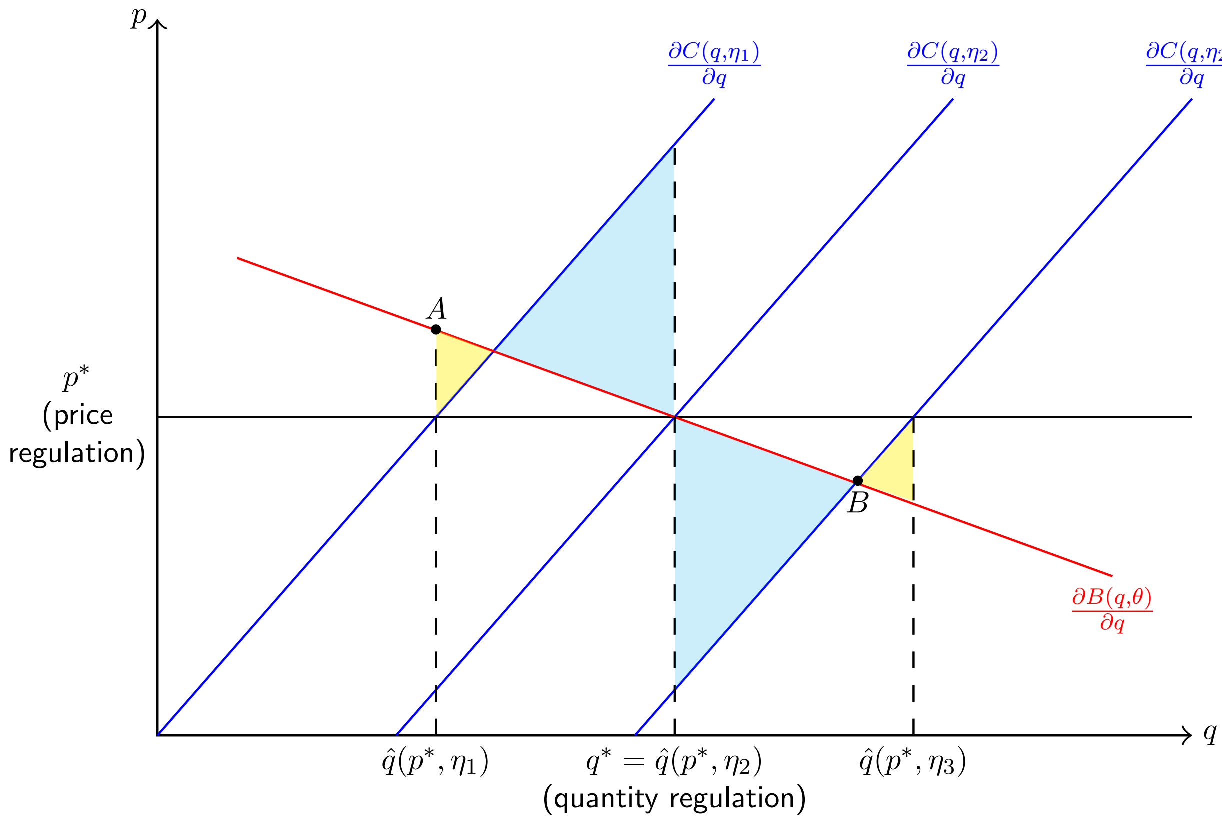

Figure 2: Case I: flat Marginal Benefit

A is an optimal point under η_1 with a quantity regulation $\hat{q}(p^*, \eta_1)$. B is optimal under η_3. Blue areas represent a dead-weight-loss with quantity regulation under η_1, and η_3, respectively. Yellow areas represent a dead-weight-loss with price regulation under η_1, and η_3, respectively.

When $\partial B/\partial q$ is steeper, price ranges over a wider interval than quantity and quantity regulation turns out to be better.

What determines whether $|B''|$ is greater or less than $|C''|$?

We might expect that $|B''|$ is large when we are dealing with pollution with *threshold effects*; i.e., there is some critical level of pollution above which things get bad, but below it things are o.k..

4. Tradeable Permits

A popular idea among environmental economists is "cap and trade".

The idea is to cap the total amount of pollution and allow firms to trade permits to pollute.

The idea is useful in environments where government is unwilling to directly tax pollution and has incomplete information on firms' costs of pollution abatement.

The issues may be illustrated in an extension of our producer-consumer externality model.

Model

There are n identical consumers with utility function

$$u = z + \varphi(x) - D(E).$$

Consumers have an endowment of the numeraire good y.

There are m competitive firms producing good x.

Each unit of x produced emits $\gamma(a)$ units of pollution where a denotes pollution abatement efforts.

Assume that $\gamma(0) = \gamma$, $\gamma'(a) < 0$, and $\gamma''(a) > 0$; for example,

$$\gamma(a) = \frac{\gamma}{1+a}.$$

The cost of producing a unit of x (in terms of the numeraire) when pollution abatement effort is a is $c + \eta a$.

Letting Z and X denote aggregate amounts of good z and good x, the efficiency problem is

$$\max \tfrac{Z}{n} + \varphi(\tfrac{X}{n}) - D(\gamma(a)X)$$
$$s.t.\ Z + (c + \eta a)X = ny$$

The efficient levels of X and a solve

$$\max y - (c + \eta a)\frac{X}{n} + \varphi(\frac{X}{n}) - D(\gamma(a)X)$$

They satisfy the first order conditions

$$\varphi'(\frac{X^e}{n}) = c + \eta a^e + nD'(\gamma(a^e)X^e)\gamma(a^e)$$

and

$$-nD'(\gamma(a^e)X^e)\gamma'(a^e) = \eta$$

The equilibrium level of X satisfies the first order condition

$$\varphi'(\frac{X^*}{n}) = c$$

and the equilibrium level of a is $a^* = 0$.

Government Intervention

The government cannot achieve efficiency by imposing a quota or a tax on X - it needs to regulate pollution directly.

It can achieve efficiency by either imposing a tax $t = nD'(\gamma(a^e)X^e)$ on emissions or providing each firm with a permit to emit at most $\rho = \gamma(a^e)X^e/m$ units of pollution.

You should verify this is the case.

This illustrates an imortant principle: it is always optimal to directly regulate the source of the externality.

Now suppose that there are two types of firms.

$m/2$ of the firms are as we have just described and the remaining $m/2$ are unable to abate their emissions (i.e., $\gamma(a) = \gamma$ for all a).

Suppose further that the government is unable to distinguish between the two types of firms.

The solution to the efficiency problem is unchanged.

All production of x should be done by the producers who can abate, but that creates no complications under the assumed CRS technology.

The efficient solution could be achieved by a tax as before, but the permit solution becomes problematic.

If the government provides each firm with a permit to emit at most $\rho = \gamma(a^e)X^e/m$ units of pollution, the efficient allocation will not be achieved.

The firms who cannot abate will continue to produce.

This difficulty is resolved by allowing the permits to be tradeable.

In equilibrium, if permits are tradeable, all the firms who cannot abate will sell their permits to those firms who can abate and the efficient allocation will be reached.

Please try to verify this.

5. Optimal Taxation with Externalities

How does the presence of externalities impact optimal taxation?

There is a literature on this dating back to Sandmo (1975) who was the first to incorporate an externality into an optimal commodity tax model.

It is straightforward to derive his result given what we have already done.

We will incorporate a producer-consumer externality into our optimal commodity taxation model.

To simplify, we focus on identical consumers.

Model

There are I identical consumers, indexed by $i = 1, ..., I$.

There are n consumer goods, indexed by $j = 1, ..., n$

Denote labor by l

Each consumer's utility function is

$$u(x_1^i, ..., x_n^i, l^i) - D(X_n).$$

where X_n is the aggregate output of good n and $D(\cdot)$ is a damage function.

Thus, the production of good n creates a negative externality.

Each good j is produced from labor with a linear technology, $X_j = l_j$

Assume competitive production so that the producer price of good j is $p_j = w$, where w is the wage rate.

Without loss of generality, let $w = 1$, which means that $p_j = 1$ for all j.

The government needs to hire T units of labor and therefore needs T units of tax revenue.

To raise the revenue, the government imposes linear taxes on goods $j = 1, ..., n$. These are $(t_1, ..., t_n)$.

Let $q = (1 + t_1, ..., 1 + t_n)$ denote the post-tax price vector.

Each consumer's problem given the price vector q and the wage w is

$$\max_{(x,l)}\{u(x,l) - D(X_n) : q \cdot x \leq wl + R\}$$

where $x = (x_1, ..., x_n)$ and R denotes non-labor income (which will be zero in the model).

This assumes that each consumer takes X_n as given when choosing his own demands.

Each consumer's indirect utility function can be written as $V(q, w, R) - D(X_n)$ and his demand function is denoted $x(q, w, R)$.

Since $w = 1$ and $R = 0$, we will just write $V(q) - D(X_n)$ and $x(q)$ in what follows.

The government's problem is:

$$\max_q I\left[V(q) - D(Ix_n(q))\right],$$

$$s.t. \sum_j^n (q_j - 1) I x_j(q) = T.$$

Analysis

The Lagrangian for the problem is

$$L = V(q) - D(Ix_n(q)) + \lambda[\sum_j^n (q_j - 1)x_j(q) - \frac{T}{I}]$$

The first order conditions are

$$\frac{\partial V}{\partial q_k} - ID' \frac{\partial x_n}{\partial q_k} = -\lambda(x_k + \sum_j^n t_j \frac{\partial x_j}{\partial q_k}) \text{ for } k = 1, ..., n$$

By *Roy's Identity*,

$$\frac{\partial V}{\partial q_k} = -\alpha x_k$$

where α is the marginal utility of income of a consumer; that is, $\alpha = \partial V / \partial R$.

Using this definition:

$$\alpha x_k + D'I \frac{\partial x_n}{\partial q_k} = \lambda(x_k + \sum_j^n t_j \frac{\partial x_j}{\partial q_k}) \text{ for } k = 1, ..., n$$

Recall the *Slutsky equation*:

$$\frac{\partial x_j}{\partial q_k} = s_{jk} - x_k \frac{\partial x_j}{\partial R}$$

where s_{jk} is the derivative of the *compensated* demand function.

We can now write:

$$\sum_{j=1}^n t_j s_{jk} - \frac{D'I}{\lambda} s_{nk} =$$

$$\frac{\alpha x_k}{\lambda} - x_k + x_k \left(\sum_j^n t_j \frac{\partial x_j}{\partial R} - \frac{D'I}{\lambda} \frac{\partial x_n}{\partial R} \right).$$

Define

$$
\begin{aligned}
b &= \frac{\alpha}{\lambda} + \sum_{j}^{n} t_j \frac{\partial x_j}{\partial R} - \frac{D'I}{\lambda}\frac{\partial x_n}{\partial R} \\
&= \text{Net } SMU \text{ of a consumer's income}
\end{aligned}
$$

Using the symmetry of the Slutsky matrix, we get

$$
\sum_{j=1}^{n} t_j s_{kj} - \frac{D'I}{\lambda} s_{kn} = -x_k\,(1-b)
$$

Dividing through by x_k we obtain

$$
-\left(\frac{\sum_{j=1}^{n} t_j s_{kj} - \frac{D'I}{\lambda} s_{kn}}{x_k}\right) = 1 - b
$$

This is analogous to the basic Ramsey Rule that we derived earlier.

To interpret the formula, define $\{t_{jp}\}_{j=1}^{n}$ as follows:

$$t_{jp} = \begin{cases} 0 & \text{if } j \neq n \\ \frac{D'I}{\lambda} & \text{if } j = n \end{cases}$$

These are interpreted as the optimal Pigouvian tax rates.

This is because if we solved the problem with $T = 0$ and allowed the government to redistribute tax revenues back to consumers via a uniform lump sum transfer, the optimal taxes would satisfy these conditions (since $\lambda = \alpha$).

Next define $\{t_{jr}\}_{j=1}^{n}$ as the solutions to

$$-\frac{\sum_{j=1}^{n} t_{jr} s_{kj}}{x_k} = 1 - b \quad \text{for all } k$$

These are interpreted as the optimal Ramsey tax rates, since if we solved the problem with $D' = 0$, the optimal rates would satisfy these conditions.

Letting t_j denote the optimal Sandmo tax rate on good j, we have

$$t_j = \begin{cases} t_{jr} & \text{if } j \neq n \\ t_{nr} + t_{np} & \text{if } j = n \end{cases}$$

Thus, the optimal Sandmo tax rate can be expressed as the optimal Ramsey rate plus a Pigouvian correction.

The optimal tax structure is therefore characterized by an additivity property.

The marginal social damage of commodity n enters the tax formula for that commodity additively and does not enter the tax formulas for the other commodities.

Thus, the fact that a commodity involves a negative externality does not justify taxing other commodities that are complementary with it, or subsidizing substitutes.

This is an important conclusion.

6. Common Property Externalities

This type of externality arises whenever there is a commonly owned resource.

Fish in the ocean, forests in LDCs are classic examples.

There is quite a bit of work on this type of externality in environmental economics.

I will outline the theory of market failure in this context, drawing on Leach's *A Course in Public Economics*.

Static Common Property Model

For concreteness consider the fishing context and assume the following:

- Infinite number of potential fishermen

- Each fisherman operates one boat at cost w

- Total catch with b boats is

$$y(s, b)$$

where s is the current stock of fish. Assume that

$$y(s, 0) = 0, \partial y/\partial s > 0, \partial y/\partial b > 0, \partial^2 y/\partial b^2 < 0$$

- Per boat catch is $y(s, b)/b$

- $x(p)$ is the aggregate demand for fish with inverse $p(x)$

Optimal Number of Boats

The optimal number of boats maximizes aggregate surplus, which is consumer surplus plus fishermen surplus.

Formally, the problem is

$$\max_b \int_0^{y(s,b)} p(x)dx - wb$$

The first order condition is:

$$\text{(FOC)}\ p(y(s,b^o))\frac{\partial y}{\partial b}(s,b^o) = w$$

This just says that the *social marginal benefit* of boats is equalized with the *social marginal cost*.

Equilibrium Number of Boats

An *equilibrium* is a price of fish and a number of boats such that demand equals supply of fish and fishermen make zero profits.

Formally, (p^*, b^*) is an equilibrium if

$$\text{(i) } x(p^*) = y(s, b^*) <=> p^* = p(y(s, b^*))$$

and

$$\text{(ii) } p^*(\frac{y(s, b^*)}{b^*}) = w$$

Result: *The equilibrium number of boats is higher than the optimal number; i.e.,* $b^* > b^0$

To see this note that in equilibrium

$$p(y(s, b^*))\frac{y(s, b^*)}{b^*} = w$$

But we know that the average product exceeds the marginal product; i.e., $y(s, b)/b > \partial y(s, b)/\partial b$, and hence

$$p(y(s, b^*))\frac{\partial y(s, b^*)}{\partial b^*} < w$$

So at equilibrium there are too many boats.

Diagrammatically, this is easy to show for $p = 1$ (i.e. perfectly elastic demand curve.)

Dynamic Common Property Model

A limitation of the static model is that it takes the stock of the resource as given.

In many applications the key question is what will happen to the stock of resource over time.

It is easy to extend the model to make it dynamic and endogenize the stock.

The stock of fish is the state variable and we assume that it evolves according to the equation

$$s_{t+1} = s_t + g(s_t) - y(s_t, b_t)$$

where $g(s_t)$ is the growth of new stock with $g(0) = 0$, $g' > 0$, $g'' < 0$.

Let the initial stock be s_1.

Planner's Problem

Consider first the planning problem of choosing the optimal number of boats in each period.

Formally, the problem is to choose $(b_t)_{t=1}^{\infty}$ to maximize discounted sum of surplus

$$\max_{(b_t)_{t=1}^{\infty}} \sum_{t=1}^{\infty} \delta^{t-1} \{ \int_0^{y(s_t,b_t)} p(x)dx - wb_t \}$$

$$s.t. s_{t+1} = s_t + g(s_t) - y(s_t, b_t)$$

where $\delta \in (0,1)$ is the discount rate.

We can solve this via recursive methods (see Stokey, Lucas and Prescott (1989)).

Think of the planner as choosing, given the current stock is s, what he wants the future stock x to be - this will imply what the number of boats has to be.

More precisely, define the function $b(x,s)$ implicitly from the equation

$$x = s + g(s) - y(s,b).$$

If the current stock of fish is s and the planner wants the stock next period to be x, the number of boats this period must be $b(x,s)$.

Note that

$$\frac{\partial b(x,s)}{\partial x} = -\frac{1}{\partial y(s,b)/\partial b} < 0$$

which tells us that the higher the future stock the lower the number of boats.

In addition,

$$\frac{\partial b(x,s)}{\partial s} = \frac{1+g'(s)-\partial y(s,b)/\partial s}{\partial y(s,b)/\partial b} > 0$$

which tells us that the higher the current stock the higher the number of boats

We can formulate the planner's problem recursively as

$$\max_x \int_0^{y(s,b(x,s))} p(z)dz - wb(x,s) + \delta V(x)$$

where $V(x)$ is the planner's *value function*.

The planner's value function satisfies the functional equation

$$V(s) = \max_x \{\int_0^{y(s,b(x,s))} p(z)dz - wb(x,s) + \delta V(x)\}$$

Denote the optimal policy function as $x^0(s)$.

The first order condition for the maximization problem implies

$$\delta V'(x^0) = -[p(y(s, b(x^0, s)))\frac{\partial y(s,b(x^0,s))}{\partial b} - w]\frac{\partial b(x^0,s)}{\partial x}$$

This says that

$$MB \text{ of additional stock} = MC \text{ (social value of lost fish consumption today)}$$

So lets look for the planner's steady state, i.e. a stock s^0 such that

$$x^0(s^0) = s^0$$

The level of boats will be

$$b^0 = b(s^0, s^0)$$

To find this, we need an expression for $\delta V'(x)$.

We know that

$$V(x) = \max_{\omega}\{\int_0^{y(x,b(\omega,x))} p(z)dz - wb(\omega,x) + \delta V(\omega)\}$$

Thus, by the *Envelope Theorem*,

$$V'(x) = p(y(x, b(x^0(x), x)))\frac{\partial y(x,b(x^0(x),x))}{\partial s}$$
$$+\{p(y(x, b(x^0(x), x)))\frac{\partial y(x,b(x^0(x),x))}{\partial b} - w\}\frac{\partial b(x^0(x),x)}{\partial s}$$

The first term in the expression captures the value of additional consumption next period while the second term captures the induced effect on the number of boats.

Thus, in steady state, letting $p^0 = p(y(s^0, b^0))$, we have:

$$\begin{aligned} & \delta[p^0 \frac{\partial y(s^0, b^0)}{\partial s} + (p^0 \frac{\partial y(s^0, b^0)}{\partial b} - w) \frac{\partial b(s^0, s^0)}{\partial s}] \\ = & -[p^0 \frac{\partial y(s^0, b^0)}{\partial b} - w] \frac{\partial b(s^0, s^0)}{\partial x} \end{aligned}$$

This can be solved for s^0.

Note that, at a steady state,

$$p^0 \frac{\partial y(s^0, b^0)}{\partial b} > w.$$

This differs from the static model because we have to take into account the impact of more boats on the future stock.

Market Equilibrium

An *equilibrium steady state* is a price p^*, a stock s^* and a number of boats b^* such that:

$$\text{(i) } x(p^*) = y(s, b^*)$$

$$\text{(ii) } \; p^*(\frac{y(s^*, b^*)}{b^*}) = w$$

$$\text{(iii) } \; g(s^*) = y(s^*, b^*)$$

(i) is demand equals supply; (ii) is fishermen make zero profits and (iii) is that the growth rate is sufficient to replenish the stock.

An interesting question is whether there exists a steady state with a positive stock or whether the outcome will be extinction - see Leach on this.

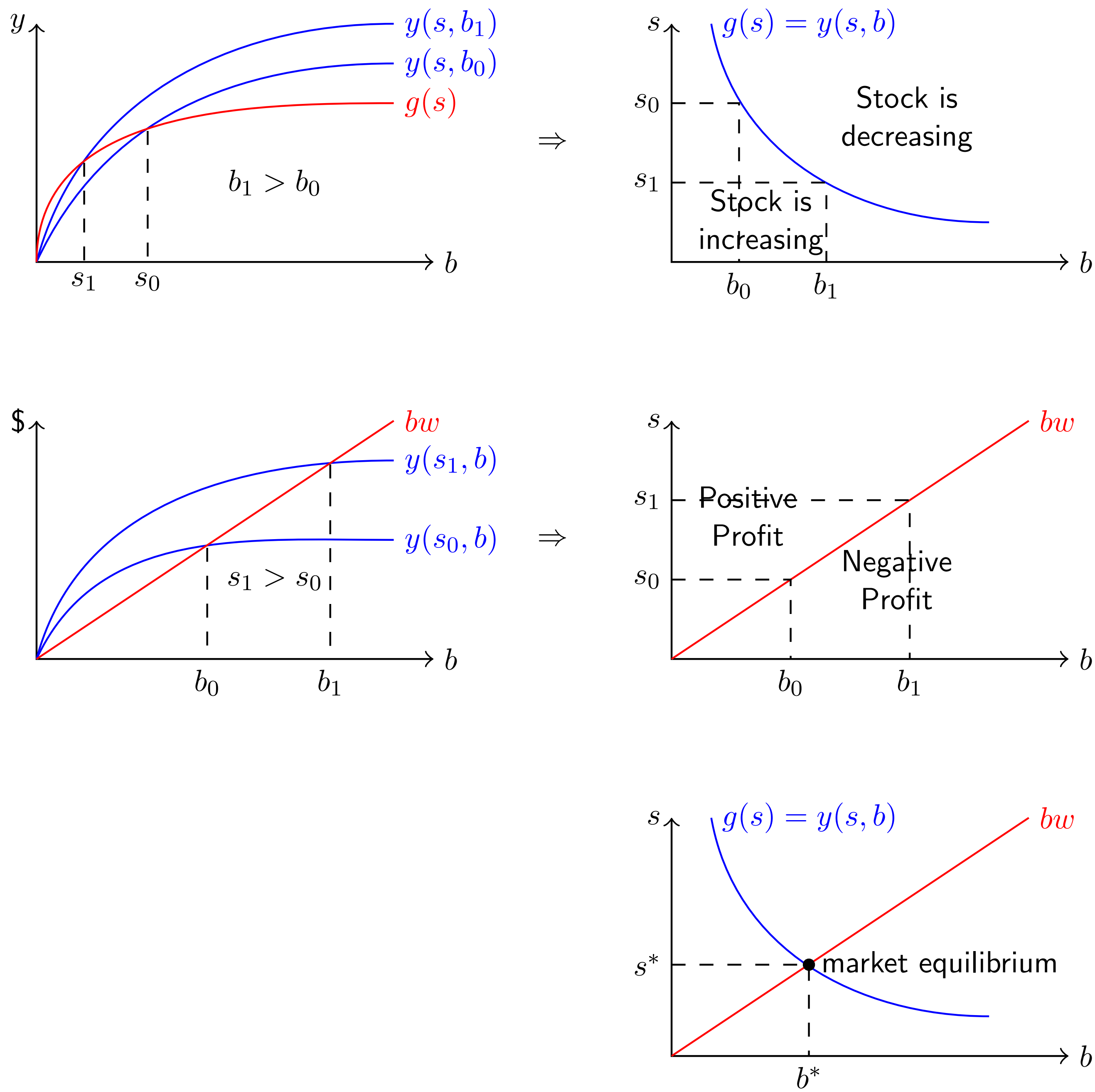

Figure 3: Stable market equilibrium on the Dynamic Common Property model

Comparing Equilibrium and Optimum

Lets study this diagrammatically, assuming that $p^* = 1$.

Condition (iii) implies negative relation between s and b. (Plot $g(s)$ and $y(s, b)$ in a diagram with s on the horizontal axis. By varying b, we can get s corresponding to each b.)

Condition (ii) implies positive relation between s and b.

In $s - b$ space, we can plot the implications of conditions (ii) and (iii) and find the market steady state. In the same diagram, we can put on the planner's steady state.

The market equilibrium will be characterized by:

i) too many boats

ii) too little stock of fish.

IV. Dynamic Optimal Taxation

Thinking through the problem of optimal taxation in dynamic economies raises a host of interesting new issues.

We will discuss:

(i) the idea of Ricardian Equivalence;

(ii) the Tax Smoothing Problem;

(iii) the Optimal Taxation of Capital;

(iv) the Problem of Time Inconsistency;

and (v) the New Dynamic Public Finance.

1. Ricardian Equivalence

Discussion of the stimulus package suggests that a debt-financed tax cut will have positive effects on the economy.

However, there is a tradition in macroeconomics that argues that the timing of taxes should not matter.

This is the idea of *Ricardian Equivalence*, named after the nineteenth century economist David Ricardo who first proposed the idea.

In modern form, it was revitalized by Barro in a 1974 article in *JPE* "Are Government Bonds Net Wealth?".

The idea of Ricardian Equivalence is simple: if government spending is financed by debt, then this will have to be paid back eventually and this means future taxes will go up.

A citizen's consumption path will be determined by his life time budget constraint and he will anticipate these higher future taxes.

Thus, a debt-financed tax cut will not alter the citizen's lifetime budget constraint and hence will not alter his consumption path.

Despite the idea's simplicity, it is worth seeing it derived formally.

We will follow the exposition in Ljungqvist and Sargent Chp 10.

Model

The economy consists of a single infinitely lived representative household.

There is a single consumption good and the household has preferences over consumption streams given by

$$\sum_{t=0}^{\infty} \beta^t u(c_t) \qquad (1)$$

where c_t is consumption in period t, $\beta \in (0,1)$ is the discount rate, and $u(\cdot)$ is increasing, strictly concave, and satisfies the Inada conditions.

Each household has an endowment sequence $\{y_t\}_{t=0}^{\infty}$ with finite present value.

There is a single risk-free asset bearing a fixed gross one period return of $R > 1$. We assume $\beta R = 1$.

Think of the asset as either a risk-free loan to foreigners or the government.

There is a government with a stream of government consumption requirements $\{g_t\}_{t=0}^{\infty}$.

The government imposes a stream of lump sum taxes $\{\tau_t\}_{t=0}^{\infty}$ on the household.

The government can also borrow and lend in the asset market.

Households

Letting b_t denote the household's ownership of the asset at time t, the household faces the sequence of budget constraints

$$c_t + \frac{b_{t+1}}{R} = y_t - \tau_t + b_t. \qquad (2)$$

The household's initial asset ownership b_0 is given.

Note that $b_t < 0$ when the household is borrowing.

When choosing $\{b_t\}_{t=0}^{\infty}$ the household faces the sequence of constraints

$$b_t \geq \widetilde{b}_t \qquad (3)$$

where

$$\widetilde{b}_t = \sum_{j=0}^{\infty} R^{-j}(\tau_{t+j} - y_{t+j}).$$

These constraints restrict the household to borrow no more than it is feasible for it to repay.

Note that even with $c_t = 0$ for ever, $\widetilde{b}_t$ is the maximum that the household can repay.

The household chooses a plan $\{c_t, b_{t+1}\}_{t=0}^{\infty}$ to maximize (1) subject to (2) and (3).

Government

Letting B_t denote the amount of government debt at time t, the government faces the budget constraint

$$\tau_t + \frac{B_{t+1}}{R} = g_t + B_t. \qquad (4)$$

If $B_t < 0$ the government is lending to the household or foreigners.

The government's period 0 debt B_0 is given.

We rule out the government running Ponzi schemes by imposing the transversality condition

$$\lim_{T \to \infty} R^{-T} B_{t+T} = 0.$$

Equilibrium

Definition: Given initial conditions (b_0, B_0), an *equilibrium* is a household plan $\{c_t, b_{t+1}\}_{t=0}^{\infty}$ and a government policy $\{g_t, \tau_t, B_{t+1}\}_{t=0}^{\infty}$ such that: (i) the government plan satisfies the government budget constraint (4), and, (ii) given $\{\tau_t\}_{t=0}^{\infty}$ the household's plan is optimal.

Ricardian Equivalence Proposition: *Given initial conditions* (b_0, B_0)*, let* $\{c_t^*, b_{t+1}^*\}_{t=0}^{\infty}$ *and* $\{g_t, \tau_t^*, B_{t+1}^*\}_{t=0}^{\infty}$ *be an equilibrium. Consider any other tax policy* $\{\tau_t^o\}_{t=0}^{\infty}$ *satisfying*

$$\sum_{t=0}^{\infty} R^{-t}\tau_t^* = \sum_{t=0}^{\infty} R^{-t}\tau_t^o. \qquad (5)$$

Then, $\{c_t^*, b_{t+1}^o\}_{t=0}^{\infty}$ *and* $\{g_t, \tau_t^o, B_{t+1}^o\}_{t=0}^{\infty}$ *is also an equilibrium where*

$$b_t^o = \sum_{j=0}^{\infty} R^{-j}(c_{t+j}^* + \tau_{t+j}^o - y_{t+j}) \qquad (6)$$

and

$$B_t^o = \sum_{j=0}^{\infty} R^{-j}(\tau_{t+j}^o - g_{t+j}). \qquad (7)$$

Proof: We first show that $\{c_t^*, b_{t+1}^o\}_{t=0}^{\infty}$ solves the household's problem under the altered government tax scheme $\{g_t, \tau_t^o, B_{t+1}^o\}_{t=0}^{\infty}$.

In general, the household's problem can equivalently be posed as choosing at time 0 a consumption plan $\{c_t\}_{t=0}^{\infty}$ to maximize (1) subject to the single present value budget constraint

$$b_0 = \sum_{t=0}^{\infty} R^{-t}(c_t - y_t) + \sum_{t=0}^{\infty} R^{-t}\tau_t. \qquad (8)$$

The household's asset holdings in each period t are then obtained by solving forward the budget constraint (2) to obtain

$$b_t = \sum_{j=0}^{\infty} R^{-j}(c_{t+j} + \tau_{t+j} - y_{t+j}). \qquad (9)$$

Observe that constraint (3) is satisfied since

$$\begin{aligned} b_t &= \sum_{j=0}^{\infty} R^{-j}(c_{t+j} + \tau_{t+j} - y_{t+j}) \\ &\geq \sum_{j=0}^{\infty} R^{-j}(\tau_{t+j} - y_{t+j}) = \widetilde{b}_t. \end{aligned}$$

Condition (5) implies that the household's present value budget constraint (8) is the same under $\{g_t, \tau_t^*, B_{t+1}^*\}_{t=0}^{\infty}$ as under $\{g_t, \tau_t^o, B_{t+1}^o\}_{t=0}^{\infty}$ and hence $\{c_t^*\}_{t=0}^{\infty}$ solves the household's problem.

Moreover, from (9), we see that the household's asset holdings are indeed given by (6).

It remains to show that the altered government tax and borrowing plans satisfy the government's budget constraint (4).

In general, the government's budget constraint (4) can be converted into a single present value budget constraint

$$B_0 = \sum_{t=0}^{\infty} R^{-t}(\tau_t - g_t). \qquad (10)$$

Government debt in each period is then obtained by solving forward the budget constraint (4) to obtain

$$B_t = \sum_{j=0}^{\infty} R^{-j}(\tau_{t+j} - g_{t+j}). \qquad (11)$$

Thus, given (5), the present value of taxes is the same under $\{\tau_t^*\}_{t=0}^{\infty}$ as under $\{\tau_t^o\}_{t=0}^{\infty}$ and (10) is satisfied.

By (11), the adjusted borrowing plan that makes (4) satisfied in each period is given by (7). ■

Discussion

An objection to this argument is that, in reality, households are not infinitely-lived and a debt-financed tax cut might transfer resources from future to current generations.

Barro's paper pointed out that the same argument goes through in an overlapping generations model provided that generations are altruistically linked via a bequest motive.

The key assumption is that the bequest motive is operational in the sense that bequests are always positive.

The Ricardian Equivalence Proposition remains true provided that the initial equilibrium has an operational bequest motive for all t and that the new tax policy must not be so different from the initial one that it renders the bequest motive inoperative.

Ljungqvist and Sargent Chp 10 demonstrate this in a model with a sequence of one-period-lived agents.

In each period t there is a one-period-lived agent who values consumption and the utility of his direct descendant.

Preferences of the period t agent are

$$u(c_t) + \beta V(b_{t+1})$$

where $b_{t+1} \geq 0$ denotes bequests to the period $t+1$ agent and $V(b_{t+1})$ is the maximized utility function of the period $t+1$ agent; i.e.,

$$\begin{aligned} V(b_{t+1}) &= \max_{(c_{t+1}, b_{t+2}) \in \Re^2_+} \{u(c_{t+1}) + \beta V(b_{t+2}) \\ &: \quad c_{t+1} + \frac{b_{t+2}}{R} \leq y - \tau_{t+1} + b_{t+1}\}. \end{aligned}$$

If the initial equilibrium does not have positive bequests, then timing of taxes does matter.

Suppose, for example, that the period t agent is not leaving bequests to the period $t+1$ agent.

If the government cuts taxes in period t and finances this by raising taxes in period $t+1$, the period t agent's consumption will increase and the period $t+1$ agent's consumption will decrease.

For more discussion of the idea of Ricardian Equivalence see Barro's paper "The Ricardian Approach to Budget Deficits."

2. Tax Smoothing

The Ricardian Equivalence Proposition assumes the government employs *lump sum* taxes.

If taxes are distortionary, the result is not true.

This is because the distortionary costs of taxation tend to be convex in the tax rate.

Thus, if the government cranks up the marginal rate of taxation on income to 100% one period and then reduces it to 0% the next period, that would not be equivalent to keeping the tax rate at 50% in each period.

Given that the Ricardian Equivalence Proposition is not valid when taxes are distortionary, the issue of how government should use debt and taxes to finance government expenditure is an interesting one.

Barro (1979) *JPE* offers a famous analysis of this problem.

His theory makes the following assumptions:

- government spending needs vary over time

- government has available a single distortionary tax

- distortionary costs of taxation are a convex function of the tax rate

His theory implies that the government should use budget surpluses and deficits as a buffer to prevent tax rates from changing too sharply.

Thus, the government should run a deficit in times of high spending needs and a surplus in times of low need.

While it is normative, empirical evidence supports this *tax smoothing theory.*

Historically, the debt/GDP ratio in the U.S. and U.K. tends to have increased in periods of high spending needs and decreased in periods of low needs.

To illustrate Barro's tax smoothing theory, I will use a simple tax smoothing model that is reminiscent of those used by Lucas and Stokey *JME* (1983) and Aiyagari et al *JPE* (2002).

This model forms the basis for my work on the political economy of fiscal policy with Marco Battaglini (see in particular "A Dynamic Theory of Public Spending, Taxation and Debt" *AER* 2008).

The model

The economy consists of a single infinitely-lived representative household.

There is a single (nonstorable) consumption good, z, produced using labor, l, with the linear technology $z = wl$.

There is a public good, g, produced from the consumption good according to the technology $g = z/p$.

The houshold consumes the consumption good, benefits from the public good, and supplies labor.

Each consumer's per period utility function is

$$z_t + A_t g_t^{\alpha} - \frac{l_t^{(1+1/\varepsilon)}}{\varepsilon + 1},$$

where $\alpha \in (0,1)$ and $\varepsilon > 0$.

The parameter A_t measures the value of the public good and it may vary across periods reflecting shocks to the society such as wars.

Assume for now that the entire sequence of shocks $\{A_t\}_{t=0}^{\infty}$ is known at time 0 and also that for all t, $A_t \in [\underline{A}, \overline{A}]$.

Citizens discount future per period utilities at rate δ.

There is a competitive labor market and competitive production of the public good.

Thus, in each period the wage rate is equal to w and the price of the public good is p.

There is also a market in risk-free one period bonds.

The assumption of a constant marginal utility of consumption implies that the equilibrium interest rate is $\rho = 1/\delta - 1$.

At this interest rate, consumers will be indifferent as to their allocation of consumption across time.

Government

The public good is provided by the government.

The government can raise revenue by levying a proportional tax on labor income.

It can also borrow and lend by selling and buying bonds.

Government policy in any period t is described by a triple $\{\tau_t, g_t, b_t\}$, where τ_t is the income tax rate; g_t is the public good; and b_t is the amount of bonds sold.

When b_t is negative, the government is buying bonds.

If the government sells b_t bonds in period t, it must repay $(1 + \rho)b_t$ in period $t + 1$.

In a period in which government policy is $\{\tau, g, b\}$, the household will supply an amount of labor

$$l^*(w(1-\tau)) = \arg\max_l \{w(1-\tau)l - \frac{l^{(1+1/\varepsilon)}}{\varepsilon+1}\}.$$

It is straightforward to show that

$$l^*(w(1-\tau)) = (\varepsilon w(1-\tau))^{\varepsilon},$$

so that ε is the elasticity of labor supply.

If the household simply consumes his net of tax earnings, he will obtain a per period utility of

$$u(w(1-\tau), g; A) = \frac{\varepsilon^{\varepsilon}(w(1-\tau))^{\varepsilon+1}}{\varepsilon+1} + Ag^{\alpha}.$$

Since the household is indifferent as to his allocation of consumption across time, his lifetime utility will equal that which he would obtain if he simply consumed his net earnings each period plus the value of his initial bond holdings.

At the beginning of period 0, the government inherits some amount of bonds to repay in the first period. This debt level is denoted by b_0.

Government policies must satisfy two feasibility constraints.

The first is that revenues must equal expenditures in each period.

Expenditure on public goods and debt repayment in period t is

$$pg_t + (1+\rho)b_{t-1}.$$

Tax revenues are

$$R(\tau_t) = \tau_t w l^*(w(1-\tau_t)) = \tau_t w(\varepsilon w(1-\tau_t))^{\varepsilon}$$

and total revenues are

$$R(\tau_t) + b_t.$$

Thus, the government's budget constraint in period t is that

$$pg_t + (1+\rho)b_{t-1} = R(\tau_t) + b_t.$$

The second constraint is that the amount of government borrowing must be feasible.

In particular, there is an upper limit on the amount the government can borrow given by $\overline{b} = \max_\tau R(\tau)/\rho$.

The optimal taxation problem

Given the initial level of government debt b_0 and the sequence of public good values $\{A_t\}_{t=0}^{\infty}$, the optimal taxation problem is to choose a sequence of policies $\{\tau_t, g_t, b_t\}_{t=1}^{\infty}$ to solve

$$\begin{array}{c} \max \sum_{t=1}^{\infty} \delta^{t-1} u(w(1-\tau_t), g_t; A_t) \\ s.t. \;\; pg_t + (1+\rho)b_{t-1} = R(\tau_t) + b_t \\ \text{and } b_t \leq \overline{b} \quad \text{for all } t. \end{array}$$

Tax Smoothing Proposition *The optimal policy is such that* $\tau_t = \tau^*$ *for all* t *and* $g_t = g^*(A_t, \tau^*)$ *where* $g^*(A, \tau)$ *satisfies*

$$\alpha A g^{\alpha-1} = [\frac{1-\tau}{1-\tau(1+\varepsilon)}]p.$$

The tax rate τ^* *is such that the initial level of debt equals the present value of future budget surpluses*

$$b_0 = \sum_{t=1}^{\infty} \delta^t [R(\tau^*) - pg^*(A_t, \tau^*)]$$

and the debt level in period t satisfies the equation

$$b_t = \sum_{j=1}^{\infty} \delta^j [R(\tau^*) - pg^*(A_{t+j}, \tau^*)].$$

Proof: Lets ignore the upper bound constraint on borrowing for now and form the Lagrangian

$$L = \sum_{t=1}^{\infty} \delta^{t-1} u(w(1-\tau_t), g_t; A_t)$$
$$\sum_{t=1}^{\infty} \lambda_t (R(\tau_t) + b_t - pg_t - (1+\rho)b_{t-1})$$

where λ_t is the multiplier on the period t budget constraint.

The first order conditions are

$$\delta^{t-1} \frac{\partial u(\cdot)}{\partial g_t} = \lambda_t p,$$

$$-\delta^{t-1} \frac{\partial u(\cdot)}{\partial \tau_t} = \lambda_t R'(\tau_t),$$

and

$$\lambda_t = (1+\rho)\lambda_{t+1}.$$

The first condition implies that

$$\alpha A_t g_t^{\alpha-1} = \frac{\lambda_t}{\delta^{t-1}} p.$$

The second condition implies that

$$\frac{\lambda_t}{\delta^{t-1}} = \frac{1-\tau_t}{1-\tau_t(1+\varepsilon)}.$$

The third condition implies that

$$\lambda_{t+1} = \delta \lambda_t.$$

We can use these conditions to conclude that the tax rate is constant through time; i.e., $\tau_t = \tau^*$ for all t.

Moreover, the public good level in each period is adjusted so that the social marginal benefit of public goods is equal to the marginal cost of public funds at the tax rate τ^*.

Thus, $g_t = g^*(A_t, \tau^*)$ where $g^*(A, \tau)$ is defined in the statement of the proposition.

The sequence of government budget constraints imply that τ^* is such that

$$b_0 = \sum_{t=1}^{\infty} \delta^t [R(\tau^*) - pg^*(A_t, \tau^*)]$$

and that

$$b_t = \sum_{j=1}^{\infty} \delta^j [R(\tau^*) - pg^*(A_{t+j}, \tau^*)].$$

Finally, note that the constraint that $b_t \leq \bar{b}$ is satisfied in each period since

$$\begin{aligned} b_t &= \sum_{j=1}^{\infty} \delta^j [R(\tau^*) - pg^*(A_{t+j}, \tau^*)] \\ &< \sum_{j=1}^{\infty} \delta^j R(\tau^*) \\ &\leq \max_{\tau} R(\tau)/\rho \end{aligned}$$

This completes the proof. ■

Note from the expressions in the proposition (and the government's period t budget constraint) that

$$b_t - \frac{b_{t-1}}{\delta} = pg^*(A_t, \tau^*) - R(\tau^*)$$

Thus, debt increases when A_t is high and decreases when A_t is low.

Stochastic Shocks

As in the model just considered, Barro's analysis assumed that spending needs were variable, but that the government had perfect foresight.

He conjectured what would happen with stochastic shocks.

The same principle should apply: in particular, he conjectured that taxes should obey a martingale; that is, $\tau_t = E\tau_{t+1}$.

Aiyagari et al (2002) study an infinite horizon, general equilibrium version of Barro's model with stochastic shocks.

They show that the tax smoothing logic does not necessarily imply a counter-cyclical theory of deficits and surpluses.

The optimal policy may be for the government to gradually acquire sufficient bond holdings to finance spending completely from interest earnings.

Excess interest earnings are rebated back to households via a cash transfer.

Taxes do not obey a martingale: they obey a supermartingale $\tau_t > E\tau_{t+1}$.

I will now illustrate these points by extending the model just analyzed to incorporate stochastic shocks in the value of the public good.

The extended model

Assume now that the value of the public good varies across periods in a random way.

Specifically, in each period, A is the realization of a random variable with range $[\underline{A}, \overline{A}]$ and cumulative distribution function $G(A)$.

Also assume that the government can also provide the household with a non-negative cash transfer denoted s_t (this allows government to rebate excess revenues back to households).

It is convenient to let

$$B(\tau_t, g_t, b_t; b_{t-1}) = R(\tau_t) - pg_t + b_t - (1+\rho)b_{t-1}.$$

We refer to B as the *net of transfer surplus*.

The government budget constraint can now be written as

$$s_t \leq B(\tau_t, g_t, b_t; b_{t-1}).$$

The optimal taxation problem

Given the initial level of government debt b_0, the optimal taxation problem is to choose a sequence of contingency plans $\{\tau_t(\cdot), g_t(\cdot), b_t(\cdot), s_t(\cdot)\}_{t=1}^{\infty}$ to maximize the representative consumer's expected utility

$$(1+\rho)b_0 + E\sum_{t=1}^{\infty}\delta^{t-1}[u(w(1-\tau_t), g_t; A_t) + s_t],$$

subject to the sequence of feasibility constraints

$$s_t \in [0, B(\tau_t, g_t, b_t; b_{t-1})] \text{ and } b_t \le \bar{b} \quad \text{for all } t.$$

If in any period t the net of transfer surplus $B(\tau_t, g_t, b_t; b_{t-1})$ were positive, the government would use it to finance transfers and hence

$$s_t = B(\tau_t, g_t, b_t; b_{t-1}).$$

Thus, we can rewrite the government's problem as choosing a sequence of contingent tax rate-public good-public debt plans $\{\tau_t(\cdot), g_t(\cdot), b_t(\cdot)\}_{t=1}^{\infty}$ to solve

$$\begin{gathered}\max E \sum_{t=1}^{\infty} \delta^{t-1}[u(w(1-\tau_t), g_t; A_t) \\ +B(\tau_t, g_t, b_t; b_{t-1})] \\ s.t. \;\; B(\tau_t, g_t, b_t; b_{t-1}) \geq 0 \\ \text{and } b_t \leq \overline{b} \quad \text{for all } t.\end{gathered}$$

This problem can be formulated recursively.

Let $v(b, A)$ denote the maximized value of the objective function when the initial level of debt is b and the value of the public good is A.

Then,

$$\begin{gathered}v(b, A) = \max_{(\tau, g, b')}\{u(w(1-\tau), g; A) + B(\tau, g, b'; b) \\ +\delta E v(b', A') : B(\tau, g, b'; b) \geq 0 \;\; \& \; b' \leq \overline{b}\}.\end{gathered}$$

Standard arguments can be applied to show that such a value function exists and that $Ev(\cdot, A)$ is differentiable and strictly concave.

From this, the properties of the optimal policies may be deduced.

The optimal policies

Let (b, A) be given. Letting λ denote the multiplier on the budget constraint, the first order conditions for the problem are:

$$1 + \lambda = \frac{1-\tau}{1-\tau(1+\varepsilon)},$$

$$\alpha A g^{\alpha-1} = [\frac{1-\tau}{1-\tau(1+\varepsilon)}]p,$$

and

$$\frac{1-\tau}{1-\tau(1+\varepsilon)} = -\delta E[\frac{\partial v(b', A')}{\partial b'}].$$

To interpret these, recall that $(1-\tau)/(1-\tau(1+\varepsilon))$ measures the cost of raising a \$1 of tax revenue via a tax hike.

The first condition says that the benefit of raising an additional unit of revenue - measured by $1+\lambda$ - must equal the marginal cost of taxation.

The second condition says that the marginal social benefit of the public good must equal its price times the marginal cost of taxation.

This is the *Samuelson Rule* modified to take into account the fact that taxation is distortionary.

The third condition says that the benefit of increasing debt in terms of reducing taxes must equal the marginal cost of an increase in the debt level.

There are two possibilities:

(i) at the optimal policy (τ, g, b') the government is providing a transfer (i.e., $B(\tau, g, b'; b) > 0$) so that $\lambda = 0$

(ii) at the optimal policy (τ, g, b') the government is not providing a transfer so that $\lambda > 0$.

Case (i)

The first condition implies that $\tau = 0$.

If $B(\tau, g, b'; b) > 0$ and $\tau > 0$ the government could reduce taxes and make up the lost revenue by reducing transfers: because taxes are distortionary, this would raise welfare.

The second condition implies that $g = g_S(A)$, where $g_S(A)$ satisfies the *Samuelson Rule*.

If $B(\tau, g, b'; b) > 0$, then extra units of the public good can be financed in a non-distortionary way by reducing transfers.

The third condition implies that $b' = b^*$ where

$$1 = -\delta E[\frac{\partial v(b^*, A')}{\partial b'}] \quad .$$

Intuitively, the marginal benefit of borrowing an additional unit is just 1 because the effect will be to just increase the transfer.

In this case, the size of the transfer is

$$B(0, g_S(A), b^*; b).$$

Case (ii)

In case (ii), the constraint is binding at the optimal policy.

The optimal policy is implicitly defined by the second and third conditions and the budget constraint

$$B(\tau, g, b'; b) = 0.$$

The tax rate is positive and the optimal level of public good reflects the fact that spending is financed by distortionary taxation.

The optimal level of borrowing reflects the fact that borrowing an additional unit means that we can reduce distortionary taxation.

When will we be in each case?

Case (i) arises when the triple $(0, g_S(A), b^*)$ satisfies the constraint that $B(0, g_S(A), b^*; b) > 0$.

Defining the function $A^*(b, b^*)$ from the equality

$$B(0, g_S(A), b^*; b) = 0,$$

· if $A < A^*(b, b^*)$, we are in case (i)

· if $A > A^*(b, b^*)$, we are in case (ii)

When $A > A^*(b, b^*)$, the tax rate-public good-public debt triple is such that $\tau > 0$, $b' > b^*$ and $g < g_S(A)$.

The debt level b^*

Further progress can be made by characterizing the debt level b^*.

To do this, we need to compute the derivative of the value function.

If $A > A^*(b, b^*)$

$$v(b', A) = \max_{\{\tau,g,z\}} \left\{ \begin{array}{c} u(w(1-\tau), g; A) + B(\tau, g, z; b') \\ +\delta E v(z, A') \\ : B(\tau, g, z; b') \geq 0 \;\; \& \;\; z \leq \bar{b} \end{array} \right\}$$

while if $A < A^*(b, b^*)$

$$\begin{aligned} v(b', A) &= u(w, g_S(A); A) + B(0, g_S(A), b^*; b') \\ &\quad +\delta E v(b^*, A'). \end{aligned}$$

Using the *Envelope Theorem,* we have that

$$\frac{\partial v(b', A)}{\partial b'} = \begin{cases} -\frac{1-\tau(b',A)}{1-\tau(b',A)(1+\varepsilon)}(1+\rho) & \text{if } A > A^*(b', b^*) \\ -(1+\rho) & \text{if } A < A^*(b', b^*) \end{cases}$$

where $\tau(b', A)$ is the optimal tax rate.

Thus, the expected marginal cost of debt is

$$-\delta E[\frac{\partial v(b',A)}{\partial b'}] = G(A^*(b', b^*)) + \int_{A^*(b',b^*)}^{\overline{A}} (\frac{1-\tau(b',A)}{1-\tau(b',A)(1+\varepsilon)})dG(A)$$

Combining this with our first order condition, the debt level b^* must satisfy

$$1 = G(A^*(b^*, b^*)) + \int_{A^*(b^*,b^*)}^{\overline{A}} \frac{1-\tau(b^*, A)}{1-\tau(b^*, A)(1+\varepsilon)} dG(A)$$

Since $\tau(b^*, A)$ exceeds 0 for all $A > A^*(b^*, b^*)$, this requires that $A^*(b^*, b^*) = \overline{A}$.

From the definition of $A^*(b^*, b^*)$, this in turn implies that

$$b^* = \underline{b} \equiv -pg_S(\overline{A})/\rho$$

At this debt level, the government's interest earnings on its bond holdings are always sufficient to finance the Samuelson level of public goods, implying that no taxation is necessary.

Summary

The optimal policy has the following form.

When the state (b, A) is such that $A < A^*(b, \underline{b})$, the tax rate is zero, the public good level is the Samuelson level and the debt level is $\underline{b}$.

Surplus revenues (positive if $A < A^*(b, \underline{b})$) are redistributed to citizens via a transfer.

When the state (b, A) is such that $A > A^*(b, \underline{b})$, the optimal policy involves positive levels of taxation, a public good level below the Samuelson level and a debt level that exceeds $\underline{b}$.

There are no surplus revenues and hence no transfer.

Dynamics

The optimal policies determine a distribution of public debt levels in each period.

In the long run, this sequence of debt distributions converges to the distribution that puts point mass on the debt level $\underline{b}$.

To understand this, first note that since $A^*(\underline{b}, \underline{b}) = \overline{A}$, it is clear that once the planner has accumulated a level of bonds equal to $-\underline{b}$, he will maintain it.

On the other hand, when the planner has bond holdings less than $-\underline{b}$ then he must anticipate using distortionary taxation in the future.

To smooth taxes he has an incentive to acquire additional bonds when the value of the public good is low in the current period.

This leads to an upward drift in government bond holdings over time.

Pulling all this together, we have the following proposition.

Asset Accumulation Proposition *The optimal policies converge to a steady state in which the debt level is* $\underline{b}$*, the tax rate is* 0*, the public good level is* $g_S(A)$*, and citizens receive* $\rho(-\underline{b}) - pg_S(A)$ *in transfers.*

Discussion

This result illustrates an interesting feature of optimal fiscal policy in a world with stochastic shocks: basically, the government may accumulate sufficient assets to be totally self-reliant.

Notice, however, that when we introduced stochastic shocks we did not change our assumptions about the structure of the bond market.

In particular, we are not allowing a market in *state-contingent* bonds.

Thus, we are in a world of *incomplete markets*.

With state-contingent debt available, the government optimally smooths taxes and relies on state contingent debt to deal with spending fluctuations (see Ljungvist and Sargent *Recursive Macroeconomic Theory* Chp 15 for an exposition).

Taxes are positive and constant just like in the case with perfect foresight.

However, it is not clear how realistic state contingent debt is.

Thus, when viewed as a positive model, the "prediction" embodied in the Proposition is not very appealing.

One way to avoid the absorbing state in which $b = \underline{b}$ is to assume that the government faces what Aiyagari et al. call "ad hoc" constraints on asset accumulation.

If the government is not allowed to accumulate more bonds than, say, $-z$ where $z \in (\underline{b}, 0)$, then even in the long run the optimal debt level will fluctuate and taxes will be positive at least some of the time.

This is because, by definition of $\underline{b}$, even when the government has accumulated $-z$ in bonds he can not finance the Samuelson level of public goods from the interest earnings when A is very high.

In these high realizations, it will be optimal to finance additional public good provision by a combination of levying taxes and reducing bond holdings.

Reducing bond holdings temporarily allows the government to smooth taxes.

The dynamic pattern of debt suggested by Barro is created by the rebuilding of bond holdings in future periods when A is low.

However, the difficulty with this resolution of the problem is obvious: why should the government be so constrained and, if it is, what should determine the level z?

3. The Optimal Taxation of Capital

The tax smoothing model just has labor income taxes.

We now discuss optimal taxation when the government can tax both labor and capital income.

The government has an exogenous revenue requirement in each period and can levy linear taxes on labor and capital income.

The government can commit to a path of labor and capital income taxes at time zero and the question is what is the optimal path of such taxes.

This problem is known as the *Ramsey Problem*, following Ramsey (1927).

The main issue is whether taxing capital income is a good idea.

Let's start with a simple model to review the issues concerning capital income taxation.

Consider an individual who lives for two periods, earns labor income y in period 1 and saves to consume in period 2.

Suppose the interest rate is r and that the government taxes labor income at rate t and income from savings at rate τ.

The individual's problem is

$$\max U(y(1-t)-s, s(1+r(1-\tau)))$$
$$s.t.\ s \in [0, y(1-t)]$$

or, equivalently,

$$\max U(C_1, C_2)$$
$$s.t.\ C_1 + \tfrac{1}{1+r(1-\tau)} C_2 \leq y(1-t)$$

The first order condition is

$$\frac{\partial U(C_1, C_2)/\partial C_1}{\partial U(C_1, C_2)/\partial C_2} = 1 + r(1 - \tau)$$

From this equation, we can see that taxing the income from savings distorts the inter-temporal allocation of consumption, by making future consumption more expensive.

Using second best logic, perhaps creating a distortion in the intertemporal allocation of consumption may be optimal given that we are already distorting labor supply by taxing income.

Thus, (assuming that he only works in period 1) the individual's problem is actually

$$\max U(C_1, C_2, l)$$
$$s.t.\ C_1 + \tfrac{1}{1+r(1-\tau)} C_2 \leq wl(1 - t)$$

and we have to worry about the distortion in l.

Interpreting C_1 and C_2 as different consumption goods, this looks analogous to the Corlett-Hague problem with two consumder goods.

Applying their result suggests that we should deviate from uniform taxation only when there is a difference in

More generally, the results from the Optimal mixed taxation literature suggest that if we can write utility as

$$U(C_1, C_2, l) = V(h(C_1, C_2), l),$$

it will not be optimal to distort the allocation of consumption across time by taxing the income from savings.

Nonetheless, this does not close the matter because this analysis does not take into account the dynamic equilibrium effects of saving on the capital stock and hence the rate of interest and the wage rate.

Thus, there is a need to embed the choice of taxes into a genuine dynamic general equilibrium model.

This was done in a famous paper by Chamley (1986) *Econometrica*.

Our exposition will follow Ljungvist and Sargent *Recursive Macroeconomic Theory* Chp 15 (See also the paper by Atkeson, Chari, and Kehoe on the reading list.)

The Model

Consider a production economy populated by a large number of identical, infinitely-lived, consumers.

Time periods are indexed by $t = 0, 1, ..., \infty$

Consumers enjoy consumption and lesiure, with preferences given by

$$\sum_{t=0}^{\infty} \beta^t u(c_t, l_t), \tag{1}$$

where β is the discount factor and u is increasing in c and l, strictly concave and satisfies the Inada conditions.

Each consumer is endowed with one unit of time that can be used for leisure l_t and labor n_t:

$$l_t + n_t = 1 \tag{2}$$

A constant returns to scale production technology transforms capital k_t and labor n_t into output via the production function $F(k_t, n_t)$.

The production function satisfies the standard Inada conditions.

Output can be used for private consumption c_t, government consumption g_t, and new capital k_{t+1}.

Government consumption is exogenously specified and constant so that $g_t = g$.

Feasibility requires that

$$c_t + g + k_{t+1} = F(k_t, n_t) + (1 - \delta)k_t, \qquad (3)$$

where δ is the depreciation rate on capital.

The Government

Government consumption is financed by proportional taxes on capital and labor income τ_t^k and τ_t^n.

The government can also trade one-period bonds. Let b_t be government indebtedness to the private sector (denominated in time t goods) at the beginning of period t.

The government's budget constraint in period t is

$$g = \tau_t^k r_t k_t + \tau_t^n w_t n_t + \frac{b_{t+1}}{R_t} - b_t \tag{4}$$

where r_t and w_t are the market determined rental rate of capital and the wage rate for labor (denominated in time t-goods) and R_t is the gross rate of return on one-period bonds held from t to $t+1$.

Interest earnings on bonds are tax exempt.

Consumers

The consumer chooses a sequence $(c_t, l_t, k_{t+1}, b_{t+1})_{t=0}^{\infty}$ to maximize (1) subject to the following sequence of budget constraints:

$$\begin{aligned} c_t + k_{t+1} + \tfrac{b_{t+1}}{R_t} = (1-\tau_t^n)w_t n_t + (1-\tau_t^k)r_t k_t \\ +(1-\delta)k_t + b_t \end{aligned} \tag{5}$$

and given initial holdings of capital k_0 and bonds b_0.

With $\beta^t \lambda_t$ as the Lagrange multiplier on the time t budget constraint, the first order conditions are:

$$u_c(t) = \lambda_t \tag{6}$$

$$u_l(t) = \lambda_t(1-\tau_t^n)w_t \tag{7}$$

$$\lambda_t = \beta\lambda_{t+1}[(1-\tau_{t+1}^k)r_{t+1} + 1 - \delta] \tag{8}$$

$$\lambda_t \frac{1}{R_t} = \beta\lambda_{t+1} \tag{9}$$

Substituting (6) into (7) and (8), we obtain

$$u_l(t) = u_c(t)(1 - \tau_t^n)w_t \tag{10}$$

and

$$u_c(t) = \beta u_c(t+1)[(1 - \tau_{t+1}^k)r_{t+1} + 1 - \delta]. \tag{11}$$

Moreover, equations (8) and (9) imply that

$$R_t = (1 - \tau_{t+1}^k)r_{t+1} + 1 - \delta. \tag{12}$$

This is a no-arbitrage condition ensuring that bonds and capital have the same rate of return.

Note that we can iterate on the consumer's budget constraints to write a single present-value constraint (see Ljungvist and Sargent)

$$\sum_{t=0}^{\infty}(\Pi_{i=0}^{t-1}\frac{1}{R_i})c_t = \sum_{t=0}^{\infty}(\Pi_{i=0}^{t-1}\frac{1}{R_i})(1 - \tau_t^n)n_t w_t + [(1 - \tau_0^k)r_0 + 1 - \delta]k_0 + b_0 \tag{13}$$

Firms

In each period, firms maximize profits

$$F(k_t, n_t) - w_t n_t - r_t k_t \tag{14}$$

Firms' first order conditions are

$$r_t = F_k(k_t, n_t) \tag{15}$$

and

$$w_t = F_n(k_t, n_t) \tag{16}$$

The Ramsey Problem

Let $c = (c_t)_{t=0}^{\infty}$, $l = (l_t)_{t=0}^{\infty}$, etc

Definition 1: A *feasible allocation* is a sequence (k, c, l, g) that satisfies (3) for all t

Definition 2: A *price system* is a 3-tuple of nonnegative bounded sequences (w, r, R)

Definition 3: A *government policy* is a 4-tuple of sequences (g, τ^k, τ^n, b)

Definition 4: A *competitive equilibrium* is a feasible allocation, a price system, and a government policy such that (a) given the price system and the government policy, the allocation solves both the firms' problem and the consumers' problem; and (b) given the allocation and the price system, the government policy satisfies the sequence of government budget constraints (4).

There are many competitive equilibria, indexed by different government policies. This multiplicity motivates the Ramsey problem.

Definition 5: Given k_0 and b_0, the *Ramsey problem* is to choose a competitive equilibrium that maximizes (1).

For reasons that will become clear, assume that the initial capital tax τ_0^k is fixed exogenously.

Zero Capital Income Tax

We can formulate the Ramsey problem as if the government chooses the after-tax rental rate of capital $\widetilde{r}_t = (1-\tau_t^k)r_t$ and the after-tax wage rate $\widetilde{w}_t = (1-\tau_t^n)r_t$.

Using (15), (16), we can express government revenues as

$$\begin{aligned}
\tau_t^k r_t k_t + \tau_t^n w_t n_t &= (r_t - \widetilde{r}_t)k_t + (w_t - \widetilde{w}_t)n_t \\
&= F_k(t)k_t + F_n(t)n_t - \widetilde{r}_t k_t - \widetilde{w}_t n_t \\
&= F(k_t, n_t) - \widetilde{r}_t k_t - \widetilde{w}_t n_t
\end{aligned}$$

Substituting this into (4), we can write the government budget constraint

$$g = F(k_t, n_t) - \widetilde{r}_t k_t - \widetilde{w}_t n_t + \frac{b_{t+1}}{R_t} - b_t$$

The Ramsey problem in Lagrangian form can be written as

$$
\begin{gathered}
L = \sum_{t=0}^{\infty} \beta^t \{ u(c_t, 1 - n_t) \\
+ \Psi_t [F(k_t, n_t) - \widetilde{r}_t k_t - \widetilde{w}_t n_t + \tfrac{b_{t+1}}{R_t} - b_t] \\
+ \theta_t [F(k_t, n_t) + (1 - \delta) k_t - c_t - g - k_{t+1}] \\
+ \mu_{1t} [u_l(t) - u_c(t) \widetilde{w}_t] \\
+ \mu_{2t} [u_c(t) - \beta u_c(t+1)(\widetilde{r}_{t+1} + 1 - \delta)] \}
\end{gathered}
\tag{17}
$$

where $R_t = \widetilde{r}_{t+1} + 1 - \delta$ as given by (12).

The choice variables are $\{(k_{t+1}, c_t, n_t, b_{t+1}, \widetilde{r}_t, \widetilde{w}_t)\}_{t=0}^{\infty}$.

Note that the household's budget constraint is not explicitly included because it is redundant when the government satisfies its budget constraint and the resource constraint holds.

The first order condition with respect to k_{t+1} is

$$
\begin{gathered}
\theta_t = \beta \{ \Psi_{t+1} [F_k(t+1) - \widetilde{r}_{t+1}] + \\
\theta_{t+1} [F_k(t+1) + 1 - \delta] \}
\end{gathered}
\tag{18}
$$

The term θ_t on the left hand side measures the marginal cost of investing more in capital at time t.

The right hand side measures the marginal benefit: the first term the increase in tax revenues and the second term the increase in output.

We can now establish:

Proposition 1 *If the solution to the Ramsey problem converges to a steady state, then in the steady state, the tax rate on capital income is zero.*

Proof: In a steady state, all endogenous variables remain constant. Using (15), the steady state version of (18) is

$$\theta = \beta\{\Psi[r - \widetilde{r}] + \theta[r + 1 - \delta]\} \tag{19}$$

The steady state version of (11) is

$$1 = \beta(\widetilde{r} + 1 - \delta). \tag{20}$$

Substituting (20) into (19) yields

$$(\theta + \Psi)(r - \widetilde{r}) = 0 \tag{21}$$

■

Intuition: taxing capital income in period $t+1$ is equivalent to taxing consumption at a higher rate in period $t+1$ than in period t.

Thus, a positive tax on capital income is equivalent to an ever-increasing tax on consumption.

Note here that we do not prove that the solution to the Ramsey problem necessarily converges to a steady state.

Primal Approach to the Ramsey Problem

In the Lagrangian formulation in (17), we reduced a pair of tax rates (τ_t^k, τ_t^n) and a pair of prices (r_t, w_t) to just one pair of numbers $(\widetilde{r}_t, \widetilde{w}_t)$ by utilizing the firms' first order conditions and equilibrium outcomes in factor markets.

In fact, we can eliminate all prices and taxes and think of the government directly choosing a feasible allocation, subject to constraints that ensure the existence of prices and taxes such that the chosen allocation is consistent with agents' optimization.

This is known as the *Primal approach* and is the standard way of solving this type of problem.

We begin with the consumer's budget constraint in present value terms (13)

$$\sum_{t=0}^{\infty} q_t^0 c_t = \sum_{t=0}^{\infty} q_t^0 (1 - \tau_t^n) w_t n_t + [(1 - \tau_0^k) r_0 + 1 - \delta] k_0 + b_0 \tag{22}$$

where for $t \geq 1$

$$q_t^0 = (\Pi_{i=0}^{t-1} \frac{1}{R_i}) \tag{23}$$

is the period 0 price of consumption in period t. We set $q_0^0 = 1$.

We then use the consumer's first order conditions to replace the prices q_t^0 and $(1-\tau_t^n)w_t$ with the consumer's marginal rates of substitution.

A stepwise summary of the primal approach is as follows:

STEP 1. Obtain the first order conditions of the consumer and firm problems, as well as any arbitrage pricing conditions. Solve these conditions for $\{q_t^0, r_t, w_t, \tau_t^k, \tau_t^n\}_{t=0}^{\infty}$ as functions of the allocation $\{(c_t, n_t, k_{t+1})\}_{t=0}^{\infty}$.

STEP 2. Substitute these expressions for taxes and prices into the consumer's present value budget constraint. This gives an intertemporal constraint involving only the allocation.

STEP 3. Solve for the Ramsey allocation by maximizing (1) subject to (3) and the "implementability condition" derived in Step 2.

STEP 4. After the Ramsey allocation is solved for, use the formulas from step 1 to find taxes and prices.

We now carry out these steps.

STEP 1

Letting λ be the Lagrange multiplier on the consumer's present value budget constraint (22), the first order conditions for the consumer's problem of maximizing (1) subject to (22) are given by:

$$\beta^t u_c(t) = \lambda q_t^0 \tag{24}$$

and

$$\beta^t u_l(t) = \lambda q_t^0 (1 - \tau_t^n) w_t \tag{25}$$

These imply that

$$q_t^0 = \beta^t \frac{u_c(t)}{u_c(0)} \tag{26}$$

and

$$(1 - \tau_t^n) w_t = \frac{u_l(t)}{u_c(t)} \tag{27}$$

The arbitrage condition (12) implies that

$$\frac{q_t^0}{q_{t+1}^0} = (1 - \tau_{t+1}^k) r_{t+1} + 1 - \delta \tag{28}$$

Firms first order conditions and factor market equilibrium imply (15) and (16)

STEP 2

Substitute equations (26), (27) and $r_0 = F_k(0)$ into (22), so that we can write the consumer's present value budget constraint as

$$\sum_{t=0}^{\infty} \beta^t [u_c(t) c_t - u_l(t) n_t] - A = 0 \tag{29}$$

where

$$\begin{aligned} A &= A(c_0, n_0, \tau_0^k) \\ &= u_c(0)\{[(1-\tau_0^k)F_k(0) + 1 - \delta]k_0 + b_0\} \end{aligned} \tag{30}$$

STEP 3

The Ramsey problem is to maximize (1) subject to (29) and the sequence of feasibility constraints (3).

Let Φ be the Lagrange multiplier on (29) and define

$$V(c_t, n_t, \Phi) = u(c_t, 1 - n_t) + \Phi(u_c(t)c_t - u_l(t)n_t). \tag{31}$$

Form the Lagrangian

$$\begin{aligned} L = \sum_{t=0}^{\infty} \beta^t \{ & V(c_t, n_t, \Phi) \\ & + \theta_t [F(k_t, n_t) + (1-\delta)k_t - c_t - g - k_{t+1}]\} \\ & - \Phi A(c_0, n_0, \tau_0^k), \end{aligned} \tag{32}$$

where $\{(\theta_t)\}_{t=0}^{\infty}$ are a sequence of Lagrange multipliers on the feasibility constraints.

The choice variables are $\{(c_t, n_t, k_{t+1})\}_{t=0}^{\infty}$

The first order conditions for this problem imply that for all $t \geq 1$

$$V_c(t) = \beta V_c(t+1)(1 - \delta + F_k(t+1)), \tag{33}$$

and

$$V_n(t) = -V_c(t) F_n(t). \tag{34}$$

In addition,

$$V_c(0) - \Phi A_c = \beta V_c(1)(1 - \delta + F_k(1)) \tag{35}$$

and

$$V_n(0) = [\Phi A_c - V_c(0)] F_n(0) + \Phi A_n \tag{36}$$

We seek an allocation $\{(c_t, n_t, k_{t+1})\}_{t=0}^{\infty}$ and a multiplier Φ that satisfies the system of difference equations formed by these four equations, (29) and the sequence of feasibility constraints (3).

STEP 4

After such an allocation has been found obtain q_t^0 from equation (26), r_t from equation (15), w_t from equation (16), τ_t^n from equation (27), and τ_t^k from equation (28).

Results from Primal Approach

Observe that if the term

$$\frac{V_c(t)}{u_c(t)} = \frac{u_c(t) + \Phi(u_{cc}(t)c_t + u_c(t) - u_{lc}(t)n_t).}{u_c(t)}$$

has the same value in periods t and $t+1$ (for $t \geq 1$), then the capital income tax in period $t+1$ is zero.

To see this note that if

$$\frac{V_c(t)}{u_c(t)} = \frac{V_c(t+1)}{u_c(t+1)}$$

then (33) can be written as

$$u_c(t) = \beta u_c(t+1)(1 - \delta + F_k(t+1)).$$

From (26) we have that

$$\frac{q_t^0}{q_{t+1}^0} = \frac{u_c(t)}{\beta u_c(t+1)}$$
$$= (1 - \delta + F_k(t+1))$$

Thus, from (28), we have that

$$(1-\tau_{t+1}^{k})r_{t+1} = F_k(t+1)$$

which implies that $\tau_{t+1}^{k} = 0$.

We can therefore establish Proposition 1 by simply noting that if the solution to the Ramsey problem converges to a steady state then

$$\frac{V_c(t)}{u_c(t)} = \frac{V_c(t+1)}{u_c(t+1)}.$$

We can also establish a stronger result. Consider the class of utility functions

$$U(c,l) = \frac{c^{1-\sigma}}{1-\sigma} + V(l) \tag{37}$$

where $\sigma \leq 1$.

Proposition 2: *For utility functions of the form (37), the optimal income tax in period t is zero for all $t \geq 2$.*

Proof: For a utility function of the form (37) it is the case that

$$\frac{V_c(t)}{u_c(t)} = 1 + \Phi(1 - \sigma)$$

■

The zero capital income tax result is also robust to a number of extensions. See Atkeson, Chari and Kehoe for discussion.

The period 0 capital income tax rate

We assumed earlier that the period 0 capital income tax rate was fixed.

What would happen if this assumption were not satisfied?

Note that

$$\frac{\partial L}{\partial \tau_0^k} = -\Phi A_\tau = \Phi u_c(0) F_k(0) k_0,$$

which is strictly positive as long as $\Phi > 0$.

The condition implies that, if there were no restraint, the optimal capital income tax in period 0 should be large enough so that $\Phi = 0$.

When $\Phi = 0$, the solution to the Ramsey problem is first best efficient implying that the government is imposing no distortionary taxation.

Thus, if there is no restriction on the period 0 capital income tax rate, the government should raise all revenues through a time 0 capital levy, then lend the proceeds to the private sector and finance government spending from the interest earnings.

All future taxes would be set equal to zero.

This is why it is necessary to impose the requirement that the period capital income tax rate is fixed.

The problem of optimal taxation just considered has been extended in a variety of ways, notably by incorporating productivity and preference shocks.

See Chari and Kehoe's survey article for more information.

4. The Problem of Time Inconsistency

Suppose that the government determines the solution to the Ramsey problem at time 0.

Suppose that at some future period t, the government asks whether it is wise to continue with the optimal solution.

To reevaluate, they resolve the Ramsey problem but from period t onwards.

If the new solution coincides with the original solution for all periods t, then the solution is *time consistent*.

If not, then the original solution is *time inconsistent*.

The solution to the Ramsey problem just analyzed is time inconsistent.

The government would like to increase the capital income tax at time t because capital is sunk at time t.

In this way, it could finance government spending with no deadweight loss.

Kydland and Prescott (1977) pointed out that the problem of time inconsistency arises in many models of policy-making.

There are numerous examples; for example, the government would like to commit ex ante not to pay terrorists to release hostages, but once hostages have been captured, it would change it mind.

When optimal policies are time inconsistent, it is not clear why the government would not change policies when the future date arises.

Since governments do not have any obvious way to commit themselves, it is not clear that the lessons of dynamic optimal taxation have any predictive content.

Moreover, if governments can in fact not commit to following a pre announced policy plan, then assuming they can is perhaps not the most sensible approach to providing policy advice.

This raises the question of what will dynamic taxation look like when governments cannot commit?

This is a complex problem as it gives rise to a repeated game between government and the citizens.

The government must in each period be choosing policy optimally given the state of the economy (i.e., the level of capital) and the citizens when choosing their savings must anticipate how government will choose future policy.

While this is closer to a question in political economy than to a normative question, sometimes thinking through the implications of lack of commitment can help identify good policies.

There are two basic approaches to modelling policy choice when governments cannot commit.

1. *Reputational Approach* - this explores the possibility that reputation can substitute for a commitment technology when governments choose sequentially.

The basic idea is that each period the government makes policy choices whose consequences include a current period return and a reputation to pass on to the next period.

Under rational expectations, any reputation that the government carries into next period must be one that it will want to confirm.

The literature adapts ideas from the literature on repeated games so that they can be applied to contexts in which a single agent (the government) behaves strategically and the remaining agents behavior (i.e., the citizens) can be summarized as a competitive equilibrium that responds nonstrategically to government's choices.

2. *Markov Approach* - this approach focuses on the Markov-perfect equilibria of the game between government and its citizens; i.e., government's choices in any period can only depend on the state of the economy in that period.

We will illustrate the Reputational Approach by going through the paper of Chari "Time Consistency and Optimal Policy Design" and the Markov approach by going through a paper by Per Krusell entitled "Time-Consistent Redistribution".

The purpose of both these papers is expositional.

The models are just designed to be the simplest possible models to think through the issues.

Thus, their purpose is simply to illustrate a method rather than to derive any significant normative lessons or make any positive predictions.

Chari's Model

Static Version

The economy consists of a large number of identical consumers and a government.

Consumers make decisions at two distinct points in time, the first and second stage.

At the first stage, consumers are endowed with w units of a consumption good from which they consume c_1 units and store k units.

If they store k units, they get back Rk units in the second stage.

In addition, consumers can work at the second stage.

If they work l units, they obtain l units of output.

The government requires G units of output in the second stage.

It raises revenues from taxes on capital and labor.

The capital tax is δ and the labor tax is τ.

Thus, a consumer's second stage income is

$$Rk(1-\delta) + l(1-\tau).$$

Each consumer's utility is $U(c_1 + c_2, l)$.

The consumer's problem is

$$\begin{gathered} \max_{(c_1, c_2, l, k)} U(c_1 + c_2, l) \\ s.t. \quad c_1 + k = w \\ c_2 = Rk(1-\delta) + l(1-\tau) \end{gathered}$$

Note that, since c_1 and c_2 are perfect substitutes, the solution is for the consumer to set $k = w$ if $R(1-\delta) \geq 1$ and $k = 0$ if $R(1-\delta) < 1$.

The government's budget constraint is

$$G = \delta Rk + \tau l.$$

Policy-making with commitment

With commitment, the government can choose policies $\pi = (\delta, \tau)$ before consumers make their first stage choices and stick with these policies.

Consumers choose their optimal plan (c_1, c_2, l, k) given π.

Let $(c_1(\pi), c_2(\pi), l(\pi), k(\pi))$ denote this optimal plan.

The government then solves

$$\begin{array}{c} \max_\pi U(c_1(\pi) + c_2(\pi), l(\pi)) \\ s.t. \quad G = \delta R k(\pi) + \tau l(\pi). \end{array}$$

This corresponds to the Ramsey problem that we have just discussed.

The optimal policy is very easy to describe.

The government sets δ so that $R(1-\delta) = 1$ and then determines τ from the constraint

$$G = \delta R w + \tau l(\delta, \tau).$$

Let $\pi^o = (\delta^o, \tau^o)$ denote the optimal policies with commitment.

Policy-making without commitment

If the government cannot commit, then it is effectively choosing the policies *after* the consumers have made their first stage choices.

It is clear that after the consumers have made their first stage choices, the optimal choice of capital tax for the government is 1 if $k > 0$.

A *sustainable equilibrium* consists of a first stage decision by consumers, a government policy, and a second stage consumer decision function such that i) consumers' decisions are optimal in each stage and ii) the government policy maximizes the government's objective function.

What does a sustainable equilibrium look like?

In a sustainable equilibrium, the first stage decision of consumers is $(c_1, k) = (w, 0)$;

The second stage consumer decision function is

$$\begin{array}{c} l(\delta,\tau) = \arg\max_l U(w + l(1-\tau), l) \\ c_2(\delta,\tau) = l(\tau)(1-\tau) \end{array};$$

and the government policy is $(\delta, \tau) = (1, \tau^*)$ where

$$\tau^* l(\tau^*) = G.$$

Thus, consumers save nothing, the government taxes capital at 100% and all revenues are raised by the labor tax.

Note that the government's utility is strictly higher with commitment - the reason is that it could have chosen the policy $(\delta, \tau) = (1, \tau^*)$ but it did not; that is, $(\delta^o, \tau^o) \neq (1, \tau^*)$.

The difficulty arises not because the government is not benevolent - it is at all times maximizing consumer welfare.

Rather the difficulty is that the government cannot commit not to change its policies when it sees an opportunity to raise consumer welfare.

Infinite period version

Imagine that this simple game was repeated for an infinite number of periods.

Assume no commitment so that, in each period t, consumers make their first stage decisions, then the government sets current tax rates, and then consumers make their second stage decisions.

How is a sustainable equilibrium defined in such a world?

Let H_{t-1} be the history of all government policies up to an including time $t-1$; that is,

$$H_{t-1} = (\pi_s)_{s=0}^{t-1}.$$

Consider the problem of a consumer at the first stage of some period t.

The consumer must choose a first stage allocation $(c_{1t}(H_{t-1}), k_t(H_{t-1}))$ and a contingency plan for setting future actions for all possible future histories.

After the first stage consumer decisions have been made, the government, faced with history H_{t-1} sets time t tax rates $\pi_t(H_{t-1})$ and chooses a contingency plan for setting all future tax rates for all possible future histories.

At the second stage of period t, the history is now $H_t = (H_{t-1}, \pi_t(H_{t-1}))$ and consumers choose a second stage allocation $(c_{2t}(H_t), l_t(H_t))$ and a contingency plan for setting all future actions for all future possible histories.

More formally, in the first stage of period t the consumer's problem is to choose a *contingency plan* defined to be a sequence of allocation rules

$$(c_{1s}(H_{s-1}), k_t(H_{s-1}), c_{2s}(H_s), l_s(H_s))_{s=t}^{\infty},$$

to solve

$$\begin{array}{c} \max \sum_{s=t}^{\infty} \beta^s U(c_{1s} + c_{2s}, l_s) \\ s.t. \quad c_{1s} + k_s = w \, \forall s \\ c_{2s} = Rk_s(1 - \delta_s(H_{s-1})) + l_s(1 - \tau_s(H_{s-1})) \, \forall s \end{array} \tag{1}$$

The histories in (1) are induced by the government's policy rules $(\pi_s(H_{s-1}))_{s=t}^{\infty}$.

For example, at the first stage of period t, consumers believe that the history the government confronts in period $t+1$ will be given by $(H_{t-1}, \pi_t(H_{t-1}))$.

The period $t+1$ policy will then be given by

$$\pi_{t+1}(H_{t-1}, \pi_t(H_{t-1})),$$

etc.

After the first stage consumer decisions have been made in period t, the governments' problem is to choose a *policy plan* defined to be a sequence of policy rules

$$(\pi_s(H_{s-1}))_{s=t}^{\infty}$$

to maximize

$$\begin{aligned} \max \sum_{s=t}^{\infty} \beta^s U(c_{1s}(H_{s-1}) + c_{2s}(H_s), l_s(H_s)) \\ s.t. \quad g \leq \delta_s(H_{s-1}) R k_s(H_{s-1}) + \tau_s(H_{s-1}) l_s(H_s) \end{aligned} \tag{2}$$

where the histories are induced from H_{t-1} by the chosen policies.

Sustainable Equilibrium

A *sustainable equilibrium* is a policy plan for the government and contingency plans for consumers such that for every history, the following conditions are met:

i) consumers' contingency plans solve (1) given the policy plan.

ii) given consumer's contingency plans, the government's policy plan solves (2).

Note that consumers take the evolution of future histories as unaffected by their actions and, in this sense, behave competitively.

The government recognizes the effect of its policies on the histories and thus on the decisions of consumers and, in this sense, does not behave competitively.

Results

The set of sustainable outcomes are characterized in the paper "Sustainable Plans" by Chari and Kehoe in *JPE* 1990.

The worst sustainable equilibrium is an infinite repetition of the static sustainable equilibrium.

With sufficiently little discounting (high β), the Ramsey policies are sustainable.

The plans supporting such outcomes specify that the government should follow Ramsey policies as long as these policies have been followed in the past.

Consumers' contingency plans specify that for such histories, they should save their entire endowments.

If the government has ever deviated from the Ramsey policies, consumers' plans specify that they save nothing.

Given such plans, the government chooses optimally to continue the Ramsey policies in each period.

Krusell's Model

There are an infinite number of periods indexed by t

There are two classes of citizen: workers and capitalists and firms with CRS production technologies.

Workers supply labor and capitalists supply capital.

The government taxes production and uses the revenues to finance a transfer to workers.

Workers

Workers supply 1 unit of labor inelastically in each period.

They consume their wage w_t and their government transfer T_t - they do not save.

Their lifetime utility is

$$\sum_{t=0}^{\infty} \beta^t u(c_{wt})$$

where

$$c_{wt} = w_t + T_t.$$

Capitalists

Capitalists save and consume but do not work.

They choose $\{c_{ct}, k_{t+1}\}_{t=0}^{\infty}$ to maximize

$$\sum_{t=0}^{\infty} \beta^t v(c_{ct})$$
$$s.t.\ c_{ct} + k_{t+1} = (1 - \delta + r_t)k_t$$

where k_0 is given.

r_t is the rental rate on capital and δ the depreciation rate.

u and v have all the usual properties and $\beta \in (0, 1)$.

Firms

Firms choose (k_t, l_t) to solve

$$(1 - \tau_t)F(k_t, l_t) - w_t l_t - r_t k_t$$

where F has all the usual properties and τ_t is the tax rate on production.

Prices and Constraints

The feasibility constraint in period t is

$$c_{wt} + c_{ct} + k_{t+1} = F(k_t, 1) + (1 - \delta)k_t.$$

The government's budget constraint is

$$\tau_t F(k_t, 1) = T_t.$$

Factor prices are

$$w_t = (1 - \tau_t) F_l(k_t, 1)$$

and

$$r_t = (1 - \tau_t) F_k(k_t, 1)$$

Note that

$$\begin{aligned} c_{wt} &= C_w(k_t, \tau_t) \\ &= F_l(k_t, 1) + \tau_t F_k(k_t, 1) k_t \end{aligned}$$

and that

$$\begin{aligned} c_{ct} &= C_c(k_t, \tau_t, k_{t+1}) \\ &= (1 - \tau_t) F_k(k_t, 1) k_t - k_{t+1} \end{aligned}$$

The Ramsey Problem

The objective of the government is

$$\sum_{t=0}^{\infty} \beta^t [\lambda v(c_{ct}) + (1-\lambda) u(c_{wt})]$$

When the government can commit to future policy, it chooses $\{\tau_t\}_{t=0}^{\infty}$ to maximize this objective subject to $\{c_{wt}\}_{t=0}^{\infty}$ and $\{c_{ct}\}_{t=0}^{\infty}$ being a competitive equilibrium.

We can rewrite it as choosing $\{k_{t+1}, \tau_t\}_{t=0}^{\infty}$ to solve

$$\max \sum_{t=0}^{\infty} \beta^t [\lambda v(C_c(\cdot)) + (1-\lambda) u(C_w(\cdot))]$$
$$v'(C_c(t)) = \beta v'(C_c(t+1))[1 - \delta + (1-\tau_{t+1}) F_k(k_{t+1}, 1)]$$

where $C_c(t) = C_c(k_t, \tau_t, k_{t+1})$, etc.

The constraint comes from the first order condition for the capitalist's problem.

The solution will not be time consistent for familiar reasons.

The Case of No Commitment

Suppose that the government seeks to maximize its objective function in each period.

We look for an equilibrium in which the current government sets the current tax correctly forecasting how future governments will set the tax.

We restrict attention to Markov perfect equilibria where the tax rate selected at time t just depends on the capital stock in period t.

This is summarized by a function $\tau_t = \Psi(k_t)$ - thus the function Ψ is the key endogenous variable.

Attention is restricted to differentiable functions $\Psi(\cdot)$

To characterize the function $\Psi(\cdot)$, we study the optimal choice of τ by a government that understands that future governments will choose according to Ψ.

The government has to anticipate how the private sector will respond - this is described by a function $k' = H(k, \tau)$

The function $H(k, \tau)$ describes next period's capital stock if the current period's tax is τ, the current period's capital stock is k and future governments set the tax according to the rule $\Psi(\cdot)$.

To solve for $H(k, \tau)$ we use the first order condition for the capitalists' optimal investment decision

$$v'(C_c(k, \tau, H)) = \beta v'(C_c(H, \Psi(H), H(H, \Psi(H)))) \\ \times[1 - \delta + (1 - \Psi(H))F_k(H, 1)] \tag{I}$$

This equation has to hold for all (k, τ).

Given the private sector's behavior and given an initial capital level k, the government solves

$$\max_{(\tau,k')} \lambda v(C_c(k, \tau, k')) + (1 - \lambda)u(C_w(k, \tau)) \\ +\beta V(k')$$

subject to

$$k' = H(k, \tau)$$

where the value function V is defined recursively by

$$V(k) = \lambda v(C_c(k, \Psi(k), H(k, \Psi(k))) \\ +(1-\lambda)u(C_w(k, \Psi(k))) + \beta V(H(k, \Psi(k)))$$

In equilibrium, the solution to the government's problem must be $\Psi(k)$.

We can thus describe the government's equilibrium behavior as solving a dynamic programming equation

$$\max_{(\tau,k')} \lambda v(C_c(k, \tau, k')) \\ +(1-\lambda)u(C_w(k, \tau)) + \beta V(k') \\ s.t.\ k' = H(k, \tau) \qquad \text{(II)}$$

More formally, a *Markov perfect equilibrium* consists of a collection of functions Ψ, H, and V such that H solves (I), V solves (II), and Ψ attains the maximum in (II).

Characterizing the equilibrium

Consider the sequential problem of choosing $\{\tau_t, k_{t+1}\}_{t=0}^{\infty}$ to solve

$$\max \sum_{t=0}^{\infty} \beta^t [\lambda v(C_c(k_t, \tau_t, k_{t+1}))$$
$$+(1-\lambda)u(C_w(k_t, \tau_t))]$$
$$s.t.\ k_{t+1} = H(k_t, \tau_t)$$

Note that this problem assumes that the current government can choose all future tax rates - which in fact it cannot.

However, despite this its solution will coincide with that of the dynamic-programming problem. (This is an application of *Bellman's principle*.)

Using the first order conditions for τ_t, τ_{t+1} and k_{t+1} yields the following condition:

$$\begin{gathered} R_\tau + R_{k'} H_\tau \\ +\beta\{R'_k - \tfrac{H'_k}{H'_\tau} R'_\tau\} = 0 \end{gathered} \tag{III}$$

where $'$s denote next period values and

$$R(k, \tau, k') = \lambda v(C_c(k, \tau, k')) \\ +(1-\lambda)u(C_w(k, \tau))$$

Moreover, in (III) this period's functions are evaluated at $(k_t, \tau_t) = (k, \Psi(k))$ and next period's functions are evaluated at $(k_{t+1}, \tau_{t+1}) = (H(k, \Psi(k)), \Psi(H(k, \Psi(k))))$

Equation (III) is a functional equation that must hold for all k. It is known as the "Generalized Euler Equation" (GEE).

A *time-consistent equilibrium* is then a pair of functions H and Ψ satisfying (I) and (III).

Interpretation of GEE

We can think of the GEE as a "variation": given values of k and k'', τ and τ' are varied in the best possible way.

If we increase τ, there will be an increase in C_w - which provides a direct boost to the worker's utility.

The capitalist is left with less resources, and reduces current consumption and savings (i.e., k').

The fall in k' has two effects: i) reduces future consumption of workers and capitalists (C'_w and C'_c)

ii) reduces τ' in order to keep k'' constant - which implies lower consumption of workers and more for capitalists.

Example

Assume that

$$u(c) = v(c) = \frac{c^{1-\sigma} - 1}{1 - \sigma}$$

for $\sigma \neq 1$.

Further assume that

$$F(k, l) = k^{\alpha} l^{1-\alpha}$$

and that

$$(\delta, \lambda, \beta, \alpha) = (1, 0, 0.95, 0.3).$$

Then, in the Ramsey problem the steady state (i.e., $(k_t, \tau_t) = (k, \tau)$) involves $\tau = 0$.

This may seem a surprising result but it is an extension of the zero capital income tax result.

In the no-commitment case, results depend on σ.

If $\sigma = 1.5$, steady state $\tau = 0.03$, while if $\sigma = 0.5$, steady state $\tau = 0.09$.

Thus, without commitment taxes will be positive in the long run leading to lower long run output.

5. New Dynamic Public Finance

Most literature on dynamic optimal taxation assumes that the government is restricted to use *linear* taxes on *current* variables like capital and labor income.

The weakness of this approach is that there is no explicit motivation for the restrictions that drive the analysis.

Why should the government be restricted to linear taxes?

Why should the government be restricted to using taxes that are functions of current variables as opposed to assets or the history of labor earnings?

This weakness has lead to a new literature about dynamic optimal taxation, called the "New Dynamic Public Finance".

Under the new approach, instead of specifying an arbitrary set of tax instruments, the analyst specifies the informational and/or enforcement frictions that limit the government's ability to tax.

Then the analyst designs a tax system that implements second best efficient allocations given these frictions.

Thus, it is like the Mirrlees approach to static optimal income taxation but applied to dynamic problems.

Indeed, this new literature is basically a dynamic extension of the Mirrlees model; that is, individuals have different earnings generating abilities which are unobserved by the government and these abilities change over time.

These changing abilities create a need for insurance as well as redistribution.

To provide an introduction to this material, we will go through the survey article "New Dynamic Public Finance: A User's Guide" by Golosov, Tsyvinski and Werning.

Model

The economy lasts for 2 periods, indexed by $t \in \{1, 2\}$

There are a continuum of individuals.

There are two goods - consumption and leisure.

In period t individuals get utility from consumption x_t and work l_t according to the utility function

$$u(x_t) - \varphi(l_t)$$

where u is increasing, strictly concave, and twice continuously differentiable and φ is increasing, strictly convex, and twice continuously differentiable.

Note that it is important to have risk aversion to capture the insurance role for policy.

Individuals have a discount rate β.

Individuals are endowed with $\bar{l}$ units of time in each period. Assume that $\varphi'(0) = 0$ and that $\lim_{l \to \bar{l}} \varphi'(l) = \infty$.

In each period individuals differ in their income generating abilities.

In period t an individual with income generating ability a earns income $y_t = al_t$ if he works an amount l_t.

There are n ability types in each period indexed by $i = 1, ..., n$.

Ability type i has ability a_i.

Let $\pi_1(i)$ denote the probability that an individual has ability type i in period 1.

This will also equal the fraction of ability type i in period 1.

Let $\pi_2(j|\, i)$ denote the probability that an individual who has ability type i in period 1 has ability type j in period 2.

There is a government which has an exogenous revenue requirement of G each period.

Individuals and the government can borrow or lend at the exogenously fixed interest rate R.

Let $(x_1(i), y_1(i))$ denote period 1 consumption and earnings of a type i individual.

Let $(x_2(j|\, i), y_2(j|\, i))$ denote period 2 consumption and earnings of a type j individual who was type i in period 1.

The resource constraint in period 1 is

$$\sum_i x_1(i)\pi_1(i) + G + S = \sum_i y_1(i)\pi_1(i)$$

where S denotes aggregate savings.

The resource constraint in period 2 is

$$\sum_i \sum_j x_2(j|\,i)\pi_2(j|\,i)\pi_1(i) + G = \sum_i \sum_j y_2(j|\,i)\pi_2(j|\,i)\pi_1(i) + (1+R)S$$

By solving out for S, we can collapse these down to one single present value resource constraint

$$\sum_i \left[x_1(i) + \frac{1}{1+R} \sum_j x_2(j|\,i)\pi_2(j|\,i) \right] \pi_1(i) + G + \frac{G}{1+R} = \sum_i \left[y_1(i) + \frac{1}{1+R} \sum_j y_2(j|\,i)\pi_2(j|\,i) \right] \pi_1(i)$$

Social welfare is defined as

$$\sum_i \left[\begin{array}{c} u(x_1(i)) - \varphi(\frac{y_1(i)}{a_i}) \\ +\beta \sum_j \left(u(x_2(j|\, i)) - \varphi(\frac{y_2(j|i)}{a_j}) \right) \pi_2(j|i) \end{array} \right] \pi_1(i)$$

This is a Utilitarian social welfare function.

The First Best

The first best allocation solves the problem of maximizing social welfare subject to the present value resource constraint.

The first best allocation satisfies the following first order conditions for all i.

$$\varphi'(\frac{y_1(i)}{a_i}) = a_i u'(x_1(i))$$

$$\varphi'(\frac{y_2(j|\,i)}{a_j}) = a_j u'(x_2(j|\,i)) \quad \text{for all } j$$

and

$$u'(x_1(i)) = \beta(1+R)\sum_j u'(x_2(j|\,i))\pi_2(j|\,i)$$

The first two conditions are the familiar consumption-leisure trade off from the static optimal income tax model.

The second is the intertemporal consumption smoothing condition.

When $\beta = 1/(1+R)$, this condition implies that $x_1(i) = x_2(j|\,i)$ for all j which implies that consumption is smoothed across time and across states.

The Second Best

Suppose that government is unable to observe individuals' ability types or labor supplies.

Assume it can observe their incomes and savings.

The government's problem can be formulated as a mechanism design problem.

The individuals make ability reports i_r and j_r to the government in the first and second period.

The government provides bundles $(x_1(i_r), y_1(i_r))$ and $(x_2(j_r|\, i_r), y_2(j_r|\, i_r))$.

For each ability type i a *reporting strategy* is a choice of a first period report i_r and a plan for the choice of a second period report $j_r(j)$.

Each ability type must have an incentive to tell the truth, so that it must be the case that for each i

$$
\begin{aligned}
& u(x_1(i)) - \varphi(\frac{y_1(i)}{a_i}) \\
& + \beta \sum_j \left[u(x_2(j|i)) - \varphi(\frac{y_2(j|i)}{a_j}) \right] \pi_2(j|i) \\
\geq \quad & u(x_1(i_r)) - \varphi(\frac{y_1(i_r)}{a_i}) \\
& + \beta \sum_j \left[u(x_2(j_r(j)|i_r)) - \varphi(\frac{y_2(j_r(j)|i_r)}{a_j}) \right] \pi_2(j|i)
\end{aligned}
$$

for all alternative reporting strategies i_r and $j_r(j)$.

The second best allocation maximizes social welfare subject to the present value resource constraint and the incentive constraints.

Results

i. Intertemporal distortions

Consider first the case with constant types; i.e., $\pi_2(i|i) = 1$.

In this case, the problem simplifies to

$$\max \sum_i \left[\begin{array}{c} u(x_1(i)) - \varphi(\frac{y_1(i)}{a_i}) \\ +\beta\left(u(x_2(i)) - \varphi(\frac{y_2(i)}{a_j})\right) \end{array} \right] \pi_1(i)$$

subject to

$$\begin{aligned} & \sum_i \left[x_1(i) + \frac{1}{1+R} x_2(i) \right] \pi_1(i) \\ & +G + \frac{G}{1+R} \\ = & \sum_i \left[y_1(i) + \frac{1}{1+R} y_2(i) \right] \pi_1(i) \end{aligned}$$

and for each $i = 1, ..., n$

$$\begin{aligned} & u(x_1(i)) - \varphi(\frac{y_1(i)}{a_i}) \\ & +\beta \left[u(x_2(i)) - \varphi(\frac{y_2(i)}{a_i}) \right] \\ \geq \quad & u(x_1(i_r)) - \varphi(\frac{y_1(i_r)}{a_i}) \\ & +\beta \left[u(x_2(i_r)) - \varphi(\frac{y_2(i_r)}{a_i}) \right] \end{aligned}$$

for all $i_r = 1, ..., n$.

Proposition 1 *With constant types, a second best allocation satisfies for all* $i = 1, ..., n$

$$u'(x_1(i)) = \beta(1 + R)u'(x_2(i))$$

Proof: To see this, note that only the total utility from consumption $u(x_1) + \beta u(x_2)$ enters the objective function and incentive constraints.

It follows that for any total utility coming from consumption $u(x_1(i)) + \beta u(x_2(i))$, it must be that resources $x_1(i) + \frac{1}{1+R}x_2(i)$ are minimized.

The result then follows immediately. ■

This is basically the optimal mixed taxation result due to Atkinson and Stiglitz in a different environment.

Now consider the general case.

Proposition 2 *A second best allocation satisfies for all* $i = 1, ..., n$ *the Inverse Euler Equation*

$$\frac{1}{u'(x_1(i))} = \frac{1}{\beta(1+R)} \sum_j \frac{1}{u'(x_2(j|\, i))} \pi_2(j|\, i)$$

Proof: To see this, consider some second best allocation and take some type i.

Imagine increasing second period consumption utility for this type in a parallel way across realizations.

That is, define $u(\widetilde{x}_2(j|\, i; \Delta)) = u(x_2(j|\, i)) + \Delta$ for $\Delta > 0$ small.

To compensate decrease first period consumption utility by $\beta\Delta$.

That is, define $u(\widetilde{x}_1(i; \Delta)) = u(x_1(i)) - \beta\Delta$ for small Δ.

Note that such variations do not affect the objective function or the incentive constraints.

Thus, for the original allocation to be optimal, it must be that $\Delta = 0$ minimizes the resources expended.

The resources expended are

$$\widetilde{x}_1(i;\Delta) + \frac{1}{1+R}\sum_j \widetilde{x}_2(j|\,i;\Delta)\pi_2(j|\,i)$$

These equal

$$\begin{aligned}&u^{-1}\left(u(x_1(i)) - \beta\Delta\right) + \\ &\frac{1}{1+R}\sum_j u^{-1}\left(u(x_2(j|\,i)) + \Delta\right)\pi_2(j|\,i)\end{aligned}$$

Minimizing this problem and evaluating at $\Delta = 0$ yields the result. ■

Proposition 3 *Suppose that for some i, there exists some j such that $0 < \pi_2(j|\,i) < 1$. Then a second best allocation which is such that $x_2(j|\,i)$ is not independent of j satisfies*

$$u'(x_1(i)) < \beta(1+R)\sum_j u'(x_2(j|\,i))\pi_2(j|\,i)$$

Proof: Apply *Jensen's Inequality* to the result in Proposition 2. ■

Thus, with variable types, in a second best allocation, the intertemporal allocation of resources is distorted.

This is perhaps the most significant result of the New Dynamic Public Finance.

It stands in stark contrast to the results of the optimal capital taxation literature.

ii. Tax Implementations

What type of tax systems will implement the second best allocations?

In this economy, period 1 taxes can be a function of period 1 income; i.e., $T_1(y_1)$.

Period 2 taxes can be a function of period 1 income, period 2 income, and savings; i.e., $T_2(y_1, y_2, s)$.

Given such a tax system, an individual of type i would solve

$$\max u(x_1) - \varphi(\tfrac{y_1}{a_i})$$
$$+\beta \sum_j \left(u(x_2(j)) - \varphi(\frac{y_2(j)}{a_j}) \right) \pi_2(j|\, i)$$

subject to

$$x_1 + s = y_1 - T_1(y_1)$$

and

$$x_2 = y_2 + s(1+r) - T_2(y_1, , y_2, s)$$

The tax system $\{T_1(y_1), T_2(y_1, y_2, s)\}$ *implements* the second best allocation if for all ability types i the solution to this problem coincides with the second best allocation.

In the static Mirrlees model, we can infer the marginal income tax rate from the distortion or "wedge" in the consumption-labor first order condition.

Thus, in the two type model for each type $i \in \{L, H\}$

$$T'(y_i) = 1 - \frac{\varphi'(y_i/a_i)}{a_i}.$$

It is tempting to think that we can infer the marginal tax rates on period 1 and 2 earnings from the consumption-labor wedges and the marginal tax rate on earnings from savings from the intertemporal wedge.

This turns out not to be the case.

Intuitively, each wedge controls only one aspect of the individual's behavior taking all other choices fixed at their optimal levels.

However, individuals choose labor and savings jointly.

Generally, there can be many different tax systems that will implement the second best allocation.

Thus, the notion of *the* optimal tax system may not be well defined.

In an infinite horizon model in which each individual's ability is the realization of an IID random variable, Albanesi and Sleet *ReStud* (2006) show that the second best allocation can be implemented with a tax system of the form $\{T_t(y_t, w_t)\}_{t=0}^{\infty}$ where w_t denotes wealth at the beginning of period t

Thus, period t taxes just depend on period t income and wealth.

In an infinite horizon neoclassical general equilibrium model, Kocherlakota *Ecma* (2005) shows that the second best allocation can be implemented with a tax system which separates capital from labor taxation.

Taxes on labor income in period t depend on the whole history of labor earnings up until period t and can be complicated non-linear functions.

Taxes on capital are linear and also history dependent.

iii. Time Inconsistency

A novel type of time inconsistency problem arises in these type of models.

We have been assuming so far that the government commits to a tax system at the beginning of period 1.

If the optimal tax system induces individuals of different ability types to earn different amounts, an individual's earnings choice in period 1 will reveal his period 1 ability type.

If individuals' ability types are persistent, an individual's period 1 earnings choice will reveal information about his period 2 ability type.

The government would like to use this information to design a better tax system in period 2.

If the government cannot commit, it will use this information.

However, individuals will know this and this will further distort their period 1 choices.

This time inconsistency problem in dynamic Mirrlees models was first pointed out by Roberts *ReStud* (1984).

He considered a T-period repeated Mirrlees model in which ability types were fully persistent.

He provided conditions under which the unique equilibrium involved the highly inefficient outcome in which all types declare themselves to be the lowest ability type in all periods and supply the lowest level of labor and receive the lowest consumption level.

V. Optimal Transfer Programs

All governments in developed countries operate programs to help the poor.

The U.S. federal goverment offers two major cash transfer programs (TANF and SSI) and three major in-kind transfer programs (Medicaid, public housing, and food stamps).

The U.S. government also offers programs to help the working poor such as the EITC and Minimum Wage.

Transfer programs for the poor are distinct from social insurance programs like Social Security, Medicare, and Unemployment Insurance.

There is a vast literature on programs for the poor, both theoretical and empirical.

The lectures will discuss the following issues.

1. Why should the government be in the business of helping the poor?

2. How should cash assistance programs be optimally designed?

3. What is the role for in-kind transfers?

4. What determines public generosity towards the poor?

1. Why Government?

Two approaches can be distinguished.

The *efficiency approach* asserts that government provision of transfers to the poor is necessary to remedy the market failure created by the fact that citizens are altruistic and the welfare of the poor is a public good.

Because the welfare of the poor is a public good, private charity can be expected to under-provide transfers to the poor.

Public provision of transfers therefore can make all citizens better off.

Related ideas are that transfers are Pareto improving because they reduce crime and social unrest.

The *distributive justice approach* views support for the poor as a moral responsibility of the government even if citizens do not care about the poor.

Different notions of distributive justice are embodied in society's social welfare function.

Under a *Utilitarian* social welfare function, the government seeks to maximize aggregate utility.

This implies a case for redistribution from the rich to the poor when there is diminishing marginal utility of income.

Under a *Rawlsian* social welfare function, the government seeks to maximize the utility of the worst off persons in society.

This yields a case for redistribution, even if there is not diminishing marginal utility of income.

2. The Design of Cash Assistance Programs

The literature on the design of cash assistance programs takes some of its inspiration from issues in the TANF program.

The structure of benefits under this program is as follows:

$$B(y) = \begin{cases} G - ry & \text{if } y \leq G/r \\ 0 & \text{if } y > G/r \end{cases}$$

where y is income, G is the *guarantee* and r is the *benefit reduction rate*.

Thus, there is considerable interest in understanding what is the optimal benefit reduction rate or the slope of the benefit function.

Benefits under this program are also *categorical*, in the sense that only single parent families are eligible.

This led to some interest in understanding the use of categorization.

Benefits are also *time limited* and many states impose *work requirements* or *workfare*.

This led to interest in time limits and workfare.

There are two basic approaches to thinking about design problems.

The *general equilibrium approach* embeds the problem of designing cash assistance into the general problem of choosing an optimal tax system for the whole society.

Thus, the cash assistance program is designed as part of the optimal tax system.

The *partial equilibrium approach* considers the problem of minimizing the fiscal cost of providing some target population (e.g., single parent families) with some minimum utility or income level.

For normative work, the minimum utility constraint makes the most sense.

However, as a positive matter, it politicians do not seem to value the leisure of the poor and hence the minimum income constraint is an interesting one to adopt for positive purposes.

A minimum income constraint is an example of a *non-welfarist objective*.

The advantage of the approach is that it does not require the modelling of the entire economy and hence permits a sharper focus on program design.

The general equilibrium approach

We have already looked at this approach when we covered optimal income taxation.

Suppose there are some citizens with either 0 or very low income generating ability.

If the social welfare function is sufficiently concave and/or the marginal utility of consumption is sufficiently diminishing, the solution to the Mirrlees model will involve providing transfers for those with zero or low earnings.

Since we know that $T'(y) \in (0, 1)$, the transfer will be phased out as income rises.

The picture will resemble a non-linear *negative income tax*.

This prescription is somewhat puzzling in light of the EITC program which provides earnings subsidies to those with low earnings.

These earnings subsidies imply that $T'(y) < 0$ for low y.

Earnings subsidies can be rationalized if the relevant labor supply decision for individuals is whether or not to work rather than how hard to work.

The former is referred to as the *extensive margin* and the latter as the *intensive margin*.

The Mirrlees model assumes that individuals respond to the tax system along the intensive margin.

However, the empirical labor supply literature has shown that extensive labor supply elasticities are significant for low income earners.

In particular, the EITC has had positive effects on labor force participation of beneficiaries.

By contrast, evidence of significant responses along the intensive margin is more limited.

Saez *QJE* (2002) analyzes the optimal income taxation problem under the assumption that the relevant margin of response is the extensive margin.

Saez's Model

There are a continuum of individuals and I types of occupations.

Individuals can be unemployed in which case they earn $w_0 = 0$.

Those in occupation i earn w_i where $0 < w_1 < .... < w_I$.

Each individual is characterized by the occupation that he can do $i \in \{0, ..., I\}$ and his disutility of work θ.

Those individuals for whom $i = 0$ have no choice but to be unemployed.

All individuals can choose to be unemployed, but individuals can only do at most one occupation.

The fraction of individuals who can do occupation i is h_i where $\sum_{i=0}^{I} h_i = 1$.

The fraction of individuals whose disutility of work is less than θ is $F(\theta)$.

The government has a revenue requirement R.

The government cannot observe individuals' types (i, θ) but can observe occupational choice.

The tax paid by those in occupation i is T_i.

The government puts weight $\mu(i, \theta)$ on the welfare of type (i, θ) individuals.

The weights are normalized so that

$$\sum_{i=0}^{I} \int_{\theta} \mu(i, \theta) dF(\theta) h_i = 1.$$

The government's problem

Individual (i, θ) will work if

$$w_i - T_i - \theta \geq -T_0.$$

The fraction of individuals working in occupation $i \in \{1, ..., I\}$ is

$$F(w_i - T_i + T_0)h_i.$$

The fraction of unemployed individuals is

$$h_0 + \sum_{i=1}^{I}(1 - F(w_i - T_i + T_0))h_i.$$

Government revenue is

$$\sum_{i=1}^{I} F(w_i - T_i + T_0)h_i T_i$$
$$+ \left[h_0 + \sum_{i=1}^{I}(1 - F(w_i - T_i + T_0))h_i\right] T_0$$

Given the taxes (T_0, T_i) the utility of a type (i, θ) ($i \neq 0$) is

$$
\begin{aligned}
& u[T_0, T_i; (i, \theta)] \\
= & \begin{cases} w_i - T_i - \theta & \text{if } \theta \leq w_i - (T_i - T_0) \\ -T_0 & \text{otherwise} \end{cases}
\end{aligned}
$$

The utility of a type $(0, \theta)$ is just

$$u[T_0; (i, \theta)] = -T_0$$

The government's objective function is

$$\sum_{i=0}^{I} \int_{\theta} \mu(i, \theta) u[T_0, T_i; (i, \theta)] dF(\theta) h_i$$

The government wants to choose a tax system $(T_0, ..., T_I)$ to maximize this objective function subject to its budget constraint.

Results

Define the *elasticity of participation* for occupation $i \in \{1, ..., I\}$ to be

$$\begin{aligned} \eta_i &= \frac{w_i - T_i + T_0}{F(w_i - T_i + T_0)h_i} \frac{\partial \left[F(w_i - T_i + T_0)h_i \right]}{\partial \left(w_i - T_i + T_0 \right)} \\ &= \frac{\left(w_i - T_i + T_0 \right) f(w_i - T_i + T_0)}{F(w_i - T_i + T_0)}. \end{aligned}$$

Define the *welfare weight* for occupation $i \in \{1, ..., I\}$ to be

$$g_i = \frac{\int_{-\infty}^{w_i - (T_i - T_0)} \mu(i, \theta) dF(\theta)}{F(w_i - T_i + T_0)}$$

and the welfare weight for the unemployed to be

$$g_0 = \frac{\int_\theta \mu(0, \theta) dF(\theta) h_0 + \sum_{i=1}^{I} \int_{w_i - (T_i - T_0)}^{\infty} \mu(i, \theta) dF(\theta) h_i}{h_0 + \sum_{i=1}^{I} (1 - F(w_i - T_i + T_0)) h_i}$$

Then we have:

Proposition *The optimal tax system satisfies for each* $i \in \{1, ..., I\}$

$$\frac{T_i - T_0}{w_i - T_i + T_0} = \frac{1}{\eta_i}\left[1 - g_i\right]$$

Proof: Substituting in the utility functions, the objective function is

$$\int_\theta \mu(0,\theta)dF(\theta)h_0(-T_0) + \sum_{i=1}^{I}\left[\begin{array}{c} \int_{-\infty}^{w_i-(T_i-T_0)} \mu(i,\theta)(w_i - T_i - \theta)dF(\theta) \\ + \int_{w_i-(T_i-T_0)}^{\infty} \mu(i,\theta)\left(-T_0\right)dF(\theta) \end{array}\right] h_i$$

The Lagrangian is

$$L = \int_\theta \mu(0,\theta)dF(\theta)h_0(-T_0) + \sum_{i=1}^{I}\left[\begin{array}{c} \int_{-\infty}^{w_i-(T_i-T_0)} \mu(i,\theta)\left(w_i - T_i - \theta\right)dF(\theta) \\ + \int_{w_i-(T_i-T_0)}^{\infty} \mu(i,\theta)\left(-T_0\right)dF(\theta) \end{array}\right] h_i$$

$$+\lambda \left[\begin{array}{c} \sum_{i=1}^{I} F(w_i - T_i + T_0) h_i T_i - R + \\ \left[h_0 + \sum_{i=1}^{I} (1 - F(w_i - T_i + T_0)) h_i \right] T_0 \end{array} \right]$$

The first order condition for T_0 is

$$\lambda \left[\begin{array}{c} h_0 + \sum_{i=1}^{I} (1 - F(w_i - T_i + T_0)) h_i \\ + \sum_{i=1}^{I} f(w_i - T_i + T_0) h_i (T_i - T_0) \end{array} \right] =$$

$$\int_{\theta} \mu(0,\theta) dF(\theta) h_0 + \sum_{i=1}^{I} \int_{w_i - (T_i - T_0)}^{\infty} \mu(i,\theta) dF(\theta) h_i.$$

The first order condition for T_i ($i \in \{1, ..., I\}$) is

$$\lambda \left\{ F(w_i - T_i + T_0) h_i - f(w_i - T_i + T_0) h_i (T_i - T_0) \right\}$$

$$= \int_{-\infty}^{w_i - (T_i - T_0)} \mu(i,\theta) dF(\theta) h_i.$$

The second condition implies that

$$T_i - T_0 = \frac{F(w_i - T_i + T_0)}{f(w_i - T_i + T_0)} - \frac{\int_{-\infty}^{w_i-(T_i-T_0)} \mu(i,\theta)dF(\theta)}{\lambda f(w_i - T_i + T_0)}$$

In addition, substituting the second condition into the first condition implies that

$$\begin{aligned} \lambda &= \int_\theta \mu(0,\theta)dF(\theta)h_0 + \sum_{i=1}^{I} \int_{\infty}^{\infty} \mu(i,\theta)dF(\theta)h_i \\ &= 1 \end{aligned}$$

Thus, we have that

$$T_i - T_0 = \frac{F(w_i - T_i + T_0)}{f(w_i - T_i + T_0)} - \frac{\int_{-\infty}^{w_i-(T_i-T_0)} \mu(i,\theta)dF(\theta)}{f(w_i - T_i + T_0)}$$

Using the definitions of η_i and g_i yields the result. ■

Assume that the government has redistributive tastes so that

$$g_0 > g_1 > ..... > g_I$$

Note that

$$\sum_{i=1}^{I} g_i F(w_i - T_i + T_0) h_i$$
$$+ g_0 \left[h_0 + \sum_{i=1}^{I} (1 - F(w_i - T_i + T_0)) h_i \right] = 1$$

Thus, there must exist i^* such that $g_i \geq 1$ for $i \leq i^*$ and $g_i < 1$ for $i > i^*$.

It follows from the Proposition that $T_i \leq T_0$ for $i \leq i^*$ and $T_i < T_0$ for $i > i^*$.

Assuming $i^* \geq 1$, the government provides a higher transfer to workers in occupation 1 than the unemployed (or, equivalently, levies a lower tax).

Thus, the optimal tax system involves an earnings subsidy.

The partial-equilibrium approach

An illustrative model of the partial equilibrium approach follows.

There is a population of potentially poor people (e.g., single parents) divided into two types according to their income generating ability or wage, denoted by a_i, where $0 < a_1 < a_2$.

Let π_i denote the fraction of type i's.

Each individual is endowed with T units of time and has a quasi-linear utility function

$$u = x - \varphi(l)$$

where x is consumption and l is labor supply.

The function φ is increasing, strictly convex and satisfies $\varphi'(0) = 0$ and $\varphi'(T) > a_2$.

In the absence of government intervention, individuals of type i would have incomes

$$\widetilde{y}_i = \arg\max\{y - \varphi(y/a_i)\}$$

and enjoy utility levels

$$\widetilde{u}_i = \widetilde{y}_i - \varphi(\widetilde{y}_i/a_i).$$

The government is concerned that all citizens utility be above some target minimal level $\underline{u}$ but desires to ensure that this objective is reached at minimum fiscal cost.

This is a utility-based objective as opposed to an income-based objective.

The problem is *utility maintenance* rather than *income maintenance*.

The income maintenance problem can also be analyzed in a similar manner.

Assume that

$$\widetilde{u}_2 > \underline{u} > \widetilde{u}_1.$$

Thus, the type 1s are needy and the type 2s are non-needy.

Assume further that the government cannot observe individuals' types.

Issues

There are a number of issues that can be studied in this type of model.

First, we can study optimal linear benefit schedules of the form used in the TANF program

$$B(y) = \begin{cases} G - ry & \text{if } y \leq G/r \\ 0 & \text{if } y > G/r \end{cases}$$

In particular, we can characterize the optimal benefit reduction rate.

Second, we can look at optimal non-linear benefit schedules using the techniques we used to study the two type Mirrlees model.

Third, we can explore the case for workfare.

I will ask you to explore these issues in a problem set.

3. The Role for In-kind Transfers

If the concern of policy-makers is the utility of the poor, then basic microeconomics suggests that cash transfers are always better than in-kind transfers.

This is because cash transfers allow recipients to choose the bundle of goods that gives them the greatest utility.

This notwithstanding, in-kind transfers to the poor are common in the U.S. and elsewhere.

The literature identifies a number of possible rationales for in-kind transfers

Paternalism

The simplest explanation is that tax-payers and policy-makers do not care about the utility of the poor, they just care about their consumption of food, housing, and healthcare.

This is related to the point made earlier concerning politicians not valuing the leisure time of the poor.

This seems to be the case, as people are far more willing to give the homeless food or shelter than they are to give them cash.

Often this may be motivated by the concern that some poor people are not the best judge of their own welfare because they have problems with addictive goods such as drugs and alcohol.

Targeting

A slightly more subtle explanation is that in-kind transfers help policy-makers target assistance to those who need it the most.

For example, suppose that the government wants to help the homeless but that being homeless is not easily observable.

Then if the government offers cash transfers to all those who claim to be homeless, then this is likely to attract more recipients than would an offer of free sleepovers at a homeless shelter.

On this argument see Blackorby and Donaldson *AER* (1988).

Pecuniary effects

An in-kind program will likely have different price effects from a cash program.

For example the building of public housing will depress housing rents by more than would an equally costly cash transfer program.

This price effect may allow the government to get more bang for its buck in terms of helping the poor.

On this argument, see Coate, Johnson, and Zeckhauser *JPubE* (1994)

Samaritan's dilemma

The most interesting explanation of in-kind transfers stems from the *Samaritan's Dilemma*.

This argument motivates the provision of in-kind transfers of insurance and/or self protection

We illustrate the Samaritan's Dilemma argument using the model of Coate *AER* (1995)

Coate's model

There are three individuals, two are rich and the other is poor.

The poor person has income y_p and faces uncertainty

With probability π he suffers a loss L

His utility function is $u(x)$ where u is increasing and strictly concave

The poor person can purchase insurance against his loss: z units of coverage costs πz

The rich individuals have incomes $y_r > y_p$.

They are altruistic towards the poor person and have utility functions $u_r^i = x + \delta u_p$ where $i \in \{1, 2\}$, x is consumption and u_p is the poor person's utility.

There is a government that acts so as to maximize the aggregate utility of the rich; i.e., to maximize the welfare function $W = u_r^1 + u_r^2$.

Because the rich are altruistic and because of the free rider problem, the government operates a transfer program for the poor person.

Let T be the transfer to the poor person and assume it is financed by taxes $T/2$ on each rich person.

The timing of the interaction between the citizens is as follows.

1) The government chooses a transfer.

2) The poor person makes an insurance choice.

3) Nature chooses whether or not the poor person experiences a loss.

4) The rich provide charitable transfers.

Cash transfers when the rich can commit

Suppose the rich can commit not to give charity to the poor person.

Since he is risk averse the poor person would fully insure himself and obtain utility $u(y_p - \pi L + T)$

The optimal government transfer is

$$T^o = \arg\max\{2y_r - T + 2\delta u(y_p - \pi L + T)\}$$

implying that

$$2\delta u'(y_p - \pi L + T^o) = 1$$

Let W^o denote the maximal level of welfare.

Cash transfers when the rich cannot commit

Suppose the rich cannot commit not to give charity to the poor person.

Suppose further that the government gives the poor person a transfer T and assume that $\delta u'(y_p - \pi L + T) \leq 1$ so that the rich would have no incentive to give transfers if the poor person fully insured.

Consider the poor person's insurance decision.

If he fully insures his utility will be $u(y_p - \pi L + T)$.

If he does not fully insure he may get transfers from the rich if he suffers a loss.

Suppose he purchases $z < L$ units of coverage.

If he suffers a loss the rich guys will choose transfers (τ_1^*, τ_2^*) where for $i = 1, 2$

$$\tau_i^* = \arg\max_{\tau_i \geq 0} \{ y_r - \frac{T}{2} - \tau_i + \delta u(y_p + (1-\pi)z - L + T + \tau_{-i}^* + \tau_i) \}$$

The aggregate transfer will be

$$\tau^*(T, z) = \max\{0, \xi(T) - (1-\pi)z\}$$

where $\xi(T)$ is such that

$$\delta u'(y_p - L + T + \xi(T)) = 1.$$

The optimal amount of insurance coverage for the poor person is

$$z^*(T) = \arg\max_{z \geq 0} \{ \pi u(y_p + T + (1-\pi)z - L + \tau^*(T, z)) + (1-\pi)u(y_p + T - \pi z) \}$$

Lemma *(i) Either* $z^*(T) = 0$ *or* $z^*(T) = L$. *(ii) If* $z^*(T^o) = 0$ *and* $T < T^o$ *then* $z^*(T) = 0$.

Proof: For (i) substitute in the expression for $\tau^*(T, z)$ into the poor person's expected utility and consider how it depends on z.

Note that the poor person's expected utility is decreasing in z as long as charitable transfers are positive and increasing thereafter.

Thus, the optimal choice of z is either no coverage or full coverage.

For (ii) simply note that the difference in the payoffs from having no coverage and full coverage is decreasing in T. ■

Proposition 1 *If* $z^*(T^o) = 0$, *the government cannot achieve the welfare level* W^o *with a cash transfer program.*

Proof: If $T > T^o$ the the government will be transferring too much from the viewpoint of the rich.

If $T < T^o$ then the poor person will not take out insurance and the allocation of resources will be inefficient.

Why inefficient?

(i) the poor person will not be fully insured.

(ii) if the bad state happens the poor person will not have enough consumption from the viewpoint of the rich because of free riding. ■

In-kind transfers when the rich cannot commit

Proposition 2 *The government can achieve the welfare level W^o with a transfer program that provides the poor person with L units of insurance coverage and a cash transfer $T^o - \pi L$.*

Alternatively, the government could simply *mandate* that the poor person buy insurance.

The key point is that the insurance decision cannot be delegated to the poor person.

4. What determines public generosity towards the poor?

Different countries and even different U.S. states offer very different levels of support for the poor.

An interesting question is why.

There are a number of papers on this question.

The simplest answer is that citizens have different degrees of altruism.

This might be driven by (say) ethnic diversity as discussed by Luttmer *JPE* (2001).

Piketty *QJE (1995)* points out that even if citizens are equally altruistic, they will choose different levels of redistribution if they have different beliefs about the incentive costs of redistribution.

He then argues that societies with the same underlying economic environments can end up with very different beliefs about the incentive costs of redistribution.

This provides a multiple equilibria explanation of why societies have different levels of poor support.

Piketty's Model

There are an infinite number of time periods indexed by $t = 0, 1, 2, ...$

There are a continuum of dynastic families indexed by $i \in [0, 1]$

Each period represents a generation and each dynasty has one offspring each period

Let y_{it} be dynasty i's pre-tax income in period t

Assume that $y_{it} \in \{y_0, y_1\}$ where $0 < y_0 < y_1$

Let L_t be the fraction of dynasties at time t who had low income in the previous period; that is,

$$L_t = m(\{i \in [0, 1] : y_{it-1} = y_0\})$$

where $m(I)$ is the measure of the set $I \subset [0, 1]$

Let $H_t = 1 - L_t$ be the fraction of dynasties at time t who had high income in the previous period

The probability that a dynasty has income y_1 depends upon effort, luck, and social origin.

Thus, letting e_{it} be dynasty i's effort in period t

$$\Pr\{y_{it} = y_1 \,|y_{it-1} = y_0, e_{it} = e\} = \pi_0 + \theta e$$

and

$$\Pr\{y_{it} = y_1 \,|y_{it-1} = y_1, e_{it} = e\} = \pi_1 + \theta e$$

where $0 < \pi_0 < \pi_1 < 1$ and $\theta > 0$.

The material welfare of dynasty i at time t is

$$U_{it} = x_{it} - \frac{(e_{it})^2}{2a}$$

where x_{it} denotes their consumption and $a > 0$.

The government operates a simple negative income tax which taxes all incomes at rate τ and redistributes the proceeds in a uniform manner.

Thus, suppose the tax rate is τ at some time t and aggregate income is Y_t. Then a dynasty with income y_0 obtains a post tax income

$$(1-\tau)y_0 + \tau Y_t$$

while a dynasty with income y_1 obtains a post tax income

$$(1-\tau)y_1 + \tau Y_t$$

The timing of actions for generation t is as follows:

(i) choose effort levels e_{it}

(ii) incomes y_{it} are realized

(iii) choose tax rate τ_{t+1}

Thus, tax rates for the current generation are determined by the past generation.

When voting over tax rates, all dynasties in period t have the same objective of maximizing the expected welfare of the lower class children; i.e.,

$$V_{t+1} = \int_{i \in L_{t+1}} U_{it+1} di$$

To figure out preferred tax rates, suppose the tax rate for generation $t+1$ is τ_{t+1}

Then the effort choice of generation $t+1$ is

$$e(\tau_{t+1}; \theta) = \arg\max \theta e(1 - \tau_{t+1})(y_1 - y_0) - \frac{e^2}{2a}$$

implying that

$$e(\tau_{t+1}; \theta) = a\theta(1 - \tau_{t+1})(y_1 - y_0)$$

Thus, the optimal tax rate is

$$\begin{aligned} \tau_{t+1}(\pi_1 - \pi_0, \theta) = & \arg\max_{\tau} (\pi_0 + \theta e(\cdot))(1 - \tau) y_1 \\ & + (1 - \pi_0 - \theta e(\cdot))(1 - \tau) y_0 \\ & + \tau [y_0 + (\pi_0 L_{t+1} + \pi_1 H_{t+1} \\ & + \theta e(\cdot))(y_1 - y_0)] - \frac{e(\cdot)^2}{2a} \end{aligned}$$

implying that

$$\tau_{t+1}(\pi_1 - \pi_0, \theta) = \frac{H_{t+1}(\pi_1 - \pi_0)}{a(y_1 - y_0)\theta^2}$$

If the parameters (π_0, π_1, θ) were known by everyone, then everybody would agree on the optimal tax rate.

Starting from any initial condition (L_0, H_0, τ_0) the economy would converge toward a unique steady-state distribution $(L_\infty, H_\infty, \tau_\infty)$.

Dynastic Learning

Now assume that dynasties have different beliefs about the structural parameters (π_0, π_1, θ).

Let $(\pi_0^*, \pi_1^*, \theta^*)$ be the true values.

What happens in the long run if the dynasties start out with different beliefs?

Assume that each dynasty learns from its own experience only.

The initial state of the economy is $(L_0, H_0, \tau_0, (\mu_{i0})_{i \in [0,1]})$ where μ_{i0} represents the prior beliefs of dynasty i - it can be any probability measure defined on the set of all logically possible (π_0, π_1, θ)

Then in period 0 dynasty i:

(i) chooses effort level $e_{i0}(\tau_0, \mu_{i0})$ to maximize its expected welfare

(ii) rationally updates its beliefs given its income y_{i0}

(iii) votes over τ_1 by supporting its socially optimal policy $\tau_{i1}(\mu_{i1}(\cdot))$ given its posterior beliefs μ_{i1}

(iv) transmits its posterior beliefs to its offspring

and so on for the next generation.

Effort choices

Let $\theta(\mu_{it})$ denote the expected value of θ given beliefs μ_{it}; i.e.,

$$\theta(\mu_{it}) = \sum_{\text{supp}(\mu_{it})} \theta \mu_{it}(\pi_0, \pi_1, \theta)$$

then dynasty i at time t will choose effort level

$$e_{it}(\tau_t, \mu_{it-1}) = e(\tau_t, \theta(\mu_{it-1}))$$

Voting decisions

Suppose that the average posterior beliefs of those in generation t (other than i) concerning θ are θ_t; i.e.,

$$\theta_t = \int_{j \neq i} \theta(\mu_{jt}) dj.$$

Then the optimal tax in period $t+1$ from the viewpoint of dynasty i in period t is

$$\begin{aligned}\tau_{it+1}(\mu_{it}) = \arg\max_{\tau} \sum_{\text{supp}(\mu_{it})} & [(\pi_0 + \theta e(\tau, \theta_t))(1-\tau)y_1 \\ & + (1 - \pi_0 - \theta e(\tau, \theta_t))(1-\tau)y_0 \\ & + \tau[y_0 + (\pi_0 L_{t+1} + \pi_1 H_{t+1} \\ & + \theta e(\tau, \theta_t))(y_1 - y_0)] \\ & - \frac{e(\tau, \theta_t)^2}{2a}]\mu_{it}(\pi_0, \pi_1, \theta)\end{aligned}$$

implying that

$$\tau_{it+1}(\mu_{it}) = \frac{H_{t+1}(\pi_1(\mu_{it}) - \pi_0(\mu_{it}))}{a(y_1 - y_0)\theta_t^2} + \frac{\theta_t - \theta(\mu_{it})}{\theta_t}$$

where $\pi_0(\mu_{it})$ and $\pi_1(\mu_{it})$ denote the expected values of π_0 and π_1 given beliefs μ_{it}

Then, applying the median voter theorem, the policy outcome is

$$\tau_{t+1} = med(\tau_{it+1}(\mu_{it})_{i\in[0,1]})$$

Belief Updating

The dynastic Bayesian updating is perfectly standard.

Consider for example dynasty i in generation t who has prior beliefs μ_{it-1}. Suppose that i is lower class ($i \in L_t$) but that its income in period t is y_1.

Then, for any $(\pi_0, \pi_1, \theta) \in$supp$(\mu_{it-1})$, we have that

$$
\begin{aligned}
& \mu_{it}(\pi_0, \pi_1, \theta) \\
= \ & \mu_{it-1}(\pi_0, \pi_1, \theta) \times \\
& \frac{\pi_0 + \theta e(\tau_t, \theta(\mu_{it-1}))}{\sum_{\text{supp}(\mu_{it-1})}[\pi_0' + \theta' e(\tau_t, \theta(\mu_{it-1}))]\mu_{it}(\pi_0', \pi_1', \theta')}
\end{aligned}
$$

Steady-State Political Attitudes

Proposition 1 *Whatever the initial condition* $(L_0, H_0, \tau_0, (\mu_{i0})_{i\in[0,1]})$ *for each dynasty* i *the belief* $\mu_{it}(\cdot)$ *converges with probability one toward some stationary belief* $\mu_{i\infty}(\cdot)$ *as* t *goes to* ∞*. The equilibrium tax rate* τ_t *converges toward some tax rate* τ_∞*.*

Does every dynasty necessarily adopt the same beliefs in steady state?

Is the long run tax rate equal to the true socially optimal tax rate?

To make the issue non-trivial assume that every dynasty's initial belief puts positive probability on the truth.

Assumption 1 *For every dynasty* i*,* $\mu_{i0}(\pi_0^*, \pi_1^*, \theta^*) > 0$.

To characterize what happens, for any tax rate τ let $S(\tau)$ be the set of beliefs $\mu(\cdot)$ such that

(i) for all $(\pi_0, \pi_1, \theta) \in$supp$(\mu)$

$$\pi_0 + \theta e(\tau, \theta(\mu)) = \pi_0^* + \theta^* e(\tau, \theta(\mu))$$

and

$$\pi_1 + \theta e(\tau, \theta(\mu)) = \pi_1^* + \theta^* e(\tau, \theta(\mu))$$

and (ii) $(\pi_0^*, \pi_1^*, \theta^*) \in$supp$(\mu)$.

Condition (i) says that when the tax rate is τ beliefs in $S(\tau)$ generate effort decisions $e(\tau, \theta(\mu))$ that lead to expected probabilities of upward mobility which are the same across all points in the support and coincide with the true probabilities.

A dynasty starting with beliefs in the set $S(\tau)$ will never modify these beliefs (assuming that the tax rate is τ.)

Proposition 2 *(a) Under Assumption 1, whatever the initial condition* $(L_0, H_0, \tau_0, (\mu_{i0})_{i\in[0,1]})$ *the long run steady state is such that:*

(i) for all i*,* $\mu_{i\infty}(\cdot) \in S(\tau_\infty)$

(ii) τ_∞ *is the median of* $med(\tau_{i\infty}(\mu_{i\infty})_{i\in[0,1]})$*.*

(b) For any beliefs distribution and tax rate $((\mu_{i\infty})_{i\in[0,1]}, \tau_\infty)$ *satisfying (i) and (ii) there exists some initial condition* $(L_0, H_0, \tau_0, (\mu_{i0})_{i\in[0,1]})$ *satisfying Assumption 1 such that the associated long run steady state is* $((\mu_{i\infty})_{i\in[0,1]}, \tau_\infty)$

The key point is that there are many beliefs that lead to no contradiction between expectation and experience.

Let $\Delta(\tau)$ be the set of all (π_0, π_1, θ) such that

$$\pi_0 + \theta e(\tau, \theta) = \pi_0^* + \theta^* e(\tau, \theta)$$

and

$$\pi_1 + \theta e(\tau, \theta) = \pi_1^* + \theta^* e(\tau, \theta)$$

Defining

$$\pi_0(\theta) = \pi_0^* + (\theta^* - \theta) e(\tau, \theta),$$

we can write this as

$$\Delta(\tau) = \{(\pi_0(\theta), \pi_1^* - \pi_0^* + \pi_0(\theta), \theta) : \theta \geq 0\}.$$

Parameters in $\Delta(\tau)$ are indistinguishable from the real parameters in the sense that if one believes these are the true parameters one will take an effort level leading to expectations about income mobility that exactly coincides with experience.

All beliefs $\mu \in S(\tau)$ have their averages $(\pi_0(\mu), \pi_1(\mu), \theta(\mu))$ in $\Delta(\tau)$.

Conversely, for any (π_0, π_1, θ) in $\Delta(\tau)$ we can find many beliefs $\mu \in S(\tau)$ whose averages equal (π_0, π_1, θ).

Suppose that we are in a long run steady state with tax rate τ_∞.

Then the dynasties will have beliefs whose averages lie in the set

$$\Delta(\tau_\infty) = \{(\pi_0(\theta), \pi_1^* - \pi_0^* + \pi_0(\theta), \theta) : \theta \geq 0\}.$$

The key point to note is that these dynasties all agree about the difference between the averages $\pi_1(\mu_{i\infty})$ and $\pi_0(\mu_{i\infty})$ - and they think it equals $\pi_1^* - \pi_0^*$

Where they disagree is on the returns to effort $\theta(\mu_{i\infty})$ - they can be arrayed along a line with the median voter being at the median position.

Those dynasties for whom $\theta(\mu_{i\infty})$ is low, vote for more redistribution and put in less effort.

It follows that a higher fraction of lower tax rate supporters have high income

Moreover, those who support lower tax rates have a higher probability of being upwardly mobile.

The model thus delivers nice predictions concerning support for redistribution as a function of current income and social class as measured by parental income.

Piketty argues that these predictions are consistent with the data.

VI. Social Insurance

Social insurance programs provide transfers based on *events* such as unemployment, disability, or age.

Examples in the U.S. are Social Security, Unemployment Insurance, Workers Compensation, and Disability Insurance.

To be eligible for benefits, a worker must have paid in to the programs - this is what makes them insurance programs.

They should be contrasted with welfare programs that provide transfers to those who are poor: social insurance benefits are not means-tested.

Social insurance spending is the biggest and most rapidly growing part of government expenditure today.

In 1953, social security comprised 3.6% of the government budget, in 2008 it comprised 20.7%.

In 1953, income security comprised 5% of the government budget, in 2008 it comprised 14.5%.

In these lectures, we will first explore motivations for the provision of social insurance.

Then we will look at work on designing unemployment insurance, social security, and disability insurance.

1. Why have social insurance?

The motivation for insurance is that it reduces risk for risk-averse individuals

Thus, Unemployment Insurance reduces the risk from involuntary unemployment

Workers Compensation and Disability Insurance reduce the risk from work-related injuries and career-ending disabilities.

Social Security reduces the risk from living to long.

But why is government intervention needed to provide this insurance?

Possible sources of market failure here are: i) Adverse selection, ii) Samaritan's Dilemma, iii) Individual optimization errors, and iv) Aggregate risk.

i) Adverse selection

Adverse selection can arise when consumers know more about their risk type than do insurance companies.

This is a natural assumption in many insurance applications: e.g., auto, health, unemployment.

Seminal paper is Rothschild and Stiglitz (1976) who analyzed what competitive equilibrium might look like in insurance markets when consumer risk types are unobservable.

They showed that market failure would result.

We will follow the treatment in Mas-Colell, Whinston, and Green: Chp 13.D and Ex 13.D.2.

Rothschild-Stiglitz model

There are two types of consumers: high risk types (H) and low risk types (L).

The fraction of high risk types is $\lambda \in (0, 1)$.

Each consumer has initial income y and faces the possibility of loss L.

Type i consumers have probability of loss π_i where $\pi_H > \pi_L$.

Each consumer is risk-averse with utility function $u(c)$ $(u' > 0$ and $u'' < 0)$ where c is consumption.

An *insurance contract* is characterized by a vector (p, T).

p is the premium and T is the transfer paid out in the event of loss.

Type i's expected utility from the contract (p, T) is

$$\pi_i u(y - p - L + T) + (1 - \pi_i)u(y - p).$$

There are two risk-neutral insurance firms who compete for consumers by offering insurance contracts.

We model the interaction between firms and consumers as a two stage game.

Stage 1: The two firms simultaneously announce sets of offered insurance contracts.

Each firm may announce any finite number of contracts.

Stage 2: Consumers choose which contract, if any, to buy.

We assume: i) if a consumer is indifferent between two contracts he chooses the one with the highest transfer;

ii) if a consumer is indifferent between choosing a contract and not, he chooses the contract; and

iii) if a consumer's most preferred contract is offered by both firms, he chooses each firm's contract with probability $1/2$.

We study the pure strategy subgame perfect Nash equilibria (SPNEs) of this game.

Observable types

If firms could observe consumers' types they can offer differ contracts to different types of consumers.

The outcome in this case is full insurance.

Proposition 1 *In any SPNE with observable types, a type i consumer accepts contract $(p_i^*, T_i^*) = (\pi_i L, L)$ and firms earn zero profits.*

Proof: It is straightforward to show that any contract (p_i^*, T_i^*) a type i consumer accepts must be such that $p_i^* = \pi_i T_i^*$ (i.e., firms earn zero profits).

If $p_i^* < \pi_i T_i^*$, a firm could do better by not offering the contract.

If $p_i^* > \pi_i T_i^*$, one firm could do better by offering a slightly lower premium and serving the whole market.

Thus, we just need to show that $T_i^* = L$.

This can be done diagrammatically.

Let c_L denote consumption in the loss state and c_N consumption in the no-loss state.

Then, with a contract of the form $(p, T) = (\pi_i T, T)$, a type i consumer's consumption in the two states is

$$c_L = y - \pi_i T - L + T$$

and

$$c_N = y - \pi_i T.$$

The latter implies

$$T = \frac{y - c_N}{\pi_i}$$

so that

$$
\begin{aligned}
c_L(c_N) &= y - \pi_i\left(\frac{y - c_N}{\pi_i}\right) - L + \frac{y - c_N}{\pi_i} \\
&= y + (1 - \pi_i)\left(\frac{y - c_N}{\pi_i}\right) - L
\end{aligned}
$$

This defines the locus of consumption pairs consistent with zero profits.

The slope of this locus is

$$
\frac{dc_L(c_N)}{dc_N} = -\frac{1 - \pi_i}{\pi_i}.
$$

The type i's consumer's indifference curves are defined by

$$
\pi_i u(c_L) + (1 - \pi_i)u(c_N) = \overline{u}
$$

These indifference curves have slope

$$
\frac{dc_L(c_N)}{dc_N} = -\left[\frac{1 - \pi_i}{\pi_i}\right]\frac{u'(c_N)}{u'(c_L)}
$$

Note that the indifference curves are tangent to the zero profit line where $c_L = c_N$, which implies that $T = L$.

Now suppose that type i consumers were accepting a contract $(p_i^*, T_i^*) = (\pi_i T, T)$ for some $T \neq L$.

Such a contract is illustrated in Fig 4.

Then either firm could deviate and earn strictly positive profits by offering a contract in the shaded area of Fig 4 such as $(\widetilde{p}, \widetilde{T})$. ■

The competitive equilibrium in the observable types case is Pareto efficient.

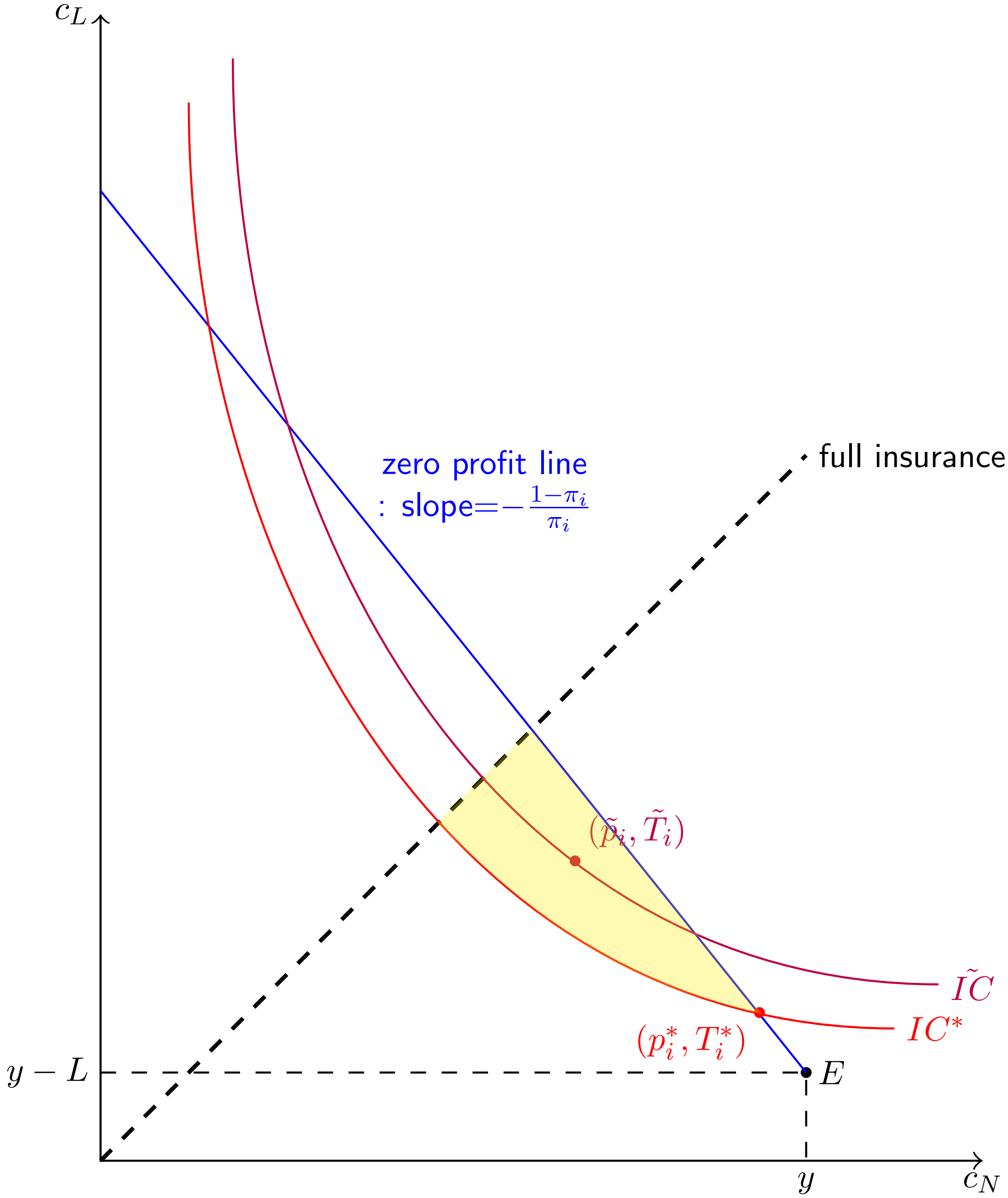

Figure 4: Prop 1. Full insurance with zero profits

E is the initial endowment without insurance. Zero profit line equates the insurance firm's revenue p_i with the expected cost $\pi_i T_i$. The type i's consumer's indifference curves are defined by $\pi_i u(c_L) + (1 - \pi_i)u(c_N) = \bar{u}$. Suppose (p_i^*, T_i^*) is optimal. Any contract in a shaded area in yellow, say, $(\tilde{p}_i, \tilde{T}_i)$ dominates (p_i^*, T_i^*) and the insurance firm gets a positive expected profit. Thus (p_i^*, T_i^*) is not optimal.

Unobservable types

If firms cannot observe consumer types, then Proposition 1 does not hold.

The reason is that high types would accept the contract intended for low types and then firms would lose money.

In principle, there are two possible types of equilibria: *pooling* and *separating*.

In a pooling equilibrium, both types accept the same contract.

In a separating equilibrium, types accept different contracts.

We characterize the possible equilibria via a series of Lemmas.

Lemma 1 *In any SPNE with unobservable types, firms earn zero profits.*

Proof: Let (p_L, T_L) and (p_H, T_H) be the contracts chosen by the two types and let the aggregate profits of the two firms be Π.

Clearly, $\Pi \geq 0$, otherwise at least one firm would be better off not participating.

Assume $\Pi > 0$. Then one firm must be making no more than $\Pi/2$.

Consider a deviation by the low profit firm in which it offers contracts $(p_L - \varepsilon_L, T_L)$ and $(p_H - \varepsilon_H, T_H)$ for $\varepsilon_L > 0$ and $\varepsilon_H > 0$ very small and such that

$$
\begin{aligned}
& \pi_i u(y - p_i + \varepsilon_i - L + T_i) + (1 - \pi_i) u(y - p_i + \varepsilon_i) \\
\geq \;& \pi_i u(y - p_{-i} + \varepsilon_{-i} - L + T_{-i}) \\
& + (1 - \pi_i) u(y - p_{-i} + \varepsilon_{-i})
\end{aligned}
$$

Contract $(p_L - \varepsilon_L, T_L)$ will attract all the low types and contract $(p_H - \varepsilon_H, T_H)$ all the high types.

Since ε_L and ε_H can be made arbitrarily small, this deviation must yield the firm close to Π. ■

Lemma 2 *No pooling equilibria exist.*

Proof: Suppose to the contrary that there is a pooling equilibrium in which the contract is (p, T).

By Lemma 1, it must be the case that $p = \overline{\pi} T$ where $\overline{\pi} = \lambda \pi_H + (1 - \lambda)\pi_L$.

Suppose that firm 1 is offering contract (p, T) as illustrated in Fig 5.

Then firm 2 has a deviation that yields a strictly positive profit.

It offers a single contract $(\widetilde{p}, \widetilde{T})$ that satisfies $\widetilde{p} > \pi_L \widetilde{T}$ lying somewhere in the shaded region of Fig 5.

This contract attracts all the low types and none of the high types.

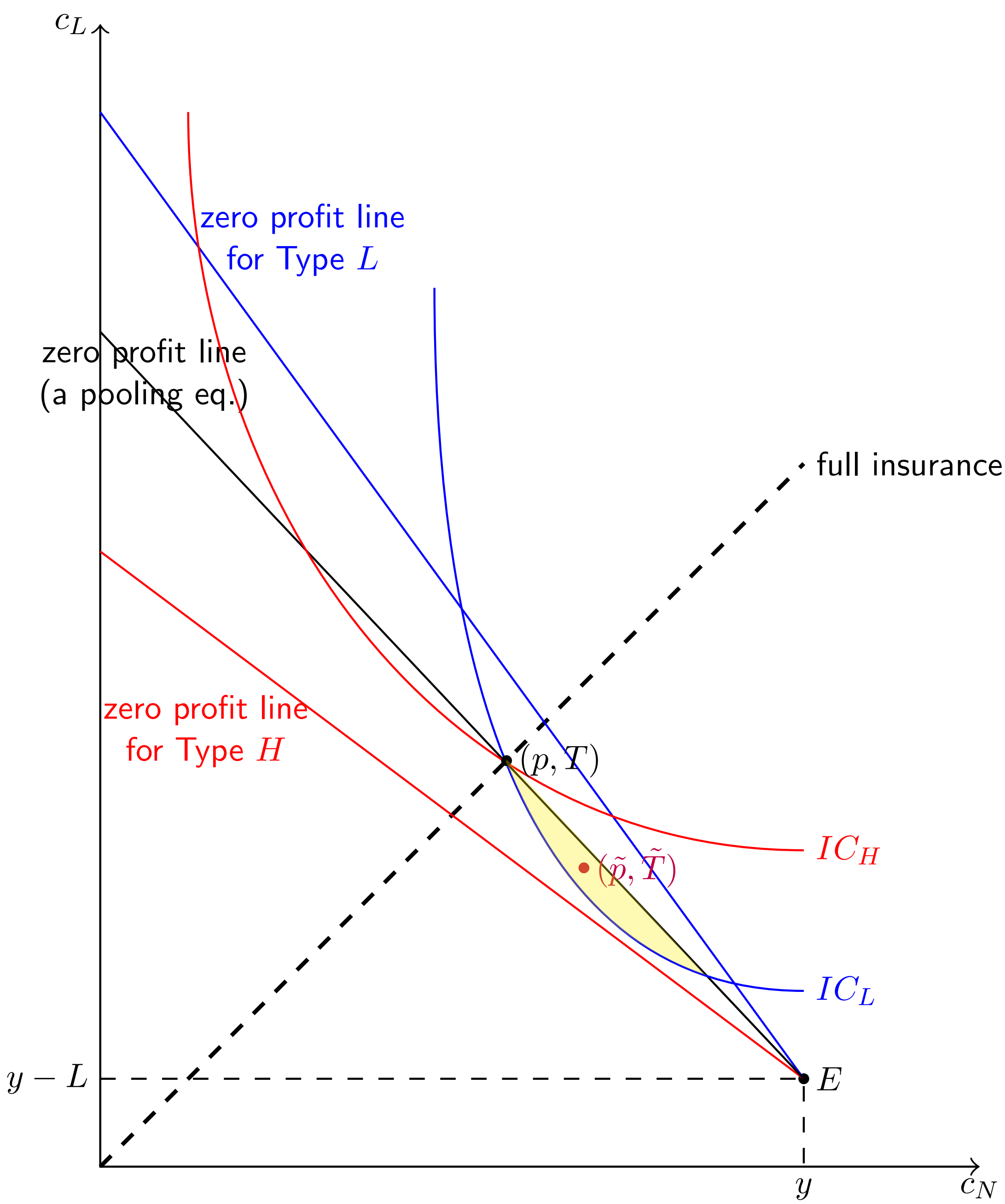

Figure 5: Lem 2. No Pooling equilibrium

Since $\pi_H > \pi_L$, zero profit line for type L is steeper. If a pooling equilibrium exists, then by Lemma 1 the equilibrium contract must be on the insurance firm's zero profit line. If an insurance firm offers $(\tilde{p}, \tilde{T})$, only type L is attracted, and the firm has a positive profit. Therefore, (p^*, T^*) is not an equilibrium.

It therefore makes positive profits - which is a contradiction. ■

Intuitively, the pooling equilibrium is broken by one firm offering a contract that has less insurance but a lower premium which attracts only the low risk types.

This is analogous to a “cream-skimming” strategy.

Lemma 3 *If* (p_L, T_L) *and* (p_H, T_H) *are the contracts chosen by the two types in a separating equilibrium. Then* $p_L = \pi_L T_L$ *and* $p_H = \pi_H T_H$.

Proof: Suppose first that $p_H > \pi_H T_H$.

Then either firm could make positive profits by offering only contract $(p_H - \varepsilon, T_H)$ where $\varepsilon > 0$ is sufficiently small.

All high risk consumers would accept this contract, and any low risk consumers who took it would contribute positive profits.

This is a contradiction.

Next suppose that $p_L > \pi_L T_L$ as in Fig 6.

If we have a separating equilibrium (p_H, T_H) must be such that $p_H < \pi_H T_H$ to guarantee zero profits.

Suppose firm 1 is offering the contract (p_H, T_H).

Then firm 2 could earn strictly positive profits by deviating and offering only a contract $(\widetilde{p}, \widetilde{T})$ in the shaded region.

This would attract only low types, because high types prefer (p_H, T_H).

This deviation would yield positive profits - which is a contradiction.

Since $p_H \leq \pi_H T_H$ and $p_L \leq \pi_L T_L$, the fact that firms make zero profits implies the result. $\blacksquare$

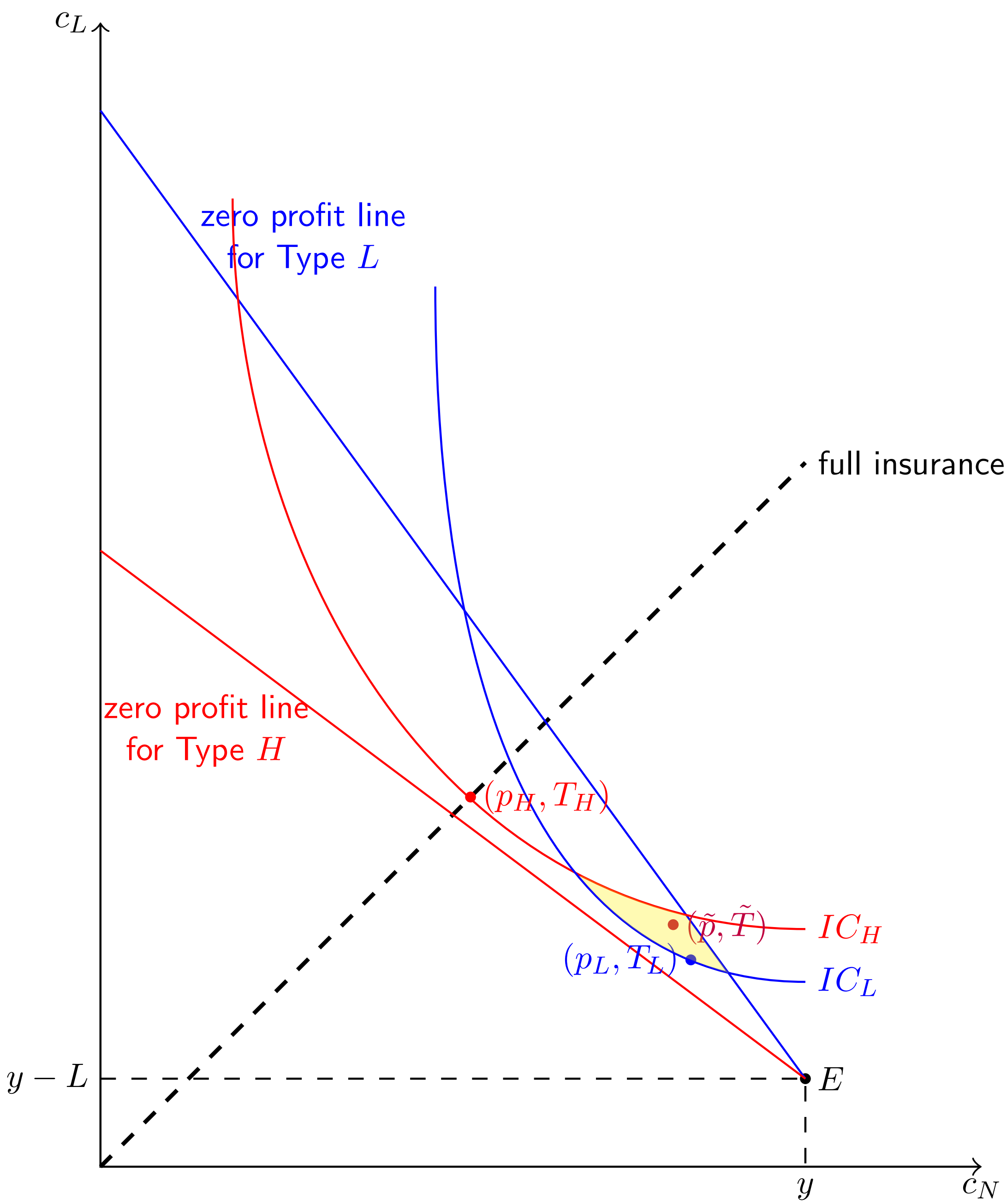

Figure 6: Lem 3. $p_i = \pi_i T_i$, $i = L, H$, under a separating equilibrium

Suppose $p_L > \pi_L T_L$, then (p_H, T_H) must be such that $p_H < \pi_H T_H$ to guarantee zero profits. A firm gets a positive profit by providing $(\tilde{p}, \tilde{T})$ which makes type L consumers better off.

Lemma 4 *If* (p_H, T_H) *is the contract chosen by high types in a separating equilibrium. Then* $(p_H, T_H) = (\pi_H L, L)$.

Proof: By Lemma 3 we have that $p_H = \pi_H T_H$.

Suppose that $T_H \neq L$ as in Fig 7.

Then a firm can make positive profits by offering a contract such as $(\widetilde{p}, \widetilde{T})$ in the shaded region.

All high risk types choose this contract and this contract yields positive profits from any consumer who accepts it.

This is a contradiction. ■

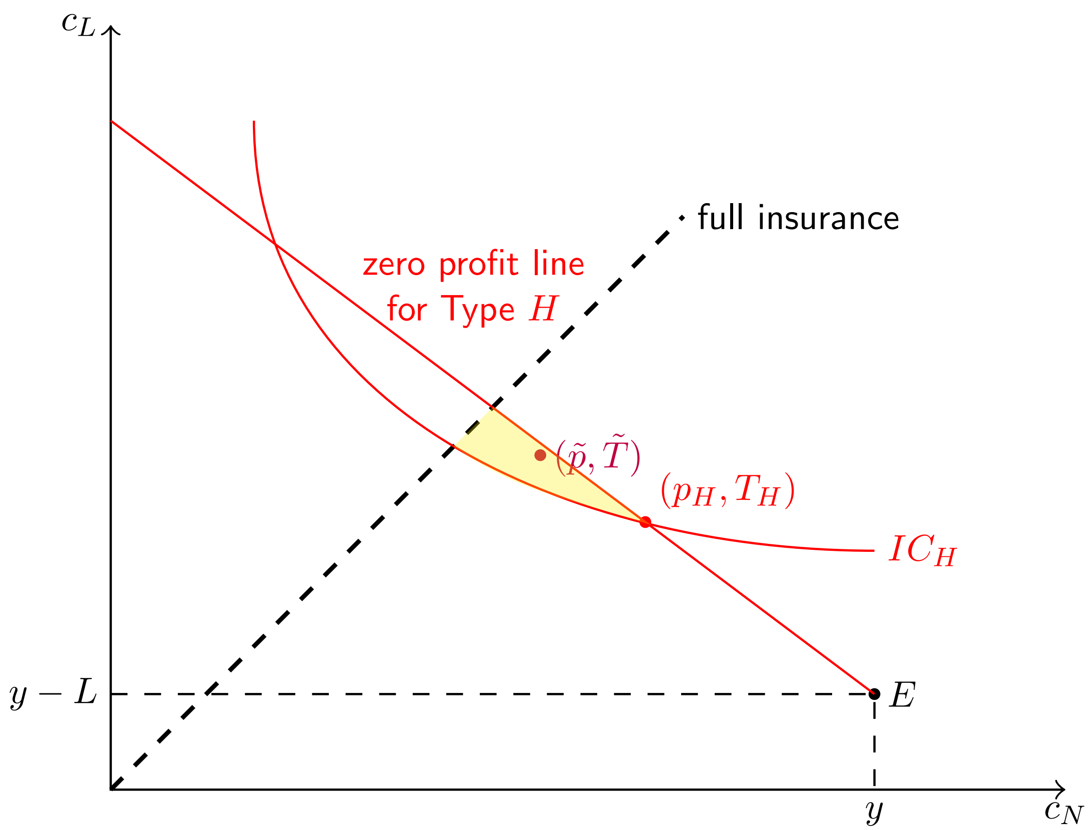

Figure 7: Lem 4. H type gets fully insured.

If $T_H \neq L$, then an insurance firm always wants to deviate this equilibrium by offering $(\tilde{p}, \tilde{T})$ which is more attractive to type H and gives the firm a positive profit.

Lemma 5 *If* (p_L, T_L) *is the contract chosen by low types in a separating equilibrium. Then* $p_L = \pi_L T_L$ *and* $T_L = \widehat{T}_L$ *where* $\widehat{T}_L$ *satisfies*

$$\pi_H u(y - \pi_L \widehat{T}_L - L + \widehat{T}_L) + (1 - \pi_H) u(y - \pi_L \widehat{T}_L) \\ = u(y - \pi_H L).$$

Proof: By Lemma 4 we have that $(p_H, T_H) = (\pi_H L, L)$.

By Lemma 3 we have that $p_L = \pi_L T_L$.

We know that high types weakly prefer (p_H, T_H) to (p_L, T_L) and thus

$$\pi_H u(y - \pi_L T_L - L + T_L) + (1 - \pi_H) u(y - \pi_L T_L) \\ \leq u(y - \pi_H L).$$

We just have to show the equality holds.

If not, the situation is as depicted in Fig 8.

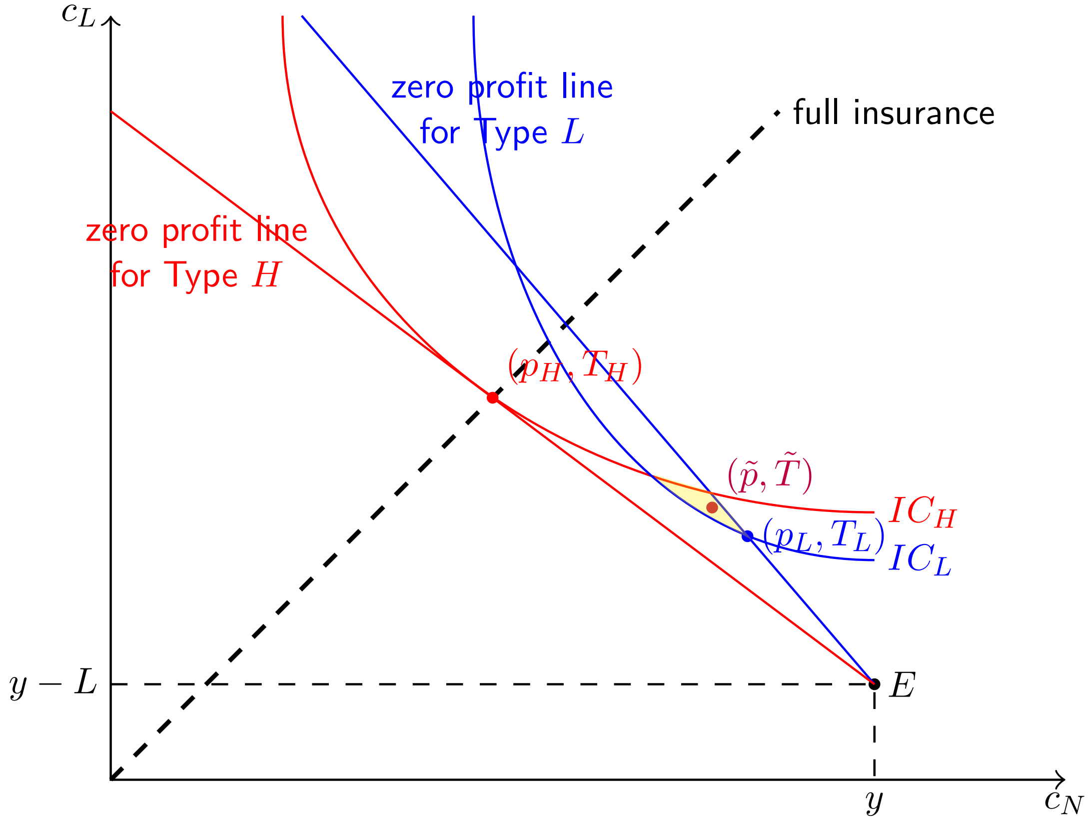

Figure 8: Lem 5. (p_L, T_L) must intersect the type H's indifference curve.

Suppose (p_L, T_L) is an equilibrium contract for type L. Then a firm wants to deviate from the equilibrium by offering $(\tilde{p}, \tilde{T})$ which attracts type L consumers and gives the firm a positive profit.

Then a firm can make positive profits by offering a contract such as $(\widetilde{p}, \widetilde{T})$ in the shaded region.

This contract attracts only low risk types and makes money on each consumer.

This is a contradiction. $\blacksquare$

Summarizing the results so far:

Proposition 2 *In any SPNE with unobservable types, type H consumers accept the contract $(p^*_H, T^*_H) = (\pi_H L, L)$, type L consumers accept the contract $(p^*_L, T^*_L) = (\pi_L \widehat{T}_L, \widehat{T}_L)$, and firms earn zero profits.*

Proposition 2 characterizes an SPNE with unobservable types *if it exists*.

But existence may be problematic.

Consider Fig 9.

Both types of consumers would prefer the pooling contract $(\widetilde{p}, L)$ to their equilibrium contracts.

Moreover, such a contract would make positive profits if all consumers took it.

Thus, one firm could deviate by offering the contract $(\widetilde{p}, L)$ and make positive profits.

In this case, there does not exist an SPNE with unobservable types.

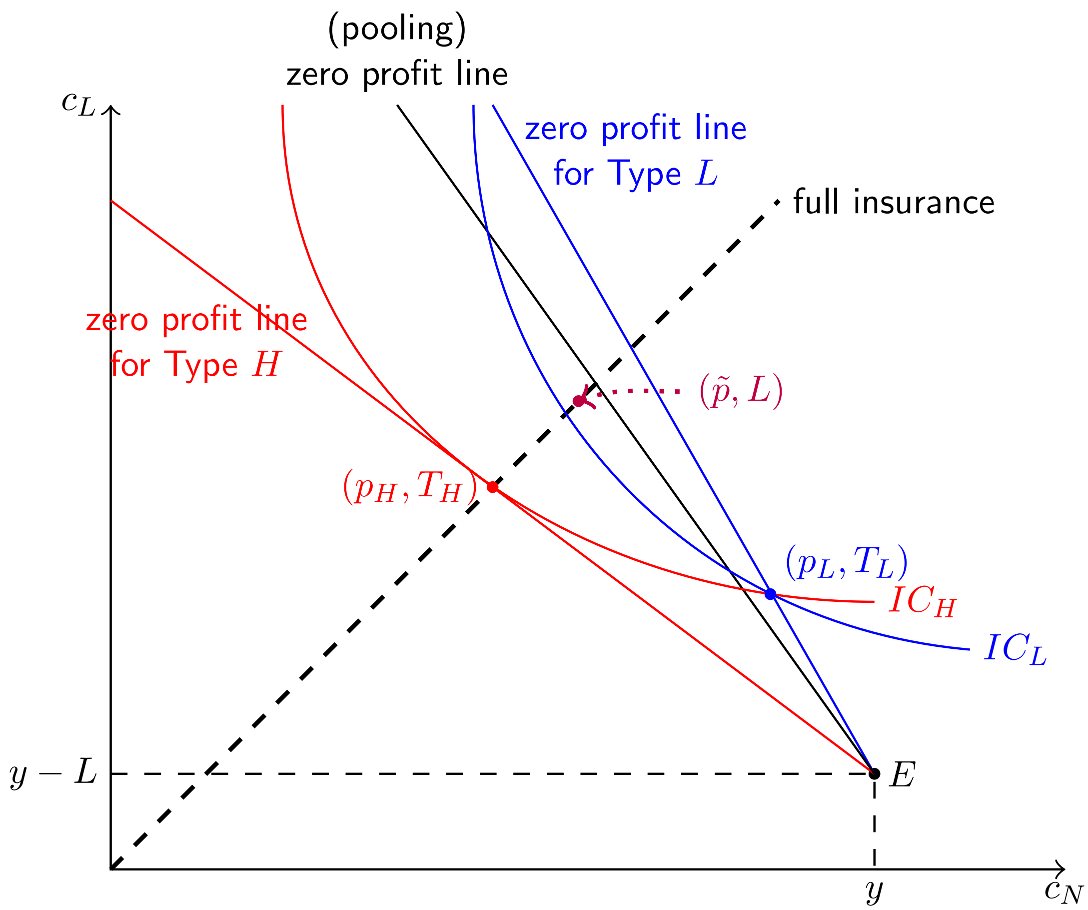

Figure 9: A case where the separating equilibrium does not exist.

Both types are better off and the insurance firm has a positive profit with the pooling contract $(\tilde{p}, L)$. However, the pooling equilibrium does not exist. This case happens when the proportion of type L consumers is large so that the pooling zero profit line is close to the zero profit line for type L.

Social insurance

If it exists, an SPNE with unobservable types leaves low risk consumers only partially insured, which is inefficient.

If equilibrium does not exist, then it is not clear what will happen: the model does not provide a prediction.

With a social insurance scheme, government could provide all individuals full insurance and finance it with a head tax $\overline{\pi}L$.

This would yield an efficient allocation.

However, when an SPNE exists, moving from the equilibrium to such a social insurance scheme will not make low types better off.

If it did, then the SPNE would not exist.

Empirical relevance

The idea that adverse selection is a real source of market failure in insurance receives empirical support.

Casual empiricism suggests it is hard to buy dental insurance, except through an employer.

There is also no market for unemployment insurance.

A natural test for adverse selection is "are those who buy more insurance more likely to suffer loss?"

This could be driven by moral hazard, but probably not in certain contexts such as death.

Finkelstein and Poterba (2004) provide evidence of adverse selection in the U.K. annuity market (annuities provide insurance against living too long).

ii) Samaritan's Dilemma

In the presence of altruism, individuals may not buy insurance even when it is available.

This is because they anticipate receiving charity from others if they suffer a loss.

Thus, for example, even if individuals could buy annuities when they retire, they may figure their relatives would take care of them if they lived too long and thus forego purchase.

More generally, the same logic implies that individuals may not save adequately for their old age (see, for example, Kotlikoff (1987) and Hansson and Stuart (1989)).

This is a historically important argument for social security.

iii) Individual optimization errors

It could also be the case that individuals would screw up their insurance decisions even if it were available and there was no charity.

This maybe because individuals have difficulty assessing probabilities or because they do not like having to think about unpleasant scenarios.

Similar logic suggests that people may undersave for retirement because they are myopic.

Such explanations become all the more plausible given the all evidence emerging from behavioral economics.

iv) Aggregate risk

Private insurance handles best situations of idiosyncratic risk where it can achieve cross-sectional pooling.

The premiums of the majority of fortunate policy-holders pay for the claims of the unlucky minority.

Private insurance does less well handling situations of aggregate risk in which entire groups are impacted by negative shocks.

Government may have an advantage in coordinating risk-sharing across large groups.

This is particularly the case for inter-generational risk sharing since the market is unable to facilitate risk-sharing between current generations and those not yet born.

Obviously, insurance contracts cannot be purchased by those not yet born.

Thus, if one generation experiences a severe recession or war, the private market cannot spread the burden of paying for these events across future generations.

The government through social security and debt can however achieve intergenerational risk-sharing (see Gordon and Varian (1988)).

2. Unemployment insurance

Most governments in developed countries provide workers with unemployment insurance.

Such schemes pay unemployed workers benefits for some period of time after they have become involuntarily unemployed.

These benefits are intended to tied them over while they search for a new job.

Typically, unemployment insurance benefits are financed by a tax paid by both workers and firms during employment.

From an economic perspective, unemployment insurance facilitates "consumption smoothing" and better job matches.

On the other hand, it may lead to less job search and higher unemployment.

It also may distort workers towards unstable jobs and lead to more on the job shirking.

There is a vast literature on unemployment insurance, both theoretical and empirical.

We will discuss optimal unemployment insurance.

Optimal level of benefits

We focus first on the optimal level of benefits.

The standard measure of the genorosity of a UI system is its *replacement rate*

$$r = \frac{net\ benefit}{net\ wage}$$

A UI system reduces a worker's gain from finding a new job to $w(1 - r)$

The replacement rate is determined by the level of UI benefits.

A high replacement rate facilitates consumption smoothing.

On the other hand, it creates *moral hazard* by making unemployment more attractive.

The optimal benefit level must trade off consumption smoothing and moral hazard.

The seminal analysis of the optimal level of UI benefits is Baily (1978).

He showed that the optimal benefit level can be expressed as a function of a small set of parameters in a static model.

His analysis was generalized by Chetty (2006).

Baily-Chetty Model

Consider a single worker who can be employed or unemployed.

The worker's income when employed is y_e and when unemployed is y_u.

The worker's wealth is W.

The worker is initially unemployed.

The probability the worker obtains employment is e where e denotes his job search effort.

The disutility of effort e is $\varphi(e)$ where $\varphi' > 0$ and $\varphi'' > 0$.

The UI system pays a benefit b to the worker if he is unemployed.

This benefit is financed by a tax t on the worker if he is employed.

The system must run an (expected) balanced budget so that

$$et = (1 - e)b.$$

The worker's utility from consumption is $u(c)$ where $u' > 0$ and $u'' < 0$.

The worker's expected utility is

$$eu(W + y_e - t) + (1 - e)u(W + y_u + b) - \varphi(e).$$

The first best

The first best problem is

$$\max_{(e,t,b)} eu(W + y_e - t) + (1-e)u(W + y_u + b) - \varphi(e)$$

$$s.t. \quad t = \frac{1-e}{e} b$$

This assumes that the insurance agency can observe and control the worker's search efforts.

Obviously, this is unrealistic, but it provides a benchmark for analysis.

The first order conditions imply that

$$b = e(y_e - y_u)$$

and

$$u'(W + ey_e + (1-e)y_u)(y_e - y_u) = \varphi'(e)$$

The solution involves perfect consumption smoothing; i.e., $c_e = c_u$ (consumption when employed equals consumption when unemployed)

The effort choice balances the disutility with the income gains.

The second best

Now assume that the insurance agency cannot observe the effort level of the worker.

We then get a moral hazard problem.

In particular, the agency cannot fully insure the worker, because he would have no incentive to undertake effort.

The worker's problem is

$$\max_e eu(W + y_e - t) + (1 - e)u(W + y_u + b) - \varphi(e)$$

Let $e(b, t)$ denote the worker's optimal effort choice and let $V(b, t)$ denote the worker's indirect utility.

The second best problem is

$$\begin{gathered} \max_{(b,t)} V(b, t) \\ s.t.\ t = \tfrac{1-e(b,t)}{e(b,t)} b \end{gathered}$$

There is a convenient formula for the optimal benefit level which we will now derive.

First solve the insurance agency's budget constraint for the tax as a function of the benefit; i.e., $t(b)$.

Then consider the unconstrained problem

$$\max_b V(b, t(b))$$

The first order condition for the optimal benefit level is

$$\frac{\partial V}{\partial b} = -\frac{\partial V}{\partial t}\frac{dt}{db}$$

Using the *Envelope Theorem*, we obtain

$$\frac{\partial V}{\partial b} = (1 - e(\cdot))u'(c_u)$$

and

$$\frac{\partial V}{\partial t} = -e(\cdot)u'(c_e)$$

Thus,

$$(1 - e(\cdot))u'(c_u) = e(\cdot)u'(c_e)\frac{dt}{db}.$$

Using the fact that

$$t(b) = \frac{1 - e(b, t(b))}{e(b, t(b))}b$$

we have that

$$\begin{aligned} \frac{dt}{db} &= \frac{1 - e(\cdot)}{e(\cdot)} - \frac{b}{e(\cdot)^2}\frac{de}{db} \\ &= \frac{1 - e(\cdot)}{e(\cdot)}(1 + \frac{\varepsilon}{e(\cdot)}) \end{aligned}$$

where ε is the elasticity of the probability of unemployment with respect to benefits; that is,

$$\varepsilon = -\frac{b}{1 - e(\cdot)}\frac{de}{db}.$$

Substituting the expression for dt/db back into the first order condition, we obtain

$$u'(c_u) = u'(c_e)(1 + \frac{\varepsilon}{e(\cdot)})$$

or

$$\frac{u'(c_u) - u'(c_e)}{u'(c_e)} = \frac{\varepsilon}{e(\cdot)}$$

We can write the marginal utility gap as

$$u'(c_u) - u'(c_e) \approx u''(c_e)(c_u - c_e)$$

Defining the *coefficient of relative risk aversion* as

$$\gamma(c) = -\frac{u''(c)c}{u'(c)}$$

we can further write

$$\begin{aligned} \frac{u'(c_u) - u'(c_e)}{u'(c_e)} &\approx -\frac{u''(c_e)(c_e - c_u)}{u'(c_e)} \\ &= \frac{\gamma(c_e)(c_e - c_u)}{c_e}. \end{aligned}$$

We have therefore proved:

Proposition *(Baily-Chetty Formula) The optimal unemployment benefit level satisfies*

$$\gamma(c_e)\frac{\Delta c}{c_e} \approx \frac{\varepsilon}{e(\cdot)},$$

where $\Delta c = c_e - c_u$ *is the drop in consumption during unemployment.*

Empirical work on UI can then be used to provide estimates of the three key variables $(\gamma(c_e), \frac{\Delta c}{c_e}, \varepsilon)$.

Optimal timing of benefits

The Baily-Chetty model sheds light on the determinants of the optimal replacement rate, but does not capture the dynamic features of UI.

In the U.S. unemployment insurance benefits are paid out at a constant level for 6 months and then reduced to zero.

In most countries, benefits are decreasing in duration, although in Sweden they increase after a while.

Intuitively, it is not obvious how they should be structured dynamically.

Shavell and Weiss (1979) *JPE* were the first to pose and analyze the optimal unemployment insurance problem from a dynamic perspective.

We follow the treatment in Ljungqvist and Sargent Chapter 21.

Shavell-Weiss Model

An unemployed worker orders stochastic processes of consumption and search effort $\{c_t, a_t\}_{t=0}^{\infty}$ according to

$$E \sum_{t=0}^{\infty} \beta^t [u(c_t) - a_t]$$

where $\beta \in (0, 1)$ and $u(\cdot)$ is increasing, smooth and strictly concave.

All jobs are alike and pay wage w per period forever.

If the worker exerts search effort a, he finds a job with probability $p(a)$.

The function $p(\cdot)$ is increasing, smooth, strictly concave, and satisfies $p(0) = 0$.

Once the worker finds a job, a is equal to 0 from then on and c is equal to w.

The worker has no savings and cannot borrow.

The insurance agency is the only source of consumption smoothing across time and states.

Autarky

Suppose the worker has no unemployment insurance - how would he behave?

Let V^e be the expected sum of discounted utility of an employed worker.

Clearly,

$$V^e = \frac{u(w)}{1-\beta}$$

Let V^u be the expected present value of utility for an unemployed worker. Then,

$$V^u = \max_{a \geq 0} \{u(0) - a + \beta[p(a)V^e + (1 - p(a))V^u]\}$$

The first order condition is

$$\beta p'(a)[V^e - V^u] \leq 1 \ \ (= \text{ if } a > 0).$$

If V_{aut} and $a_{aut} > 0$ denote the solutions to the autarky problem, we have that

$$V_{aut} = u(0) - a_{aut} + \beta[p(a_{aut})V^e + (1 - p(a_{aut}))V_{aut}].$$

and

$$\beta p'(a_{aut})[V^e - V_{aut}] = 1$$

These solutions are straightforward to compute....

The first best

Consider the unemployment insurance problem when the insurance agency can observe and control both the worker's consumption and his search effort.

Assume the insurance agency wants to design an unemployment insurance contract to give the worker discounted expected value $V > V_{aut}$.

The agency wants to provide V in the most efficient way; that is, to minimize expected discounted costs of providing V.

Let $C(V)$ denote the minimized expected discounted costs.

This will be a strictly convex function of V because of risk aversion.

Given V, the agency assigns a first period consumption-effort pair (c, a) and a promised continuation value V^u should the worker not find employment.

Thus, we think of V as a state variable and (c, a) as the policy functions.

In this way, we can pose the problem recursively.

The triple (c, a, V^u) will be functions of V and will satisfy the Bellman equation

$$C(V) = \min_{c,a,V^u}\{c + \beta(1 - p(a))C(V^u)\}$$
$$s.t.\ V = u(c) - a + \beta[p(a)V^e + (1 - p(a))V^u]$$

The constraint is known as the "promise-keeping" constraint - it guarantees that the unemployed worker's future payoff is V.

V^e - the payoff from being employed is as defined above.

Letting θ be the multiplier on the promise keeping constraint, the first order conditions for (c, a, V^u) are given by:

$$\theta = \frac{1}{u'(c)}$$

$$C(V^u) = \theta[\frac{1}{\beta p'(a)} - (V^e - V^u)]$$

$$C'(V^u) = \theta$$

Moreover, note that by the *Envelope Theorem*

$$C'(V) = \theta$$

The third and fourth condition together with the strict convexity of $C(\cdot)$ imply that $V^u = V$

Applying this repeatedly, we see that the worker's continuation value is held constant during his entire spell of unemployment.

Substituting in $V^u = V$, we see that c and a are constant during the entire unemployment spell.

Thus, the worker's consumption is "fully smoothed" during the unemployment spell.

But the worker's consumption is not smoothed across states of employment and unemployment assuming that $V < V^e$.

The second best

Assume that the agency cannot observe a, but it can still observe and control c.

The problem is now

$$C(V) = \min_{c,a,V^u}\{c + \beta(1-p(a))C(V^u)\}$$
$$s.t.\ V = u(c) - a + \beta[p(a)V^e + (1-p(a))V^u]$$
$$\beta p'(a)[V^e - V^u] \leq 1 \ \ (= \text{ if } a > 0).$$

The additional constraint is an "incentive constraint" which reflects the fact that the worker will choose his search effort optimally.

Assume that the function $C(V)$ continues to be strictly convex - this is now an assumption, which may not be satisfied.

Letting η be the multiplier on the incentive constraint, we have the first order conditions for (c, a, V^u) are given by:

$$\theta = \frac{1}{u'(c)}$$

$$\begin{aligned} C(V^u) &= \theta[\frac{1}{\beta p'(a)} - (V^e - V^u)] \\ &\quad -\eta\frac{p''(a)}{p'(a)}(V^e - V^u) \\ &= -\eta\frac{p''(a)}{p'(a)}(V^e - V^u) \end{aligned}$$

$$C'(V^u) = \theta - \eta\frac{p'(a)}{1 - p(a)}$$

Note that we have used the incentive constraint to simplify the second condition.

Moreover, it is still the case that by the *Envelope Theorem*

$$C'(V) = \theta$$

Since the second equation implies that $\eta > 0$, the third and fourth equation implies that

$$C'(V) > C'(V^u).$$

It follows that, under the assumed strict convexity, we must have that $V > V^u$ implying that the worker's continuation value is decreasing during his entire spell of unemployment.

Since V^u is decreasing through time, it follows that the worker's consumption is decreasing over his spell of unemployment and that his search effort is increasing.

The duration dependence of benefits is designed to provide incentives to search.

Example

Assume that a period is a week and set $\beta = 0.999$

Assume further that

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma}$$

and let $\sigma = 0.5$.

Let $w = 100$ and

$$p(a) = 1 - \exp(-ra)$$

where r is such that $p(a_{aut}) = 0.1$.

Figure 21.2.1 depicts the *replacement ratio* c/w and the search effort as a function of weeks of unemployment.

Further research

One weakness of this model is that it assumes that unemployed workers have no assets and can do no consumption smoothing of their own.

It may be reasonable to assume that unemployed workers cannot borrow, but it seems that they should be able to engage in self-insurance by accumulating a buffer stock of savings.

When they become unemployed, they will start decumulating these assets to smooth their consumption.

This may suggest that benefits should not fall with the duration of unemployment, because the longer a worker is unemployed the lower will be his asset levels.

For a formal exploration of this argument see "Asset Based Unemployment Insurance" *IER* 2012 by Pontus Rendahl.

He shows that benefits should be conditional on wealth.

As wealth levels fall over the duration of unemployment, benefit levels rise.

See also Shimer and Werning "Liquidity and Insurance for the Unemployed" *AER* 2008 for a related analysis in a different model.

On the empirical front, you might also want to read Chetty "Moral Hazard vs Liquidity in Unemployment Insurance" *JPE* 2008.

He is concerned with the classic question of whether unemployment insurance leads to longer spells of unemployment because of moral hazard.

There is certainly plenty of evidence that larger unemployment benefits lead to longer unemployment spells.

But he argues that this is coming primarily from a "liquidity effect" rather than a "moral hazard effect".

Under the moral hazard effect, people cut back job search efforts because higher benefits make the unemployed state more attractive.

Under the liquidity effect, higher benefits lead people to set a higher reservation wage because they are no longer so desperate to get cash.

The liquidity effect relies on the fact that people cannot perfectly smooth consumption due to imperfect credit markets.

Finally, there is a bunch of work on the distortions arising from the fact that the payroll tax which finances UI is only partially experience-rated.

That is, the payroll tax (which is levied on employers) rises as firms have more layoffs, but not on a one-for-one basis.

This partial experience rating subsidizes layoffs.

2. Social security

The U.S. Social Security program taxes workers to provide income support for the elderly.

It is a *pay-as-you-go* system as opposed to a *fully funded* system

Social security is the largest single source of income for the elderly.

2/3 of the elderly derive more than 1/2 of their income from Social security

For 1/5 of the elderly, social security is the only source of income.

To be eligible to receive Social security, individuals must have worked and paid payroll taxes for 10+ years and must be 62 or older.

Eligible individuals receive an annuity payment that lasts until death which is calculated as a function of average lifetime earnings.

To analyze the optimal level of social security benefits, one needs a theory of why the program is necessary in the first place.

Feldstein (1985) provides an analysis under the assumption that the program is necessary because some individuals are myopic and do not save enough.

On the other hand, social security reduces the savings of the non-myopic.

Since savings are productive and social security benefits are not, this creates a distortion.

The optimal level of benefits trades off the benefit of helping the myopic with the costs of distorting the savings of the non-myopic.

See also Krueger and Kubler (2005) *AER* who view the function of social security as spreading of aggregate risk across generations.

Hansson and Stuart (1989) *AER* take a Samaritan's Dilemma perspective.

Feldstein-style model

The following model is a simplified version of Feldstein's and draws on Andersen and Bhattacharya, *Economic Theory* 2010.

The model is infinite horizon, overlapping generations.

Periods are indexed by $t = 0, ..., \infty$.

Individuals work in the first period of their lives and retire in the second.

The size of the population is constant and normalized to 1.

Individuals have one unit of labor which they supply inelastically.

The wage rate is w and is exogenous.

Savings yield an exogenous return $\rho \geq 0$; i.e., a dollar saved in the first period of life, yields $1+\rho$ in the second.

The government operates a *pay-as-you-go* social security system.

It taxes labor earnings at rate θ and provides the aged a benefit b.

The government's budget must balance in each period, so that

$$b = \theta w.$$

The only decision individuals make is how much to save.

Individuals' true lifetime utility is

$$\ln c_1 + \ln c_2$$

where c_1 is consumption when young and c_2 is consumption when old.

There are two types of individuals: *myopes* and *life-cyclers*.

The fraction of myopes in the population is μ.

Myopes do no saving and life-cyclers save so as to maximize lifetime utility.

Life-cyclers solve the problem

$$\max_{s \geq 0} \ln\left(w(1-\theta) - s\right) + \ln(s(1+\rho) + b).$$

The solution is

$$s(\theta, b) = \begin{cases} 0 & \text{if } \frac{w(1-\theta)}{2} \leq \frac{b}{2(1+\rho)} \\ \frac{w(1-\theta)}{2} - \frac{b}{2(1+\rho)} & \text{otherwise} \end{cases}.$$

Observe that higher social security benefits reduces saving of life-cyclers

$$\frac{\partial s}{\partial b} = -\frac{1}{2(1+\rho)} < 0$$

If saving is productive (i.e., $\rho > 0$), this crowding out creates an aggregate cost for the economy, since social security is not productive.

Note also that when the government budget constraint is taken in to account, the condition for savings to be positive

$$\frac{w(1-\theta)}{2} > \frac{b}{2(1+\rho)},$$

is equivalent to

$$w\left(\frac{1+\rho}{2+\rho}\right) > b.$$

Letting $v(\theta, b)$ denote the life-cycler's indirect utility, by the Envelope Theorem, we have that

$$\frac{\partial v(\theta, b)}{\partial b} = \frac{1}{s(\theta, b)(1+\rho) + b}$$

and

$$\frac{\partial v(\theta, b)}{\partial \theta} = \frac{-w}{w(1-\theta) - s(\theta, b)}$$

Note that

$$\begin{aligned}
\frac{dv(\frac{b}{w}, b)}{db} &= \frac{1}{s(\frac{b}{w}, b)(1+\rho) + b} - \frac{1}{w - b - s(\frac{b}{w}, b)} \\
&= \frac{1}{(\frac{w(1+\rho)}{2} - \frac{\rho b}{2})} - \frac{1}{\left(\frac{w}{2} - \frac{\rho b}{2(1+\rho)}\right)} \\
&= -\frac{\rho}{(w(1+\rho) - \rho b)} < 0
\end{aligned}$$

Thus, higher social security benefits reduce the lifetime utility of life-cyclers assuming that $\rho > 0$.

Intuitively, this is because they crowd out productive savings.

Optimal social security benefits

Using the fact that the government budget constraint implies that $\theta = b/w$, we can write the government's problem as

$$\max_b \mu \left[\ln (w-b) + \ln b\right] + (1-\mu)v(\frac{b}{w}, b).$$

Using the expression for $dv(\frac{b}{w}, b)/db$ derived above, the first order condition is

$$\mu \left[\frac{1}{b} - \frac{1}{w-b}\right] - (1-\mu)\left[\frac{\rho}{(w(1+\rho)-\rho b)}\right] = 0$$

Using the fact that

$$\frac{1}{b} - \frac{1}{w-b} = \frac{w-2b}{b(w-b)},$$

the optimal benefit level is such that

$$\mu \frac{w-2b}{b(w-b)} = (1-\mu)\left[\frac{\rho}{(w(1+\rho)-\rho b)}\right].$$

This implies that the optimal benefit level satisfies

$$\mu\left(w-2b\right)\left(w(1+\rho)-\rho b\right)=\rho(1-\mu)b(w-b).$$

This is a quadratic equation that can be solved for a closed form solution (Problem Set?).

When everyone is a myope ($\mu = 1$), then $b = w/2$.

When everyone is a life-cycler ($\mu = 0$), then assuming $\rho > 0$, $b = 0$.

If $\rho = 0$, then there is no cost to the life-cyclers of raising benefits, so that $b = w/2$.

Assuming $\rho > 0$, then it is clear from the first order condition, that as we increase the fraction of myopes, the optimal benefits must increase.

The model therefore nicely illustrates the idea that the optimal level of social security trades off the benefit of helping the myopic with the costs of distorting the savings of the non-myopic.

Of course, the model does not explain why the government uses a pay-as-you-go scheme as opposed to a fully-funded scheme.

Under a fully-funded system, the government saves on behalf of the citizens and the costs of distorting savings are avoided.

3. Disability insurance

Golosov and Tysvinsky *JPE* 2006 provide a nice analysis of optimal disability insurance.

In the U.S., the government provides citizens with disability insurance through the social security system.

The social security disability insurance program provides income to a large number of citizens (6 million) and costs a huge amount of money ($61 billion in 2001).

G & T's paper first shows how a dynamic optimal social insurance mechanism can be implemented with a simple, realistic set of tax instruments - a so-called "asset-tested disability system".

They then parametrize their model and numerically evaluate features of the optimal system and the welfare gains from implementing their system (relative to a stylized version of the current US system).

We will focus on the first part of the paper - but the second part is also interesting.

The paper is part of the *New Dynamic Public Finance* which we have already discussed.

Golosov-Tysvinsky model

Consider an individual who lives for T periods and has preferences defined over lifetime consumption and labor represented by

$$E \sum_{t=1}^{T} \beta^{t-1}[u(c_t) + v(l_t)]$$

where E is the expectation operator, β the discount rate and c_t and l_t are consumption and labor in period t

u is increasing and strictly concave and v is decreasing. Moreover, $v(0) = 0$.

If the individual becomes disabled at time t his productive ability is then equal to zero and he cannot work.

Moreover, his ability stays at zero for the rest of his life.

If the individual is able his productive ability in period t is θ_t.

The sequence $\{\theta_1, ..., \theta_T\}$ is known at date 1.

Let

$$\pi_1 = \Pr(\text{able at } t = 1)$$

$$\pi_t = \Pr(\text{ able at } t | \text{ able at } t-1) \;\; t \geq 2$$

$$\begin{aligned} \Pi_{s,t} &= \Pr(\text{ able at } t | \text{ able at } s-1) \\ &= \pi_s \cdot \cdots \cdot \pi_t \end{aligned}$$

$$\Pi_t = \Pr(\text{able at } t) \;\; t \geq 1$$

$$\Pi_0 = 1$$

Let c_t denote consumption of the individual at age t if he is able and let l_t his labor supply.

Let x_t^s denote consumption of the individual at age t if he became disabled at $s \leq t$.

The government cannot observe whether the individual is disabled but can observe income.

The individual learns whether or not he is able at the beginning of each period.

The interest rate R is exogenous and satisfies $\beta = 1/(1+R)$.

The wage rate is exogenous and equal to w.

An allocation of consumption and labor (c, l, x) is *feasible* for the individual if the expected present value of consumption is less than the expected present value of output; that is,

$$
\begin{aligned}
& \sum_{t=1}^{T} \beta^{t-1} \Pi_t c_t + \sum_{s=1}^{T} \Pi_{s-1}(1-\pi_s) \sum_{t=s}^{T} \beta^{t-1} x_t^s \\
\leq \quad & \sum_{t=1}^{T} \beta^{t-1} \Pi_t w \theta_t l_t
\end{aligned}
$$

Allocations must also respect incentive compatability constraints because the age at which the individual becomes disabled is private information.

Incentive-compatability constraints require that in each period the expected utility of working is higher than the utility of claiming disability.

Thus, for all $s = 1, ..., T$

$$
\begin{aligned}
& u(c_s) + v(l_s) + \sum_{t=s+1}^{T} \beta^{t-s} \Pi_{s+1,t}[u(c_t) + v(l_t)] \\
& + \sum_{t=s+1}^{T} \Pi_{s+1,t-1}(1-\pi_t) \sum_{i=t}^{T} \beta^{i-s} u(x_i^t) \\
\geq \ & \sum_{t=s}^{T} \beta^{t-s} u(x_t^s)
\end{aligned}
$$

The social planner's problem is

$$\max_{(c,l,x)} \sum_{t=1}^{T} \beta^{t-1} \Pi_t [u(c_t) + v(l_t)]$$
$$+ \sum_{s=1}^{T} \Pi_{s-1}(1-\pi_s) \sum_{t=s}^{T} \beta^{t-1} u(x_t^s)$$

subject to the feasibility constraint and the incentive constraints.

Characterization of the optimum

The first best

If there are no incentive constraints, the optimal solution is such that for all t, s ($s \leq t$)

$$c_t = x_t^s = \overline{c}$$

and

$$wu'(c_t) = -v'(l_t)/\theta_t \ \text{ for all } t$$

Thus, the first best has full insurance (consumption smoothing) and the earnings of the individual when able optimally balance the marginal disutility of labor with the marginal benefit of consumption.

The second best

Result 1: *In the second best solution, the following properties are satisfied: (i) the feasibility constraint holds with equality; (ii) the incentive constraints all hold with equality; and (iii) in each period,* $c_t > x_t^t$.

These properties are all intuitive.

Result 2: *In the second best solution,*

$$wu'(c_t) = -v'(l_t)/\theta_t \quad \textit{for all } t.$$

Thus, the consumption-labor margin is undistorted when the individual is able.

This is analogous to the result from the two type Mirrlees model concerning the labor supply of high types.

Result 3: *In the second best solution, for each period* $t < T$

$$\frac{1}{u'(c_t)} = [\frac{\pi_{t+1}}{u'(c_{t+1})} + \frac{1-\pi_{t+1}}{u'(x_{t+1}^{t+1})}]$$

This is the *Inverse Euler Equation* for the problem.

It has the following important implication:

Result 4: *In the second best solution, for each period* $t < T$ *such that* $\pi_{t+1} \in (0,1)$

$$u'(c_t) < \pi_{t+1}u'(c_{t+1}) + (1-\pi_{t+1})u'(x_{t+1}^{t+1})$$

This result follows from Jensen's Inequality and Result 3.

It implies that the allocation of consumption across time is distorted for the able.

In particular, the marginal utility of consumption in the current period is lower than the expected marginal utility of consumption in the next period.

This means that the individual is consuming too much in the current period relative to the first best or, equivalently, he is saving less than the first best amount.

This intertemporal distortion of savings is caused by the incentive constraints.

If the individual saves more from period t to $t+1$ then because the marginal utility of consumption is decreasing, it is harder to prevent him form masquerading as disabled.

Accordingly, it is desirable to deter savings by taxation or even asset limits.

We already mentioned this point when we discussed the New Dynamic Public Finance.

Result 5: *In the second best solution, for all* s, t *and* t'

$$x_t^s = x_{t'}^s$$

Thus, consumption is constant once the individual becomes disabled.

This is because all uncertainty is resolved once the individual becomes disabled.

Implementing the optimum

How can the government decentralize the second best optimal allocation?

Golosov and Tysvinski show that the government can use what they term an *asset-tested disability insurance system*.

The system has two key features:

(i) disability transfers depend upon the length of pre-disability work history.

(ii) disability transfers are paid only to individuals whose assets are below a pre-specified minimum.

Formally, the implementation problem is to design a system of taxes and disability transfers so that competitive equilibrium allocations are second best.

First we need to define what is a competitive equilibrium given a tax/transfer system?

Definition 1: *Given a tax system* $\{\tau_t(\cdot)\}$ *allocations of consumption, labor supply, and savings* $(\widetilde{c}, \widetilde{l}, \widetilde{x}, \widetilde{k})$ *constitute a competitive equilibrium if they solve the following problem:*

$$\max_{(c,l,x,k)} \sum_{t=1}^{T} \beta^{t-1} \Pi_t [u(c_t) + v(l_t)]$$
$$+ \sum_{s=1}^{T} \Pi_{s-1}(1-\pi_s) \sum_{t=s}^{T} \beta^{t-1} u(x_t^s)$$
$$s.t. \; \forall t \;\; c_t + k_t \leq$$
$$w\theta_t l_t + (1+R)k_{t-1} + \tau_t(\{\theta_i l_i, k_{i-1}\}_{i=1}^{t})$$
$$\forall t \geq s \qquad x_t^s + k_t^s \leq (1+R)k_{t-1}^s +$$
$$\tau_t((\{\theta_i l_i\}_{i=1}^{s-1}, \{\theta_i l_i = 0\}_{i=s}^{t}), (\{k_{i-1}\}_{i=1}^{s}, \{k_i^s\}_{i=s}^{T}))$$

where $k_{s-1}^s = k_{s-1}$ *and feasibility is satisfied.*

Then they say that the tax system $\{\tau_t(\cdot)\}$ *implements* the second best optimal allocation (c^*, l^*, x^*) if (c^*, l^*, x^*) is equal to the competitive equilibrium allocation just defined.

The formal definition of an asset-tested disability insurance system is as follows:

Definition 2: *An asset-tested disability insurance system* $(\overline{k}, S, T_a)$ *consists of: (i) a sequence of asset tests* $\overline{k}(i)$ $i = 1, ..., T$; *(ii) a sequence of lump-sum transfers of the form* $S_d(t, i) = T_d(i) - w\theta_t l_t$, $1 \leq i \leq t \leq T$, *where* $S_d(t, i)$ *is the transfer received in period* t *by a consumer who became newly disabled in period* i *with assets not exceeding* $\overline{k}(i)$*; and (iii) a lump sum tax* T_a *that is paid each period by a consumer who is still working or who had assets exceeding* $\overline{k}(i)$ *when he declared disability.*

Theorem: *For any second best optimal allocation* (c^*, l^*, x^*), *there exists an asset-tested disability insurance system* $(\overline{k}, S, T_a)$ *that implements it.*

Part (i) describes the asset tests - basically, if the individual becomes disabled in period i he must have assets below $\overline{k}(i)$ in order to get disability benefits in the future.

Part (ii) describes the disability benefits.

If the individual declares disability in period i and then earns in period $t \geq i$ then his earnings are subject to a 100% tax.

So, obviously, under this scheme, individuals who have declared disability will not work again (even if they could).

The disability benefit an individual receives depends upon the date at which he first stopped earning.

Part (iii) describes the tax that the able pay.

This tax is also paid by those who are disabled whose assets exceed the limit.

Under this system, individuals choose to save only up to the assets limit.

Saving more will eliminate their ability to benefit from the disability insurance.

When they become disabled, they claim disability and receive the transfer $T_d(i)$ in each period for the rest of their lives.

They also supplement their transfer by decumulating their assets.

The asset limits are key to implementing the second best optimal allocation.

Golosov and Tysvinski show in a two period example that the second best optimal allocation cannot be implemented with a linear tax on savings.

Again, we discussed this issue when we discussed the New Dynamic Public Finance.

VII. State and Local Public Finance

Around 50% of government spending in the US is accounted for by state and local governments.

State and local governments provide a host of services, including education, police, parks, and roads.

State and local governments differ from national governments because citizens can move freely between states and localities.

This raises a host of interesting issues.

These lectures will introduce a number of basic ideas from the state and local literature.

We will begin by introducing the classic paper of Tiebout *JPE* (1956) and the literature the paper inspired.

Next we will discuss property taxation and zoning.

After that we will discuss the distributional implications of local service provision.

Finally, we will discuss the idea of capitalization.

1) The Tiebout model

The basic normative issue in state and local public finance is should we have local governments and, if so, what should they provide.

Tiebout argued that a system of local governments providing local public goods financed by local taxes has an efficiency advantage over a more centralized system.

A *local public good* is one which benefits only the local community rather than the community as a whole.

For example, local radio vs national radio; police vs national defense; local hospitals vs cancer research.

Tiebout argued that local public goods differed from public goods in an important way.

Namely, that if they were provided via a system of competing local governments, citizens could "vote with their feet".

As Tiebout saw it, the fundamental problem with public goods was that government would be very unlikely to choose optimal levels.

This is because they would have no way of knowing what peoples' preferences were.

However, with local public goods and local governments, local competition would lead to optimal choices because people could "vote with their feet".

If a citizen-voter does not like the level of national defense spending in a community, there is not much he can do to express his displeasure.

He can vote out the incumbent government, but this is a crude tool because there are so many issues on which to base his vote.

If he does not like the level of education spending in his community, he can vote with his feet and move.

This leads to two conclusions: (i) people can sort into communities where the public good - tax mix reflects their preferences; and (ii) the fact that people can move disciplines the behavior of local governments.

We can illustrate the idea diagrammatically.

Tiebout argued informally that under the following seven assumptions, this system of competition would lead to efficient provision:

Assumption 1: no moving costs

Assumption 2: full information

Assumption 3: large number of communities

Assumption 4: no locational constraints created by employment

Assumption 5: no spillovers between communities

Assumption 6: average cost of providing public services as a function of population is U-shaped - implying there is a cost minimizing population size

Assumption 7: communities with population sizes below (above) the cost minimum seek to expand (contract).

Under these assumptions, efficient provision of public services would prevail, Tiebout asserted.

All the communities would be optimally sized, and citizen-voters would live in communities that provided public good-tax bundles that were optimal for them.

Tiebout argued that these assumptions were not much more extreme than those underlying perfect competition in private goods markets (no frictions, full information, no monopoly, etc).

Tiebout's argument is important because the forces that he identifies are clearly realistic.

Reaction to Tiebout

Tiebout's argument was far from rigorous.

It was not clear precisely what he was assuming about most of the main ingredients of the argument.

However, the importance of the underlying idea spawned a large literature trying to make it more precise and understand exactly the assumptions that were necessary for the system of local competition that he identified to actually achieve efficient provision of local public goods.

The models used in this literature are so-called *Tiebout models*.

They are static Arrow-Debreu general equilibrium models which in addition have distinct regions of habitation.

Each region has a government which provides local public goods and collects taxes to pay for them.

There is perfect consumer mobility between regions.

Consumers are fully informed about prices, taxes and public services in each region and choose to live in the region that provides them with the highest utility.

Governments are assumed to have varying objectives - catering to the median voter; maximizing property values; maximizing tax surplus; etc.

An *allocation* in a Tiebout model specifies: a consumption bundle for each consumer; a production plan for each firm; a public good bundle for each regional government; and a region of habitation for each consumer.

A *Tiebout equilibrium* consists of an allocation; a price for each commodity; and a tax system for each region such that:

(i) consumers are choosing their consumption bundles optimally

(ii) consumers are choosing their regions of habitation optimally (taking as given taxes and public good levels)

(iii) firms maximize profits

(iv) markets clear

(v) each regional government balances its budget

(vi) each regional government's tax and public goods plan maximizes its objective (whatever that is)

Study of these models led researchers to realize that Tiebout's model of competing local governments does not have the same efficiency properties as market competition except under very restrictive assumptions.

Bewley (1981)

A famous paper which points this out is Bewley (1981) who provides a long list of counter-examples to Tiebout's claims.

He then provides some assumptions under which Tiebout's argument is correct.

Lets look at some of his examples to see the sort of issues that come up

Example 1: (Economies of Scale)

There are 2 consumers and 2 regions.

There is a single public good g and labor l.

Each consumer is endowed with a unit of labor and cares only about his public good consumption.

To produce 1 unit of the public good requires 1 unit of labor; thus, in each region j

$$g_j = l_j$$

where g_j is the level of the public good in region j and l_j is the labor used.

The following situation describes a Tiebout equilibrium.

The price of the public good and of labor is 1.

There is a single consumer in each region.

The tax in each region is equal to 1 and the government uses the tax revenues to provide 1 unit of the public good.

In equilibrium each consumer is indifferent between living in either region.

Each government is using its revenues optimally, given its citizen's tastes.

This equilibrium is inefficient because both consumers would be better off if they lived in the same region.

The fundamental reason for the inefficiency is that people do not take into account the economies of scale that result when they move into a region.

Example 2: (Mismatched Consumers)

There are 4 consumers (A, B, C, D) and 2 regions.

There are four types of public services (g_A, g_B, g_C, g_D) and labor l.

Each consumer is endowed with 1 unit of labor

The public services are produced using labor and the production technology is such that

$$n_j(g_{Aj} + g_{Bj} + g_{Cj} + g_{Dj}) = 2l_j$$

where g_{kj} is the amount of service k provided in region j, n_j is the number of consumers in region j and l_j is the quantity of labor used in region j.

Note here that the resources used to produce the public service are increasing in the number of users.

This illustrates the important distinction between *public services* and *public goods*.

The consumers' utilities are

$$\begin{aligned} u_A &= 2g_{Aj} + g_{Bj} \\ u_B &= g_{Aj} + 2g_{Bj} \\ u_C &= 2g_{Cj} + g_{Dj} \\ u_D &= g_{Cj} + 2g_{Dj} \end{aligned}$$

The following is a Tiebout equilibrium. Consumers A and C live in region 1 and Consumers B and D live in region 2

$(g_{A1}, g_{B1}, g_{C1}, g_{D1}) = (1, 0, 1, 0)$ and
$(g_{A2}, g_{B2}, g_{C2}, g_{D2}) = (0, 1, 0, 1)$

The price of each type of public service and of labor is 1. The tax in each region is equal to 1.

Note that no consumer wishes to move.

The governments are choosing taxes and public services optimally given their citizens' preferences.

However, this equilibrium is not efficient.

Consider the following alternative allocation.

Consumers A and B live in region 1 and Consumers C and D live in region 2

$(g_{A1}, g_{B1}, g_{C1}, g_{D1}) = (1, 1, 0, 0)$ and
$(g_{A2}, g_{B2}, g_{C2}, g_{D2}) = (0, 0, 1, 1)$

The price of each type of public service and of labor is 1.

The tax in each region is equal to 1.

The difficulty is that consumers are mismatched and cannot sort themselves out by migration alone.

Bewley notes that the inefficiencies exhibited in these two examples would not arise if local governments set public good levels anticipating the migration changes they would induce.

But even with those governments, problems still arise.

If local governments do set public good levels anticipating migration, then they must be given an objective function.

What is plausible? Population maximization? Property value maximization? Profit maximization?

Example 3: (Anonymous taxation)

There are 2 consumers (A and B) and 2 regions.

There is a single public good g and labor l.

Each consumer is endowed with a unit of labor.

To produce 1 unit of the public good requires 1 unit of labor; thus, in each region j

$$g_j = l_j$$

where g_j is the level of the public good in region j and l_j is the labor used.

The consumers' utilities are

$$\begin{aligned} u_A &= g \\ u_B &= g - 3l \end{aligned}$$

The following situation describes a Tiebout equilibrium.

A lives in Region 1 and B lives in Region 2.

The tax in Region 1 is equal to 1 and the government uses the tax revenues to provide 1 unit of the public good.

The tax in Region 2 is equal to 0 and the government provides no public good.

However, the situation is clearly inefficient, because B would be better off if he lived in region 1 and payed no taxes.

But there is no way to achieve efficiency even if the governments could anticipate population changes.

Suppose that one government set a tax τ and provided 2τ units of the public good hoping to attract both consumers.

Then A would require that $\tau > 1/2$ and B would require that $2\tau - 3\tau > 0$.

Hence there is no way of getting the consumers to live together, given that they have to pay the same taxes.

Sufficient Conditions

Bewley goes on and on, in similar vein.

He then concludes his analysis on a more positive note by identifying some assumptions under which Tiebout's argument works.

His sufficient conditions are:

(i) Public services as opposed to public goods.

(ii) As many regions as there are consumer types.

(iii) Profit maximizing anticipatory governments but must all make zero profits in equilibrium.

(iv) Free trade between regions.

Under these conditions, a Tiebout equilibrium exists and it is efficient.

But Bewley, argues - this is scarely convincing since it does not seem a remotely sensible model.

2) Property Taxation and Zoning

The primary source of revenue for U.S. local governments has historically been the property tax.

Given this, the efficiency properties of a system of local governments financing public services via property taxes have long been of interest.

The literature identifies two sources of distortions from property taxation.

First, property taxation will distort housing choices, leading households to consume too little housing.

Second, property taxation will distort public service choices because households do not face the true tax price for services.

Hamilton (1975) argued that if local governments can implement zoning ordinances specifying minimum housing qualities, these distortions will not arise.

Property taxes will be "benefit taxes" - non-distortionary user fees for public services.

Coate (2010) argues that while this is in principle possible, local governments are unlikely to choose zoning ordinances in the right way.

Property taxes may indeed be benefit taxes, but zoning ordinances will distort housing choices, leading households to consume too much housing.

This lecture will illustrate these ideas in the dynamic Tiebout model developed by Coate (2010).

Coate's model

Consider a geographic area consisting of two communities, indexed by $i \in \{1, 2\}$.

Time is infinite, with periods indexed by $t \in \{0, ..., \infty\}$.

A population of size 1 resides in the area, but in each period new households arrive and old ones leave.

A household residing in the area in any period will need to remain there in the subsequent period with probability μ.

The only way to live in the area is to own a house in one of the communities.

Housing

Houses come in two types: large and small.

Houses are durable, but a fraction $d \in (0, 1 - \mu)$ of the stock in each community is destroyed at the end of each period.

New houses can be built and the cost of building a house of type $H \in \{L, S\}$ is C_H where $C_L > C_S$.

Each community has enough land to accomodate a population of size 1 and land has no alternate use.

The stock of old houses of type H in community i at the beginning of a period is O_{Hi} and new construction is N_{Hi}.

New construction is completed at the beginning of each period and new and old houses are perfect substitutes.

Public services

A public service is provided in each community.

The service level in community i is g_i.

The cost is cg_i per household.

Households

Each household receives a per-period income y.

When living in the area, households have preferences defined over housing, services, and consumption.

Households differ in their preferences for large houses, measured by θ.

A type θ household with consumption x and services g obtains a period payoff of $x + B(g) + \theta$ if it lives in a large house and $x + B(g)$ otherwise.

When not living in the area, a household's payoff just depends on its consumption.

Households discount future payoffs at rate δ and can borrow and save at rate $1/\delta - 1$.

Housing markets

Housing markets open at the beginning of each period.

Demand comes from new households moving into the area and remaining residents who need new houses or who want to move.

Supply comes from owners leaving the area, residents who want to move, and new construction.

New construction is supplied by competitive construction firms.

The price of houses of type H in community i is P_{Hi}.

Public finance

Service provision in community i is financed by a tax τ_i on the value of property $\sum_H P_{Hi}(O_{Hi} + N_{Hi})$.

Community i must balance its budget in each period implying that

$$\tau_i \sum_H P_{Hi}(O_{Hi} + N_{Hi}) = cg_i \sum_H (O_{Hi} + N_{Hi}).$$

The majority-preferred service level is implemented.

Timing

Each period begins with a stock of old houses $\mathbf{O} = (O_{L1}, O_{S1}, O_{L2}, O_{S2})$ of aggregate size $1 - d$.

Existing residents learn whether they will remain in the area and new households arrive.

Housing markets open, prices $\mathbf{P} = (P_{L1}, P_{S1}, P_{L2}, P_{S2})$ are determined, and new construction

$$\mathbf{N} = (N_{L1}, N_{S1}, N_{L2}, N_{S2})$$

takes place. Aggregate new construction is d.

Knowing $\mathbf{P}$ and $\mathbf{O} + \mathbf{N}$, residents choose services g_1 and g_2.

At the end of the period, a fraction d of the housing stock is destroyed so next period's stock is $\mathbf{O}' = (1 - d)(\mathbf{O} + \mathbf{N})$.

Equilibrium

The model has a recursive structure with the stock of old houses the aggregate state variable.

An *equilibrium* consists of a price rule $\mathbf{P}(\mathbf{O})$, a new construction rule $\mathbf{N}(\mathbf{O})$, public service rules $(g_1(\mathbf{O}), g_2(\mathbf{O}))$, and, for each household type, a value function and a housing demand function, such that three conditions are satisfied.

First, *household optimization*.

Second, *housing market equilibrium*.

Third, *majority rule*.

Household optimization

Households only have one decision to make.

If they need to reside in the area, they must choose what type of house to buy (large or small) and in which community (1 or 2).

They will take into account their preference for large houses as measured by θ, the current and future price of houses, the taxes they will pay, and the level of services that are provided.

Housing market equilibrium

Since new construction is supplied by competitive construction firms and land is free, the supply of type H houses is perfectly elastic at price C_H.

The price of type H houses in community i will equal C_H if new construction takes place, but can fall below C_H if there is no new construction.

Prices must be such that households buying type H houses are indifferent between communities.

Majority rule

The preferred service level for residents of type H houses in community i is

$$g^*(\rho_{Hi}(\mathbf{O})) = \arg\max_g \{B(g) - \rho_{Hi}(\mathbf{O})g\},$$

where $\rho_{Hi}(\mathbf{O})$ is the *tax price* of services faced by these residents.

This tax price is

$$\rho_{Hi}(\mathbf{O}) = \frac{cP_{Hi}(\mathbf{O})}{P_{Li}(\mathbf{O})\lambda_i(\mathbf{O}) + P_{Si}(\mathbf{O})(1 - \lambda_i(\mathbf{O}))},$$

where $\lambda_i(\mathbf{O})$ is the fraction of post-construction houses that are large.

The majority-preferred level is

$$g_i(\mathbf{O}) = \begin{cases} g^*(\rho_{Li}(\mathbf{O})) & \text{if } \lambda_i(\mathbf{O}) \geq 1/2 \\ g^*(\rho_{Si}(\mathbf{O})) & \text{if } \lambda_i(\mathbf{O}) < 1/2 \end{cases}.$$

Equilibrium steady states

Given an equilibrium, a stock of old houses is a *steady state* if new construction is such as to maintain the stock constant.

Proposition 1 *If O^* is a steady state, the fraction of large houses in each community is the same; that is, $\lambda_1(O^*) = \lambda_2(O^*) = \lambda^*$. If $\lambda^* \geq 1/2$, the public service level in each community is $g_L^* \equiv g^*(\frac{cC_L}{C_L\lambda^*+C_S(1-\lambda^*)})$ and households live in large houses if*

$$\theta \geq (1-\delta(1-d))(C_L - C_S) + \frac{c(C_L - C_S)}{C_L\lambda^* + C_S(1-\lambda^*)} g_L^*.$$

If $\lambda^* < 1/2$, the public service level is $g_S^* \equiv g^*(\frac{cC_S}{C_L\lambda^*+C_S(1-\lambda^*)})$ and households live in large houses if

$$\theta \geq (1-\delta(1-d))(C_L - C_S) + \frac{c(C_L - C_S)}{C_L\lambda^* + C_S(1-\lambda^*)} g_S^*.$$

Sketch of Proof: If $\mathbf{O}^*$ is an equilibrium steady state, then, $\mathbf{N}(\mathbf{O}^*) = d\mathbf{O}^*/(1-d)$.

Since $O^*_{Hi} > 0$ for all Hi, it must be the case that there is new construction of both types of houses in both communities.

Given perfectly elastic supply, housing prices must equal construction costs so that $\mathbf{P}(\mathbf{O}^*) = (C_L, C_S, C_L, C_S)$.

It must also be the case that the fraction of large houses in each community is the same; that is, $\lambda_1(\mathbf{O}^*) = \lambda_2(\mathbf{O}^*) = \lambda^*$.

For if, say, community 1 had a greater fraction of large houses, the public service surplus enjoyed by large house owners in community 1 would be higher than in community 2.

Since the price of houses is the same, no-one would buy a large house in community 2.

Since both house prices and the fraction of large houses are the same across the two communities, the service levels and taxes are also the same.

If a majority of households own large houses ($\lambda^* \geq 1/2$), then, the public service level will be g_L^* and households live in large houses only if their preference exceeds the stated expression.

If a majority of households own small houses ($\lambda^* < 1/2$), the public service level is g_S^* and households live in large houses only if their preference exceeds the stated expression. ■

Property tax distortions

The equilibrium steady states are inefficient.

From an efficiency perspective, households should live in large houses if

$$\theta \geq \theta^e \equiv (1 - \delta(1 - d))(C_L - C_S).$$

The steady state fraction of large houses is therefore *too low*.

The efficient level of public services is

$$g^e \equiv g^*(c) = \arg\max \{B(g) - cg\}.$$

Public services are therefore *under-provided* if large home owners are a majority and *over-provided* otherwise.

Zoning

Suppose from period 0 onwards, one community enforces a zoning ordinance requiring all newly constructed houses to be large.

Proposition 2 *In steady state, all houses in the zoned community are large and all houses in the unzoned community are small. The public service level in each community is g^e and households live in the zoned community only if $\theta \geq \theta^e$.*

This is Hamilton's argument at work.

Sketch of Proof: Suppose that community 1 is the zoned community.

If $\mathbf{O}^*$ is a steady state, then, under zoning, it must be the case that $O^*_{S1} = 0$ and hence $\lambda_1(\mathbf{O}^*) = 1$.

It must also be the case that $O^*_{L2} = 0$ and hence that $\lambda_2(\mathbf{O}^*) = 0$.

To see why, suppose, to the contrary, that $O^*_{L2} > 0$.

Then it must be the case that the steady state price of large houses in both communities is C_L.

Since the price of small houses in community 2 is C_S, the tax price of public services is lower for large house owners in community 1.

But this means public service surplus enjoyed by large house owners in community 1 is higher than in community 2 which would mean no large house owner would buy in community 2.

Since $P_{L1}(\mathbf{O}^*) = C_L$ and $\lambda_1(\mathbf{O}^*) = 1$ and $P_{S1}(\mathbf{O}^*) = C_S$ and $\lambda_2(\mathbf{O}^*) = 0$ it follows that

$$(g_1(\mathbf{O}^*), \tau_1(\mathbf{O}^*)) = (g^*(c), \frac{cg^*(c)}{C_L})$$

and that

$$(g_2(\mathbf{O}^*), \tau_2(\mathbf{O}^*)) = (g^*(c), \frac{cg^*(c)}{C_S}).$$

It follows that a household of type θ will prefer living in a large house in community 1 to a small house in community 2 if

$$\theta + B(g^*(c)) - \tfrac{cg^*(c)}{C_L} C_L - C_L + \delta(1-d)C_L$$
$$\geq B(g^*(c)) - \tfrac{cg^*(c)}{C_S} C_S - C_S + \delta(1-d)C_S$$

or, equivalently, if their preference θ exceeds θ^e. ■

Benefit taxation

Observe that in the zoning steady state, the property taxes paid by households in both communities equal $cg^*(c)$.

Each household therefore pays a tax equal to the cost of the services it consumes.

Property taxes are therefore *benefit taxes*.

Endogenous zoning

The above analysis assumes that one community has zoning.

In reality zoning decisions are made by the households who reside in the community.

If communities could choose whether to implement zoning, would the efficient outcome result?

To endogenize zoning, suppose at the end of each period, residents vote whether to impose a zoning ordinance for the next period.

The vote takes place before housing stock is destroyed, so that all voters own houses.

Community i's zoning regulation is $Z_i \in \{0, 1\}$.

The state variables are now $\mathbf{O}$ and the zoning regulations $\mathbf{Z} = (Z_1, Z_2)$ chosen by residents in the prior period.

An equilibrium now includes a description of the zoning rules $\mathbf{Z}'(\mathbf{O}, \mathbf{Z})$.

When voting, residents anticipate how zoning changes housing prices, tax prices, and service levels in the next period and beyond.

A household owning a type H house in community i who expects to continue owning the same type of house next period will favor zoning if

$$\begin{gathered}(1-\mu-d)[P_{Hi}^1 - P_{Hi}^0] \\ +\mu\left[(B(g_i^1) - \rho_{Hi}^1 g_i^1) - (B(g_i^0) - \rho_{Hi}^0 g_i^0)\right] \\ +\delta\mu\Delta V > 0\end{gathered}$$

where P_{Hi}^1 is price with zoning and P_{Hi}^0 is price without, etc....and ΔV denotes the difference in continuation utility.

Note that since $d < 1-\mu$, households benefit from an increase in the value of their homes.

Given an equilibrium, $(\mathbf{O}^*, \mathbf{Z}^*)$ is a *steady state* if new construction is such as to maintain the stock of old houses at $\mathbf{O}^*$ and if citizens maintain the zoning rules $\mathbf{Z}^*$.

Steady state $(\mathbf{O}^*, \mathbf{Z}^*)$ is *efficient* if *either* $\mathbf{Z}^*$ equals $(1, 0)$ and $\mathbf{O}^*$ equals $(1-d)(1-F(\theta^e), 0, 0, F(\theta^e))$, *or* the reverse.

In an equilibrium with endogenous zoning, do old housing stocks and zoning rules converge to an efficient steady state?

Coate shows that there exists no equilibrium with endogenous zoning which has a steady state that is both efficient and satisfies a local stability property.

Sketch of proof: At an efficient steady state, one community has no zoning and consists of all small houses, the other has zoning and consists of all large houses.

If the small house community deviated by imposing zoning, the price of small houses would rise.

Public service surplus in the deviating community could only increase because prior to the deviation all houses are small.

Thus, the short run impact of the deviation is positive, if households benefit from higher prices for their houses.

This is the case if $d < 1 - \mu$.

The stability property rules out harmful long run effects of the deviation. ■

What does happen when zoning decisions are endogenous?

Coate provides examples in which equilibrium involves both communities always imposing zoning.

In these equilibria, the steady state is that all houses are large.

In steady state, property taxes are benefit taxes because all houses have the same price, but the housing choice is distorted upwards.

Thus the distortion is the opposite of that arising with property taxes and no zoning.

3) Distributional Implications of Local Service Provision

While local government provision of services may allow for efficient sorting, it may also have negative distributional consequences.

This is particularly the case for local provision of K-12 education.

The concern is that local communities will stratify according to income, and higher income communities will provide better quality schooling for their children.

This inequity in education will result in growing social inequality.

Over the last 25 years, a series of state supreme court rulings and concern over public education, have led many states to enact reforms with the aim of reducing inequities in public education.

This process began with the Serrano ruling in California in 1971 and by 1998 had overturned the school-financing systems in 16 states.

The effect of this litigation, both actual and threat, has been to increase the role of the state and decrease that of local provision.

There has therefore been a move from local to centralized financing.

A number of researchers have analyzed the impact of such reforms theoretically.

Most of these analyses have been in the context of static Tiebout models.

Fernandez and Rogerson (1998) *AER* present a nice analysis which traces out the dynamic implications of such reforms for the steady state income distribution and welfare.

The paper combines a static Tiebout model with an overlapping-generations structure.

Fernandez and Rogerson's model

Infinite horizon, with time periods denoted by $t = 1, ..., \infty$

2-period OLG

A continuum of individuals of measure 1 is born in each time period

Each individual belongs to a household consisting of one parent and one child

All decisions are made by parents each of whom has identical preferences given by

$$u(c, h) + Ew(y_c)$$

where c is consumption of a numeraire private good, h is consumption of housing services, and y_c is the child's income when it becomes a parent.

$u(c, h)$ is strictly concave, increasing in both arguments and defines preferences over c and h that are homothetic

$w(y_c)$ is strictly concave and increasing

Parents' incomes belong to the discrete set $\{y_1, ..., y_I\}$ where $y_1 < .... < y_I$.

A parent's income is determined by the quality of the education that it had as a child

The probability that a parent has income y_i if it had an education of quality q is $\phi_i(q)$

Let

$$v(q) = \sum_i \phi_i(q)w(y_i)$$

Then, preferences can be defined over c, h, and q:

$$u(c, h) + v(q)$$

Assume that $v(q)$ is increasing and strictly concave.

Parents must choose a community in which to live - there are two such communities indexed by $j = 1, 2$ and referred to as C_1 and C_2.

In the basic model, each community provides education, financing it by a tax on housing spending (i.e., a property tax).

Each community is characterized by a proportional tax on housing spending t_j, a quality of education q_j and a net-of-tax housing price p_j.

Let $\pi_j = (1 + t_j)p_j$ denote the gross-of-tax price of housing

The tax revenues finance education and q_j just equals per-pupil tax revenue

There are no private schools

The supply of housing in community j is $H_j^s(p_j)$.

Parents rent housing for themselves and their child - the owners of housing are outside the model.

In each period, the following happens:

(i) Parents choose where they are going to live (at this time they know their incomes)

(ii) Residents in each community choose a tax rate (majority voting)

(iii) Parents make their housing choices, pay their taxes, send their child to school, and consume.

Solving the model

(i) Within a period

For any given period and income distribution, we can solve for what happens.

This part of the model is basically a static Tiebout model in which public services are financed by property taxation and households differ in their incomes.

There are a large class of models of this type - see Epple, Filimon, and Romer (1993) for a discussion of the theoretical issues.

From a parent's perspective, a community is completely characterized by the pair (π, q); i.e., the gross-of-tax price of housing and the schooling quality.

Households take these as given when choosing their community.

A parent with income y facing the pair (π, q) has an indirect utility function

$$V(\pi, q; y) = \max u(c, h) + v(q)$$
$$\pi h + c \leq y$$

Given the pair of options $\{(\pi_1, q_1), (\pi_2, q_2)\}$ each parent will choose the community which provides the greatest utility.

Let $h(\pi, y)$ be the parent's housing demand function.

By homotheticity, this can be written as:

$$h(\pi, y) = g(\pi)y$$

where $g(\pi)$ is a decreasing function

Given a set of residents of mass N_j and a tax rate t_j in C_j, the quality of schooling q_j and the pre-tax price of housing p_j must satisfy

$$N_j g(\pi_j)\mu_j = H_j^s(p_j)$$

and

$$t_j p_j g(\pi_j)\mu_j = q_j$$

where μ_j is mean income in C_j.

We can solve the first equation for p_j for any given choice of t_j

The second equation then gives the associated level of q_j.

In this way, we can write school quality and post-tax housing prices as functions $q_j = q(t_j, \mu_j, N_j)$ and $\pi_j = \pi(t_j, \mu_j, N_j)$

To progress further, some more structure needs to be imposed on preferences:

Assumption: *For all* (q, π),

$$\frac{v'(q)}{u_c(y - \pi h(\pi, y), h(\pi, y))h(\pi, y)}$$

is increasing in y.

Note that

$$V(\pi, q; y) = u(y - \pi h(\pi, y), h(\pi, y)) + v(q)$$

Thus,

$$\frac{\partial V(\pi, q; y)}{\partial q} = v'(q)$$

and, by the Envelope Theorem

$$\frac{\partial V(\pi, q; y)}{\partial \pi} = u_c(y - \pi h(\pi, y), h(\pi, y))h(\pi, y)$$

Thus,

$$\begin{aligned} & \frac{v'(q)}{u_c(y - \pi h(\pi, y), h(\pi, y))h(\pi, y)} \\ = \quad & -\frac{\partial V(\pi, q; y)/\partial q}{\partial V(\pi, q; y)/\partial p} \\ = \quad & \text{MRS between } q \;\&\; \pi \end{aligned}$$

This assumption therefore says that higher income individuals have steeper indifference curves in the (q, π) plane.

This type of assumption is necessary to ensure existence of an equilibrium in this type of model - see Epple, Filimon, and Romer (1993).

It is often referred to as the *single-crossing property*.

Under this assumption, we obtain the result that the two communities must be *stratified*.

To be more precise, if $\{(\pi_1^*, q_1^*), (\pi_2^*, q_2^*)\}$ is an equilibrium then $(\pi_1^*, q_1^*) > (\pi_2^*, q_2^*)$ and all parents in community C_1 have higher income than those in community C_2.

Thus, community 1 has higher quality schools, higher gross-of-tax housing prices, and a higher income population.

Lower income households are deterred from moving to community 2, despite its better quality education, because the housing is more expensive.

Again, such stratification is a standard prediction in these models when single-crossing is assumed.

The equilibrium tax rates in the two communities will be those preferred by the parent with the median income.

The preferred tax rate of a parent with income y in community C_j solves the problem

$$max_t V(\pi(t, \mu_j, N_j), q(t, \mu_j, N_j); y)$$

The first order condition for the problem implies that

$$u_c(\cdot)h(\cdot)\pi_t(\cdot) = v'(\cdot)q_t(\cdot)$$

The right hand side is the marginal benefit of taxation and the left hand side is the marginal cost

It is not possible to say in general whether t_1 will be greater or less than t_2.

While community 1 has higher quality education, it has a bigger tax base because households spend more on housing.

(ii) Across periods

The state variable of the model is the income distribution of the parents $\lambda = (\lambda_1, ..., \lambda_I)$ where λ_i is the fraction of parents with income y_i.

Let $\Lambda(\lambda)$ denote the set of income distributions that could be generated by the within-period equilibrium if the income distribution of the parents were given by λ.

Thus, if the income distribution of the parents in period t is λ^t, then the income distribution in period $t+1$ is an element of the set $\Lambda(\lambda^t)$

An income distribution λ^* is a *steady state* if $\lambda^* \in \Lambda(\lambda^*)$.

The policy experiment

We are interested in the implications of a switch from a local community based system of educational finance to a centralized system.

The model we have described is of a local system.

To model a centralized system, Fernandez and Rogerson suppose that the state government levies a property tax t that applies to both communities and provides a uniform quality of schooling across the communities.

The tax rate is chosen by the parents in both communities (again by majority voting).

Lets think intuitively about the implications of this change.

If school quality and taxes are equalized across districts, then housing prices must be equal across communities

The quality of schooling will be determined by the tax rate the voter with the median income prefers

But the median income depends upon the income distribution and hence is endogenous.

Thus, it may take a while for a steady state to be reached

The nice thing about this model is that allows the long run implications of this change to be worked out.

Nonetheless, it is not possible to say much analytically, so they calibrate and simulate, computing the steady state income distributions under the two systems

Results

Moving to centralized financing increases both average income and education spending as a fraction of total consumption.

See Tables 1,2,3 & 4 of the paper for the details.

To get a welfare measure, they compute the expected utility of a parent whose income is a random draw from the steady state income distribution.

This is the same thing as assuming a utilitarian social welfare function.

Welfare is higher under centralized financing.

For a money measure of the welfare gain they compute the percent by which the vector of income $\{y_1, ..., y_I\}$

would have to be reduced each period in the centralized financing system to make aggregate utility the same as it was under local financing.

The amount is 3.2%.

They argue that this means that the gains from centralized financing are significant.

4) Capitalization

Tiebout's idea that people vote with their feet suggests that the benefits of local government amenities and the costs of a community's tax obligations should be capitalized into housing prices.

There is a huge literature, mostly empirical, exploring variants of this idea.

This literature follows the seminal work of Oates entitled "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis" which was published in the 1969 *JPE.*

Oates provided evidence that property tax rates and school spending levels were capitalized into house prices in New Jersey municipalities.

On the theory side, most treatments of the idea are static and use simple supply and demand models.

This has the disadvantage of obscuring exactly what should be capitalized.

In this lecture, I present a simple dynamic model of capitalization.

The model

Consider a community such as a municipality or a school district.

The time horizon is infinite, with periods indexed by t.

There is a set of potential residents of the community of size 1.

Potential residents are characterized by the strength of their desire to live in the community which is measured by the preference parameter θ that takes on values between 0 and $\overline{\theta}$.

The fraction of potential residents with preference parameter below $\theta \in [0, \overline{\theta}]$ is $\theta / \overline{\theta}$.

There is turnover, so that in each period new households join the group of potential residents and old ones leave.

The probability that a household currently a potential resident will be one in the subsequent period is μ.

Thus, in each period, a fraction $1 - \mu$ of households leave the pool of potential residents and are replaced by an equal number of new ones.

The only way to live in the community is to own a house.

There are a fixed number of houses sufficient to accomodate a population of size H where H is less than the size of the pool of potential residents (i.e., 1).

These houses are infinitely durable.

The government of the community provides a durable public good which depreciates at rate $\delta \in (0, 1)$.

In a period in which the community's level of the public good is g, the government invests in $I(g) \geq 0$ units.

The investment function $I(g)$ is assumed to be exogenous.

Investment costs c per unit and is financed by a tax on the residents.

The investment takes time to build and is not available for use until the next period.

The community does not pay for the investment until it is complete and hence taxes to finance the investment are also not levied until the next period.

Each household in the pool of potential residents receives an exogenous income per period sufficient to pay taxes and purchase housing in the community.

When living in the community, households have preferences defined over the public good and private consumption.

A household with preference parameter θ with private consumption x obtains a period payoff of $\theta + x + B(g)$ if it lives in the area when the level of the public good is g.

The public good benefit function $B(g)$ is increasing, smooth, strictly concave, and satisfies $B(0) = 0$.

When not living in the community, a household's per period payoff is just given by $\overline{u}$.

Households discount future payoffs at rate β and can borrow and save at rate $\rho = 1/\beta - 1$.

There is a competitive housing market which opens at the beginning of each period.

Demand comes from new households moving into the community and supply comes from owners leaving the community.

The price of houses is denoted P.

The timing of the model is as follows.

Each period, the community starts with a public good level g and a tax obligation T (which may be zero).

The public good level is the sum of the depreciated level from the prior period and any investment approved in the prior period.

The tax obligation is to finance any investment undertaken in the prior period.

At the beginning of the period, those in the pool of potential residents learn whether they will be remaining in the pool and new households join.

Households in the pool must then decide whether to live in the area.

The housing market opens and the equilibrium housing price $P(g,T)$ is determined.

The government levies taxes on residents sufficient to meet its tax obligation and residents obtain their payoffs from living in the area.

These depend not only on their preference θ but also on the public good level and taxes.

The government undertakes investment $I(g)$.

The community's public good level and tax obligation in the next period (g', T') is $((1-\delta)g + I(g), cI(g))$.

Housing market equilibrium and capitalization

We now explain how the housing market equilibriates given the investment path $I(g)$.

First consider the decisions of households.

At the beginning of any period, households fall into two groups: those who resided in the community in the previous period and those who did not, but could in the current period.

Households in the first group own homes, while those in the second group do not.

Households in the first group who leave the set of potential residents will sell their houses and obtain a continuation payoff of

$$P(g, T) + \frac{\underline{u}}{1-\beta}.$$

The remaining households in the first group and all those in the second must make a location decision $l \in \{0,1\}$, where $l = 1$ means that they live in the community.

This decision will depend on their preference parameter θ and on the level of public goods and taxes.

Since selling a house and moving is costless, there is no loss of generality in assuming that all households sell their property at the beginning of any period.

This makes each household's location decision independent of its property ownership state.

It also means that the only future consequences of the current location choice is through the selling price of housing in the next period.

To make this more precise, let $V_\theta(g,T)$ denote the expected payoff of a household with preference parameter θ at the beginning of a period in which it belongs to the set of potential residents but does not own a house.

Then, we have that

$$V_\theta(g,T) = \max_{l\in\{0,1\}} \left\{ \begin{array}{c} l\left(\theta + B(g) - T/H - P(g,T) + \beta P(g',T')\right) \\ +(1-l)\underline{u} + \beta[\mu V_\theta(g',T') + (1-\mu)\frac{\underline{u}}{1-\beta}] \end{array} \right\}.$$

Inspecting this problem, it is clear that a household of type θ will choose to reside in the community if and only if

$$\theta + B(g) - T/H - P(g,T) + \beta P(g',T') \geq \underline{u}. \quad (3)$$

We can now describe the equilibrium.

Given an initial state (g,T), the price of housing $P(g,T)$ adjusts to equate demand and supply.

The demand for housing is the fraction of households for whom (3) holds.

Given the uniform distribution of preferences, this fraction is

$$1 - \frac{\underline{u} - (B(g) - T/H - P + \beta(g', T'))}{\overline{\theta}}.$$

The supply of housing is, by assumption, perfectly inelastic at H.

The equilibrium price of housing therefore satisfies

$$1 - \frac{\underline{u} - (B(g) - T/H - P(g, T) + \beta P(g', T'))}{\overline{\theta}} = H.$$

To characterize the housing market equilibrium, define the present value of public goods surplus $S(g, T)$ recursively as follows:

$$S(g, T) = B(g) - T/H + \beta S(g', T').$$

Intuitively, $S(g, T)$ represents the discounted value of future public good surplus for a household who will be living in the community permanently starting in a period in which the community has public good level g and tax obligation T.

Capitalization Theorem *There exists a constant* K *such that the equilibrium housing price is given by*

$$P(g, T) = K + S(g, T).$$

The Theorem tells us that the equilibrium price can be expressed as the sum of a constant and the discounted value of future public good surplus.

The theorem therefore implies that the value of future public good levels and tax obligations is fully capitalized into the price of housing.

The constant K is tied down by the requirement that the marginal household with preference $(1 - H)\overline{\theta}$ is just indifferent between living and not living in the community.

Discussion

Note that what is capitalized is the entire stream of future public good benefits and tax obligations.

Denoting period 0 as the current period and future periods by $t = 1, ..., \infty$, we have that

$$S(g, T) = \sum_{t=0}^{\infty} \beta^t (B(g_t) - T_t/H)$$

where $\langle g_t, T_t \rangle_{t=0}^{\infty}$ is the sequence of policies defined inductively by the equations

$$(g_0, T_0) = (g, T)$$

and for all $t \geq 1$

$$(g_t, T_t) = ((1 - \delta)g_{t-1} + I(g_{t-1}), cI(g_{t-1}))\,.$$

Comparing across different communities, the theorem suggests that we can look to housing prices to see how households value the streams of future public good benefits and tax obligations that their local governments generate.

Consistent with its static focus, the empirical literature often assumes that housing prices reflect residents' valuations of current public good levels and taxes.

This is legitimate only if we assume that the community is in a steady state in which $I(g) = c\delta g = T$.

In this case,

$$S(g, T) = \frac{B(g) - T/H}{1 - \beta}.$$

It is also important to note that the full capitalization result follows from the assumption that the supply of houses is fixed.

Proof of Capitalization Theorem

Consider a period in which the community's level of public good is g and its tax obligation is T.

The supply of housing is H. Potential residents just differ in their preference for living in the community θ.

Clearly, those with higher θ will have a greater willingness to pay to live in the community.

Thus, the fraction H of potential residents with the highest preference parameters will live in the community.

The marginal resident will therefore have preference parameter $\overline{\theta}(1-H)$ since

$$\frac{\overline{\theta}-\overline{\theta}(1-H)}{\overline{\theta}}=H.$$

This will be the case in each and every period irrespective of the community's level of public good and tax obligation.

It follows that, in equilibrium, all types $\theta \in [0, \overline{\theta}\,(1-H))$ never reside in the community which implies that

$$V_\theta(g,T) = \frac{\underline{u}}{1-\beta}.$$

Types $\theta \in [\overline{\theta}\,(1-H)), \overline{\theta}]$, on the other hand, will reside in the community as long as they remain in the pool of potential residents.

For these types, therefore, irrespective of g and T

$$V_\theta(g,T) = \begin{array}{c} \theta + B(g) - T/H - P(g,T) + \beta P(g',T') \\ +\beta[\mu V_\theta(g',T') + (1-\mu)\frac{\underline{u}}{1-\beta}] \end{array}.$$

We now develop an expression for the value functions of these resident households.

Let $\theta \in [\overline{\theta}\,(1 - H)), \overline{\theta}]$.

Let 0 index the current period and let future periods be indexed by $t = 1, ..., \infty$.

Let $\langle g_t, T_t \rangle_{t=1}^{\infty}$ be the sequence of policies defined inductively by the equations

$$(g_1, T_1) = ((1 - \delta)g + I(g), cI(g))$$

and for all $t \geq 2$

$$(g_t, T_t) = ((1 - \delta)g_{t-1} + I(g_{t-1}), cI(g_{t-1})) \;\; .$$

We know that

$$V_\theta(g, T) = \begin{array}{c} \theta + \beta(1-\mu)\frac{u}{1-\beta} + B(g) - T/H - P(g,T) \\ +\beta\left[P(g_1, T_1) + \mu V_\theta(g_1, T_1)\right] \end{array} . \tag{4}$$

But, since the household will reside in the community in period 1 if it remains in the pool, we also know that

$$
\begin{aligned}
&\beta\left[P(g_1,T_1)+\mu V_\theta(g_1,T_1)\right] \\
= &\begin{array}{c} \beta(1-\mu)P(g_1,T_1) \\ +\beta\mu\left[\begin{array}{c} \theta+\beta(1-\mu)\frac{\underline{u}}{1-\beta}+B(g_1)-T_1/H \\ +\beta\left\{P(g_2,T_2)+\mu V_\theta(g_2,T_2)\right\} \end{array}\right] \end{array}.
\end{aligned}
$$

Moreover, period 1's housing price $P(g_1,T_1)$ satisfies the equilibrium condition

$$
1-\frac{\underline{u}-(B(g_1)-T_1/H-P(g_1,T_1)+\beta P(g_2,T_2))}{\overline{\theta}}=H,
$$

which implies that

$$
P(g_1,T_1)=\overline{\theta}(1-H)-\underline{u}+B(g_1)-T_1/H+\beta P(g_2,T_2).
$$

Substituting this into the above expression, we can write

$$
\begin{aligned}
&\beta\left[P(g_1,T_1)+\mu V_\theta(g_1,T_1)\right] \\
= \;&\kappa_1(\theta)+\beta\left[B(g_1)-T_1/H\right] \\
&+\beta^2\left[P(g_2,T_2)+\mu^2 V_\theta(g_2,T_2)\right]
\end{aligned}
$$

where

$$
\kappa_1(\theta)=\beta\left\{\begin{array}{c} (1-\mu)\left[\overline{\theta}(1-H)-\underline{u}\right] \\ +\mu\left[\theta+\beta(1-\mu)\frac{\overline{u}}{1-\beta}\right] \end{array}\right\}.
$$

Again, since the household will reside in the community in period 1 if it remains in the pool, we also know that

$$\begin{aligned} & \beta^2 \left[P(g_2,T_2)+\mu^2 V_\theta(g_2,T_2)\right] \\ = \quad & \begin{array}{c} \beta^2(1-\mu^2)P(g_2,T_2) \\ +\beta^2\mu^2 \left[\begin{array}{c} \theta+\beta(1-\mu)\frac{\underline{u}}{1-\beta}+B(g_2)-T_2/H \\ +\beta\left\{P(g_3,T_3)+\mu V_\theta(g_3,T_3)\right\} \end{array}\right] \end{array}. \end{aligned}$$

Equilibrium in the housing market implies that

$$P(g_2,T_2)=\overline{\theta}(1-H)-\underline{u}+B(g_2)-T_2/H+\beta P(g_3,T_3).$$

Substituting this in, we can write

$$\begin{aligned} & \beta^2\left[P(g_2,T_2)+\mu^2 V_\theta(g_2,T_2)\right] \\ = \quad & \kappa_2(\theta)+\beta^2\left[B(g_2)-T_2/H\right] \\ & +\beta^3\left[P(g_3,T_3)+\mu^3 V_\theta(g_3,T_3)\right] \end{aligned}$$

where

$$\kappa_2(\theta)=\beta^2\left\{\begin{array}{c} (1-\mu^2)\left[\overline{\theta}(1-H)-\underline{u}\right] \\ +\mu^2\left[\theta+\beta(1-\mu)\frac{\underline{u}}{1-\beta}\right] \end{array}\right\}.$$

By similar logic, for all periods $t \geq 3$, we have that

$$\begin{aligned} & \beta^t \left[P(g_t, T_t) + \mu^t V_\theta(g_t, T_t) \right] \\ = \; & \kappa_t(\theta) + \beta^t \left[B(g_t) - T_t/H \right] \\ & + \beta^{t+1} \left[P(g_{t+1}, T_{t+1}) + \mu^{t+1} V_\theta(g_{t+1}, T_{t+1}) \right] \end{aligned}$$

where

$$\kappa_t(\theta) = \beta^t \left\{ \begin{array}{c} (1-\mu^t) \left[\overline{\theta}(1-H) - \underline{u} \right] \\ +\mu^t \left[\theta + \beta(1-\mu) \frac{\underline{u}}{1-\beta} \right] \end{array} \right\}.$$

Successively substituting these expressions into (4), reveals that

$$\begin{aligned} V_\theta(g, T) = & \theta + \beta(1-\mu) \frac{\underline{u}}{1-\beta} + B(g) - T/H - P(g, T) \\ & + \sum_{t=1}^{\infty} \left\{ \kappa_t(\theta) + \beta^t \left(B(g_t) - T_t/H \right) \right\} \\ = & \sum_{t=0}^{\infty} \kappa_t(\theta) + S(g, T) - P(g, T). \end{aligned}$$

Now, it must be the case that in equilibrium the marginal household, which is the household with preference $\overline{\theta}(1-H)$, is just indifferent between residing in the community or not.

Thus, it must be the case that

$$\sum_{t=0}^{\infty} \kappa_t(\overline{\theta}(1-H)) + S(g,T) - P(g,T) = \frac{\overline{u}}{1-\beta}.$$

This implies that

$$P(g,T) = \sum_{t=0}^{\infty} \kappa_t(\overline{\theta}(1-H)) - \frac{\overline{u}}{1-\beta} + S(g,T).$$

Defining

$$K = \sum_{t=0}^{\infty} \kappa_t(\overline{\theta}(1-H)) - \frac{\overline{u}}{1-\beta},$$

yields the result. ■

Capitalization and optimal investment

Now suppose that investment in the public good were chosen each period to maximize the expected payoffs of those households residing in the community.

One might worry that, given that they may have to leave the community (with probability $1-\mu$), the residents would be short-sighted and underinvest.

However, since they will be rewarded for investing by getting a higher price for their houses when they leave, capitalization actually provides them with optimal incentives.

To see this consider a period in which the current level of the public good is g.

If the residents invest I units, then a resident of type $\theta \in [\overline{\theta}(1-H)), \overline{\theta}]$ will obtain an expected continuation payoff

$$P((1-\delta)g+I, cI)+\mu V_{\theta}((1-\delta)g+I, cI)+(1-\mu)\frac{\overline{u}}{1-\beta}$$

In the proof of the Capitalization Theorem, we showed that for households who would reside in the community (i.e., households for whom $\theta \in [\overline{\theta}(1-H)), \overline{\theta}]$), we have that for some constant $\kappa(\theta)$

$$V_\theta(g, T) = \kappa(\theta) + S(g, T) - P(g, T).$$

Given that, by the Theorem, there exists some constant K such that the equilibrium price of housing is such that

$$P(g, T) = K + S(g, T),$$

it follows that

$$V_\theta(g, T) = \kappa(\theta) - K.$$

Thus, we can write the expected continuation payoff of a resident of type $\theta \in [\overline{\theta}(1-H)), \overline{\theta}]$

$$
\begin{aligned}
& P((1-\delta)g + I, cI) + \mu V_\theta((1-\delta)g + I, cI) \\
& \quad + (1-\mu)\frac{\overline{u}}{1-\beta} \\
= & K + S((1-\delta)g + I, cI) + \mu[\kappa(\theta) - K] \\
& \quad + (1-\mu)\frac{\overline{u}}{1-\beta}
\end{aligned}
$$

The residents will therefore choose an investment level to solve public good surplus; i.e.,

$$\max_{I \geq 0} S((1-\delta)g + I, cI).$$