1. Commodity Taxation I: Ramsey Rule
2. Commodity Taxation II: Production Efficiency
3. Income Taxation I: Mirrlees Model
We use what we know about incidence and efficiency costs to analyze optimal design of commodity taxes

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

The literature on optimal commodity taxation focuses on linear tax systems

Non-linear tax systems are studied with income taxation
The starting point is the Second Theorem of Welfare Economics: any Pareto optimal outcome can be achieved as a competitive equilibrium with appropriate lump-sum transfers of wealth.

This result requires the same assumptions as the First Theorem plus one more (and some convexity requirements):

1. Complete markets (no externalities)
2. Perfect information
3. Perfect competition
4. Lump-sum taxes/transfers across individuals feasible

If all assumptions hold, there is no equity-efficiency trade-off and the optimal tax problem is trivial. Implement LSTs that meet distributional goals given revenue requirement.

Problem: information
To set optimal LSTs, government needs to know individual characteristics (ability or earning capacity).

Impossible to make people reveal ability at no cost: If asked, they have incentives to misreport skill levels.

Tax instruments must be a function of observable economic outcomes (income, property, consumption of goods).

This distorts prices, affecting behavior and generating DWL.

Information constraints force us to move from 1st best world of the second welfare theorem to 2nd best world with inefficient taxation: It is impossible to redistribute or raise revenue for public goods without generating efficiency costs.
Central Results in Optimal Tax Theory

1. Ramsey (1927): inverse elasticity rule


5. Chamley, Judd (1983): no capital taxation in infinite horizon models
Government sets taxes on uses of income in order to fulfill two goals:

1. Raise total revenue $R$

2. Minimize utility loss for agents in economy. Or to minimize excess burden of taxation.
Ramsey Model: Key Assumptions

1. Lump sum taxes are prohibited
2. Cannot tax all commodities (leisure untaxed)
3. Production prices $p$ are fixed. Consumer prices are $q = p + t$
Set-up

- A population of identical individuals (equivalent to a single representative consumer) $\implies$ No equity concerns
- The consumer maximizes $u(x)$, where $x = (x_0, \ldots, x_N)$, subject to $q_1x_1 + \ldots + q_Nx_N \leq Z$, where $Z$ is non-wage income
- From the point of view of government is the same to set $(t_1, \ldots, t_N)$ than to set $(q_1, \ldots q_N)$, given $(p_1, \ldots p_N)$
- The FOCs are: $\frac{\partial u}{\partial x_i} = \alpha q_i$, where $\alpha = \frac{\partial V}{\partial Z}$ is the marginal utility of income for the individual
- We get (Marshallian) demands $x_i(q, Z)$ and indirect utility function $V(q, Z)$, where $q = (q_1, \ldots q_N)$
Excess burden (using Equivalent Variation)

- Government solves:

$$\max_q \{ V(q, Z) \} \text{ s.t. } tx = \sum_{i=1}^{N} t_i x_i(q, Z) \geq R.$$

- Equivalent to minimize the excess burden of taxation:

$$\min_q EB(q) = E(q, V(q, Z)) - E(p, V(q, Z)) - R,$$

subject to the same revenue requirement.

- Why? Because $E(p, V(q, Z))$ is monotonically increasing in $V(q, Z)$ and $R$ and $Z \equiv E(q, V(q, Z))$ are constant. Maximizing $V(q, Z)$ is the same as minimizing $-E(p, V(q, Z))$. The excess burden of taxation is evaluated at the utility level $V(q, Z)$ that holds in the presence of taxation.
If we could tax all commodities, the problem is trivial. Consider uniform taxation, where \( q = \phi p \) and \( \phi > 1 \) is chosen such that \((\phi - 1)p = \phi p x = R\).

Excess burden is \( Z - E(p, V(q, Z)) - R \). Substituting:

\[
E(\phi p, V(\phi p, Z)) - E(p, V(\phi p, Z)) - (\phi - 1)p x(\phi p, Z)
\]

Since \( E(., .) \) is homogeneous of degree 1 on prices:

\[
\phi E(p, V(\phi p, Z)) - E(p, V(\phi p, Z)) - (\phi - 1)p x(\phi p, Z)
= (\phi - 1)E(p, V(\phi p, Z)) - (\phi - 1)p x(\phi p, Z)
\]

Finally, using the fact that:

\[
E(p, V(\phi p, Z)) \equiv px^C(\phi p, V(\phi p, Z)) \equiv px(\phi p, Z)
\]

\[
(\phi - 1)p x(\phi p, Z) - (\phi - 1)p x(\phi p, Z) = 0
\]
With uniform taxation, excess burden is zero. This is equivalent to imposing a LST.

The constraint for the household is $\phi p_x = Z$ or:

$$p_x = Z - \frac{(\phi - 1)Z}{\phi}$$

If we could tax all commodities, uniform taxation would be optimal.

Next we see what can we do if we cannot tax all commodities.
The Lagrangian of government is:

$$L_G = V(q, Z) + \lambda [\sum_i t_i x_i(q, Z) - R]$$

Using Roy's identity ($\frac{\partial V}{\partial q_i} = -\alpha x_i$):

$$(\lambda - \alpha) x_i + \lambda \sum_j t_j \frac{\partial x_j}{\partial q_i} = 0$$
Optimal tax rates have to satisfy a system of $N$ equations and $N$ unknowns:

$$\sum_j t_j \frac{\partial x_j}{\partial q_i} = -\frac{x_i}{\lambda} (\lambda - \alpha) \quad \text{for } i = 1, \ldots, N$$

Here note that one commodity cannot be taxed (think of commodity 0)

Also note that $\partial x_j / \partial q_i = \partial x_j / \partial t_i$

Using a perturbation argument we can arrive at the same formula
Ramsey Formula: Perturbation Argument

- Suppose government increases $t_i$ by $dt_i$

- The effect of this tax increase on social welfare is the sum of effect on government revenue and on private surplus

- The marginal effect on government revenue is:
  \[ dR = x_i dt_i + \sum_j t_j dx_j \]

- The marginal effect on private surplus is:
  \[ dU = \frac{\partial V}{\partial q_i} dt_i = -\alpha x_i dt_i \]

- At an optimum these two effects cancel each other:
  \[ dU + \lambda dR = 0 \]
Recall the Slutsky equation:

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

We substitute into formula above:

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

\[ \Rightarrow \frac{1}{x_i} \sum_j t_j \frac{\partial h_i}{\partial q_j} = -\frac{\theta}{\lambda} \]

where \( \theta = \lambda - \alpha - \lambda \frac{\partial}{\partial Z} (\sum_j t_j x_j) \)
The term $\theta$ is independent of $i$ and measures the value for the government of introducing a 1 euro LST:

$$\theta = \lambda - \alpha - \lambda \frac{\partial (\sum_j t_j x_j)}{\partial Z}$$

Three effects:

1. Direct value for the government is $\lambda$ (recall this is the Lagrange multiplier of the government budget constraint).
2. Loss in welfare for the individual is $\alpha$ (this is the Lagrange multiplier of the consumer’s problem).
3. Behavioral effect $\rightarrow$ loss in tax revenue of $\frac{\partial (\sum_j t_j x_j)}{\partial Z}$.
Intuition for Ramsey Formula

\[
\frac{1}{x_i} \sum_j t_j \frac{\partial h_i}{\partial q_j} = -\frac{\theta}{\lambda}
\]

- Suppose that \( E \) is small so that all taxes are also small. Then the tax \( t_j \) on good \( j \) reduces the consumption of good \( i \) (holding utility constant) by approximately:

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

- The summatory on the left represents the total reduction in consumption of good \( i \) (because of all taxes)

- Dividing by \( x_i \) yields the percentage reduction in consumption of each good \( i = \text{"index of discouragement"} \) of the tax system on good \( i \)

- The Ramsey formula says that these indexes must be equal across goods at the optimum
Using elasticities, we can write Ramsey formula as:

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

where \( \varepsilon_{ij}^c \) is the compensated elasticity (from Hicksian demands). We have normalized \( p_j = 1 \).

In the special case in which all cross-elasticities are zero we obtain the classic inverse elasticity rule:

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

More elastic goods should have lower taxes.
Consider a case with only three goods 0, 1, 2 where 0 is untaxed. We have two equations (there are 2 taxes):

\[
\begin{align*}
\frac{t_1}{1 + t_1} \varepsilon_{11}^c + \frac{t_2}{1 + t_2} \varepsilon_{12}^c &= \frac{\theta}{\lambda} \\
\frac{t_1}{1 + t_1} \varepsilon_{21}^c + \frac{t_2}{1 + t_2} \varepsilon_{22}^c &= \frac{\theta}{\lambda}
\end{align*}
\]

Rearranging, we obtain:

\[
\begin{align*}
\frac{t_1}{(1 + t_1)} &= \frac{\varepsilon_{22}^c - \varepsilon_{12}^c}{\varepsilon_{11}^c - \varepsilon_{21}^c} \\
\frac{t_2}{(1 + t_2)} &= \frac{\varepsilon_{22}^c - \varepsilon_{12}^c}{\varepsilon_{11}^c - \varepsilon_{21}^c}
\end{align*}
\]
Uniform Taxation?

- Using the property that $\varepsilon_{11}^c + \varepsilon_{12}^c + \varepsilon_{10}^c = 0$:

$$\frac{t_1}{(1 + t_1)} = \frac{\varepsilon_{20}^c + \varepsilon_{21}^c + \varepsilon_{12}^c}{\varepsilon_{10}^c + \varepsilon_{12}^c + \varepsilon_{21}^c}$$

- We have $t_1 = t_2$ only when $\varepsilon_{10}^c = \varepsilon_{20}^c$. This happens if both goods are equally complementary with respect to the untaxed good (the numeraire).

- If $\varepsilon_{10}^c < \varepsilon_{20}^c$, good 1 should be taxed at a higher rate. This result is called **Corlett-Hague theorem**: With preferences that are not separable between taxable goods and the untaxed good, government should deviate from uniform taxation by taxing more heavily goods that are more complementary to the untaxed good.
Corlett-Hague Theorem

- When the untaxed good is leisure, this means higher taxes for goods that are complementary to leisure (lower taxes for goods that are complementary to labor)
- Classical example: swimsuits and pizza. Swimsuits should be taxed more than pizza
- Deaton finds a necessary and sufficient condition under which we get uniform taxation as a prescription of the Ramsey rule. It is called “Quasi-separability” of the utility function: marginal rates of substitution between goods on a given indifference curve must be independent of the consumption of the untaxed good (for example, Cobb-Douglas)
Limitations of Ramsey Formula

- According to Ramsey Formula government should tax inelastic goods to minimize efficiency costs

- We do not take into account redistributional issues (representative consumer)

- Problem: necessities tend to be more inelastic than luxury goods

- Optimal Ramsey tax system is likely to be regressive

- Diamond (1975) extends Ramsey model to take redistributive concerns into account
Diamond 1975: Many-Person Model

- $H$ individuals with utilities $u^1, ..., u^h, ..., u^H$

- Aggregate consumption of good $i$ is:

$$X_i(q) = \sum_h x^h_i$$

- Government chooses tax rates $t_i$ and a lump sum transfer $T \geq 0$ to maximize social welfare:

$$\max W(V^1, ..., V^H) \text{ s.t. } \sum_{i=1}^{N} t_i X_i \geq R + T$$

- Consider the effect of increasing the tax on good $i$ by $dt_i$
Diamond: Perturbation Argument

- Effect of perturbation on revenue:

\[ dR = X_i dt_i + \sum_j t_j dX_j = dt_i \left[ X_i + \sum_j t_j \frac{\partial X_j}{\partial q_i} \right] \]

- Effect on individual \( h \)'s welfare (recall \( \frac{\partial V^h}{\partial q_i} = -\alpha^h x_i^h \)):

\[ du^h = \frac{\partial V^h}{\partial q_i} dt_i = -\alpha^h x_i^h dt_i \]

- Effect on total private welfare:

\[ dW = \sum_h \left( -\frac{\partial W}{\partial V^h} \right) \alpha^h x_i^h dt_i = -dt_i \left[ \sum_h \beta^h x_i^h \right] \]

where \( \beta^h = \left( \frac{\partial W}{\partial V^h} \right) \alpha^h \) is \( h \)'s social marginal utility of wealth

- At the optimum:

\[ dW + \lambda dR = 0 \]
Solving yields formula for optimal tax rates:

$$\sum_j t_j \frac{\partial X_j}{\partial q_i} = -\frac{X_i}{\lambda} [\lambda - \frac{\sum h \beta^h x^h_i}{X_i}]$$

If government does not care for redistribution: $\beta^h = \alpha$ constant: Back to Ramsey Formula

With redistributive tastes, $\beta^h$ will be lower for higher income individuals

The new term $\left(\frac{\sum h \beta^h x^h_i}{X_i}\right)$ is the average social marginal utility, weighted by consumption of good $i$
Diamond Formula: Special Case

- When all uncompensated cross price elasticities are zero, optimal tax rates satisfies:
  \[
  \frac{t_i}{1 + t_i} = \frac{1}{\epsilon_{ii}} \left( 1 - \frac{\sum_h \beta^h x_i^h}{\lambda X_i} \right)
  \]

- Again \( t_i \) must be inversely proportional to the elasticity, but the term in brackets is not constant across goods

- For goods that are consumed by people with a high weight in the SWF, the term \((\sum_h \beta^h x_i^h) / (\lambda X_i)\) is large

- The optimal tax rate for these goods is lower (elasticities being the same)

- The opposite happens for goods consumed by people with low weight in the SWF (the rich)
Optimal Transfer

- In this model, it is optimal for the government to pay a uniform transfer $T$ on top of tax rates.

- With redistributive tastes, this uniform transfer is strictly positive ($T > 0$).

- Without redistributional considerations, $T = -R$.

- This is ruled out by the constraint $T \geq 0$: the poor cannot afford to pay a LST.
Diamond and Mirrlees (1971) relax the assumption that producer prices are fixed. They model production

Two main results:

1. Production efficiency: even in an economy where first-best is unattainable, optimal policy maintains production efficiency (production efficiency: cannot produce more output with same inputs, cannot produce same output with fewer inputs)

2. Characterize optimal tax rates with endogenous prices and show that Ramsey rule can be applied
Model

- Robinson Crusoe economy: one consumer, one firm, two goods (labor, consumption). Labor is used as input to produce consumption that is sold by the firm to the consumer \( (x = f(l)) \)

- Government cannot use a LST (we have in mind a case with many individuals)

- Government directly chooses allocations and production subject to requirement that allocation must be supported by an equilibrium price vector

- Government levies tax \( t \) on consumption to fund revenue requirement \( R \)

- Individual budget constraint: \( (1 + t)c \leq l \) (pre-tax prices are 1)

- By changing \( t \) we get demand as a function of tax rates: the Offer curve
Figure 3

Source: Diamond and Mirrlees (1971)
Social Planner’s problem

Government’s problem is:

\[
\max_{\tau} V(q) = u(c(q), l(q))
\]

subject to 2 constraints:

1. Revenue constraint: \( tc \geq R \)

2. Production constraint: \( x = f(l) \)

Replace these constraints by \((l, c) \in H\) where \(H\) is the production set (taking into account the revenue constraint)
Production Set with Revenue Requirement

Figure 4

Source: Diamond and Mirrlees (1971)
Lumpsum tax payment to cover requirement in excess of profits

Source: Diamond and Mirrlees (1971)
2nd Best: Optimal Distortionary Tax

Figure 5

Source: Diamond and Mirrlees (1971)
Key insight: allocation resulting from optimal distortionary tax is still on the boundary of the production set (production efficiency): Despite the need to distort consumption choices to raise revenue (or redistribute), the production side should not be distorted.

The equilibrium price vector $q$ places the consumer on the boundary, subject to revenue requirement.

With a LST, there is tangency between the boundary of the production set and consumer’s indifference curve, yielding higher welfare.
With many firms, the condition of production efficiency is that the marginal rate of technical substitution (MRTS) between any two inputs is the same for all firms (recall that $MRTS_{KL} = \frac{\partial f}{\partial K} \frac{\partial f}{\partial L}$. Here $f$ is the production function).

With two firms, $x$ and $y$ and two inputs, $K$ and $L$, with an optimal tax schedule:

$$MRTS^x_{KL} = MRTS^y_{KL}$$

Without taxation, profit maximizing firms in competitive markets always reach production efficiency.

The same result is obtained with taxation, provided all firms face the same after-tax prices for inputs, so input taxes are not differentiated between firms.
No Taxation of Intermediate Goods

- **Intermediate goods**: goods that are neither direct inputs or outputs to individual consumption

- Taxes on transactions between firms would distort production

- Consider two industries, with labor as the primary input: intermediate good $A$, final good $B$

- Industry $A$ uses labor ($l_A$) to produce good $A$ with a one-to-one technology: $x_A = l_A$

- Industry $B$ uses good $A$ and labor $l_B$ to produce final good $B$: $x_B = F(l_B, x_A)$. CRS technology

- With wage rate $w$, producer price of good $A$ is $p_A = w$

- Suppose that good $A$ is taxed at rate $\tau$. The cost for firm $B$ of good $A$ is $w + \tau$
Firm \( B \) chooses \( l_B \) and \( x_A \) to maximize \( F(l_B, x_A) - wl - (w + \tau)x_A \)

We get \( F_I = w \) and \( F_A = w + \tau > F_I \)

Aggregate production is inefficient. Reduce \( l_B \) and increase \( l_A \) a small amount

Then \( x_A \) increases. Total production of good \( B \) increases. Tax revenue rises (government budget constraint satisfied)
To sum up, the Diamond-Mirrlees result gives a rationale for

- Non-taxation of intermediate goods: VAT as correct form of indirect taxation
- Non-differentiation of input taxes between firms
- No tariffs should be imposed on goods and inputs imported or exported by the production sector

A key assumptions is that consumption side is independent of the production side
Notation and Basic Concepts

- Let \( T(z) \) denote tax liability as a function of earnings \( z \).

- Transfer benefit with zero earnings is \( -T(0) \geq 0 \) (demogrant or lumpsum grant).

- **EXAMPLE (Linear Tax):** \( T(z) = T(0) + \tau z \), with \( T(0) \leq 0 \) and \( 0 < \tau < 1 \). After tax or net income is:

  \[
  z - T(z) = -T(0) + (1 - \tau)z
  \]

- Here the marginal tax rate is constant \((\tau)\).

- Generically, marginal tax rate is \( T'(z) \): individual keeps \( 1 - T'(z) \) for an additional euro of earnings.
Notation and Basic Concepts

- When moving from zero to $z$ earnings, net income rises from $-T(0)$ to $z - T(z)$. Individual keeps a fraction:

\[
\frac{z - T(z) + T(0)}{z} = 1 - \frac{T(z) - T(0)}{z}
\]

- We call $\tau_p = \frac{T(z) - T(0)}{z}$, the participation tax rate. Individual keeps fraction $1 - \tau_p$ of earnings when moving from 0 to $z$:

\[
z - T(z) = -T(0) + z - [T(z) - T(0)] = -T(0) + z(1 - \tau_p)
\]

- If $-T(0) = 5000$, $z = 10000$, $T(10000) = 2000$, $\tau_p = 0.7$ ($1 - \tau_p = 0.3$)

- Break-even earnings point $z^*$: point at which $T(z^*) = 0$
No Behavioral Responses

- All individuals have same utility $u(c)$, strictly increasing and concave, where $c$ is after-tax income.

- We take income $z$ as fixed for each individual: $c = z - T(z)$. Income $z$ has density $h(z)$.

- Utilitarian government maximizes the sum of utilities:

$$\int_0^\infty u(z - T(z))h(z)dz$$

- Subject to the budget constraint $\int T(z)h(z)dz \geq R$ (multiplier $\lambda$).
The Lagrangian is:

\[ \mathcal{L} = [u(z - T(z)) + \lambda T(z)]h(z) \]

The FOC is:

\[ [-u'(z - T(z)) + \lambda]h(z) = 0 \]

We have \( u'(z - T(z)) = \lambda \Rightarrow z - T(z) = c \) constant for all \( z \)

Then \( c = \bar{z} - R \), where \( \bar{z} = \int zh(z)dz \) average income

Government should confiscate all income and equalize all after-tax incomes
Issues with Simple Model

- A marginal tax rate of 100% would destroy incentives to work. The assumption that $z$ is exogenous is not realistic.
- Optimal income tax theory incorporates behavioral responses since the work of Mirrlees (1971).
- Even without behavioral responses people could object a tax of 100%. Can be seen as confiscatory.
- There are political constraints that put a bound on what governments can do.
Second Welfare Theorem

- People have different abilities to earn income
- According to the 2nd Welfare Theorem, any Pareto efficient allocation can be reached as a competitive equilibrium, provided that a suitable redistribution of initial endowments through individualized lump-sum taxes can be done
- Because of informational problems, government cannot do this. Need to use distortionary taxes and transfers based on observables (income and consumption) to redistribute
- There is a conflict between efficiency and equity
Efficiency-Equity Trade-Off

- Tax revenue can be used to transfer programs aimed to reduce inequality. This is desirable if society is concerned by inequality and poverty.
- Taxes and transfers reduce incentives to work. High tax rates create inefficiency if individuals react to taxes.
- This generates an equity-efficiency trade-off.
The crucial point in Mirrlees (1971) is that we have to take into account distortions on labor supply.

There is a continuum of individuals who differ in productivity (wage) $w$ distributed with density $f(w)$.

Individuals maximize utility $u(c, l)$, subject to the constraint $c = wl - T(wl)$, where $c$ is consumption (or after-tax income), $l$ is labor supply, and $T(wl)$ is a tax-transfer schedule. Solution yields indirect utility $V(w)$.

Government cannot observe $w$. Only observes $wl$.
The problem of the government is to find a tax schedule \( T(\cdot) \) to maximize:

\[
SWF = \int G(u(c, l)) f(w) dw
\]

s.t. resource constraint:

\[
\int T(wl) f(w) dw \geq E
\]

and individual FOC:

\[
w(1 - T') u_c + u_l = 0
\]

where \( G(\cdot) \) is increasing and concave.

The general solution to this problem is very complex and provides little intuition. Due to this, Mirrlees carried out several simulations using a Cobb-Douglas utility function, a log-normal wage distribution and a classical utilitarian SWF.
Mirrlees 1971: Results

- Optimal tax schedule is approximately linear above a certain threshold. Other authors as Tuomala (1984) found this result is not robust.

- Get $T(\cdot) < 0$ at the bottom (transfer) and $T(\cdot) > 0$ further up (tax).

- Due to complexity, a few general results:
  1. $0 \leq T'(\cdot) \leq 1$
  2. The marginal tax rate $T'(\cdot)$ should be zero at the top if skill distribution bounded.
After Mirrlees

- Deep impact on information economics
- Little impact on practical tax policy until recently
- Piketty (1997), Diamond (1998) and Saez (2001) connect Mirrlees approach to empirical literature
- Obtain formulas in terms of labor supply elasticities
- Three lines: 1) Revenue-maximizing linear tax (Laffer curve); 2) Top income tax rates; 3) Full income tax schedules
Laffer Curve

• With a constant tax rate $\tau$, **reported income** $z$ depends on $1 - \tau$ (net-of-tax rate)

• Total tax revenue is $R(\tau) = \tau \cdot z(1 - \tau)$. This has an inverse-U shape: $R(\tau = 0) = R(\tau = 1) = 0$

• Tax rate $\tau^*$ that maximizes $R$:

$$0 = R'(\tau^*) = z - \tau^* \frac{dz}{d(1 - \tau)}$$

$$\Rightarrow \tau^*_{\text{MAX}} = \frac{1}{1 + \varepsilon}$$

where $\varepsilon = \left[\frac{(1 - \tau)}{z}\right] \frac{dz}{d(1 - \tau)}$ is the elasticity of reported income

• Strictly inefficient to have $\tau > \tau^*$ (if $\varepsilon = 0$, $\tau^*_{\text{MAX}} = 1$; if $\varepsilon = 1$, $\tau^*_{\text{MAX}} = 1/2$; etc.)
Now assume that the marginal tax rate is constant ($\tau$) above some fixed income threshold $\bar{z}$

We derive the optimal value of $\tau$ using a perturbation argument

Assume no income effects $\varepsilon^c = \varepsilon^u = \varepsilon$

Total number of individuals above $\bar{z}$ is $N$

We denote by $z^m(1 - \tau)$ their average income (depends on net-of-tax rate $1 - \tau$)
A small change $d\tau > 0$ above $\bar{z}$ has three effects:

1. A **mechanical increase** in tax revenue of $dM = N[z^m - \bar{z}]d\tau$

2. A **behavioral response**:

$$dB = N\tau dz^m = -N\tau \frac{dz^m}{d(1 - \tau)} d\tau$$

$$= -N \frac{\tau}{1 - \tau} \frac{1 - \tau}{z^m} \frac{dz^m}{d(1 - \tau)} z^m d\tau = -N \frac{\tau}{1 - \tau} \bar{\epsilon} z^m d\tau$$
3. **A welfare effect:** If government values marginal consumption of rich at \( \bar{g} \in (0, 1) \), \( dW = -\bar{g} \, dM \). Here \( \bar{g} \) depends on the curvature of \( u(c) \) and on the SWF.

- Summing the three terms:

\[
dM + dW + dB = dM(1 - \bar{g}) + dB = Nd\tau \left\{ (1 - \bar{g})[z^m - \bar{z}] - \bar{e} \frac{\tau}{1 - \tau} z^m \right\}
\]

- Optimal \( \tau \) is such that \( dM + dW + dB = 0 \) \( \Rightarrow \)

\[
\frac{\tau^*_\text{TOP}}{1 - \tau^*_\text{TOP}} = \frac{(1 - \bar{g})(z_m / \bar{z} - 1)}{\bar{e}(z_m / \bar{z})}
\]
Optimal Top Rate

- We can solve:
  \[ \tau^{\star}_{\text{TOP}} = \frac{(1 - \bar{g})(\frac{z_m}{\bar{z}} - 1)}{(1 - \bar{g})(\frac{z_m}{\bar{z}} - 1) + \bar{\epsilon} \frac{z_m}{\bar{z}}} \]

- \( \tau^{\star}_{\text{TOP}} \) decreases with \( \bar{g} \) and with \( \bar{\epsilon} \). It increases with \( z_m / \bar{z} \) (the thickness of top tail)

- Not an explicit formula since \( z_m / \bar{z} \) also a function of \( \tau \)
The top tail of the income distribution is closely approximated by a Pareto distribution. Density function is \( f(z) = a \cdot k^a / z^{1+a} \) where \( a > 1 \) is the Pareto parameter (also \( F(z) = 1 - k^a / z^a \)).

With a Pareto distribution the ratio \( z_m / \bar{z} \) is the same for all \( \bar{z} \) in the top tail and it is equal to \( a/(a - 1) \).

For the US, reasonable values are \( \bar{z} = 400,000 \) and \( z_m = 1.2 \) million, so that \( z_m / \bar{z} = 3 \) and \( a = 1.5 \). For France, \( a \) slightly higher between 2.2 and 2.3.
Since $z_m/\bar{z} = a/(a - 1)$, with a Pareto distribution the formula can be simplified into:

$$\Rightarrow \tau^*_{\text{TOP}} = \frac{1 - \bar{g}}{1 - \bar{g} + a \cdot \bar{e}}$$

Example: $\bar{e} = 0.5$, $\bar{g} = 0.5$, $a = 1.5 \Rightarrow \tau^*_{\text{TOP}} = 40\%$

In France, taking $\bar{e} = 0.5$, $\bar{g} = 0.5$, $a = 2.2$, we get $\tau^*_{\text{TOP}} = 31.25\%$
Top earner gets $z^T$, and second earner gets $z^S$

Then $z^m = z^T$ when $\bar{z} > z^S \Rightarrow z^m/\bar{z} \rightarrow 1$ when $\bar{z} \rightarrow z^T \Rightarrow$

$$dM = Nd\tau[z^m - \bar{z}] \rightarrow 0 < dB = Nd\bar{z}\epsilon \frac{\tau}{1 - \tau}z^m$$

Optimal $\tau$ is zero for $\bar{z}$ close to $z^T$

This result has little practical relevance since it applies to the top earner only
To find the revenue maximizing top tax rate we assign 0 weight to welfare of top incomes

- With an utilitarian SWF $\bar{g} = u_c(z^m) \rightarrow 0$ when $\bar{z} \rightarrow \infty$
- With a Rawlsian SWF $\bar{g} = 0$ for any $\bar{z} > \max(z)$

If $\bar{g} = 0$, $\tau_{\text{TOP}} = \tau_{\text{MAX}} = 1 / (1 + a\bar{\varepsilon})$

Example: $a = 2$ and $\bar{\varepsilon} = 0.5 \Rightarrow \tau = 50\%$. Result extremely sensible elasticity value ($a = 2$, $\bar{\varepsilon} = 2 \Rightarrow \tau = 20\%$)

The Laffer rate can be obtained as a special case with $\bar{z} = 0$

$$z^m / \bar{z} = \infty = a / (a - 1) \Rightarrow a = 1 \Rightarrow \tau_{\text{MAX}} = 1 / (1 + \bar{\varepsilon})$$
• Now consider general problem of setting optimal \( T(z) \)

• Let \( H(z) = \text{CDF of income [population normalized to 1]} \) and \( h(z) \) its density [endogenous to \( T(.) \)]

• Let \( g(z) = \text{social marginal value of consumption for taxpayers with income } z \) in terms of public funds

• Let \( G(z) \) be the \textbf{average} social marginal value of consumption for taxpayers with income above \( z \) \[ G(z) = \int_z^\infty g(s)h(s)\,ds/(1 - H(z)) \]
FIGURE 3 – Local Marginal Tax Rate Perturbation

Source: Saez 2001
Consider small reform: increase \( T' \) by \( d\tau \) in small interval \((z, z + dz)\)

- Mechanical revenue effect

\[
dM = dzd\tau(1 - H(z))
\]

- Before the change, an individual with income \( z + dz \) pays taxes \( \Omega + \tau dz \), where \( \Omega \) represents taxes paid by an individual with income \( z \)

- After the change, she will pay taxes \( \Omega + (\tau + d\tau)dz \). The difference is \( d\tau dz \)

- Mechanical welfare effect

\[
dW = -dzd\tau(1 - H(z))G(z)
\]
Behavioral effect: substitution effect $\delta z$ inside small band $[z, z + dz]$:

$$dB = h(z)dz \cdot T' \cdot \delta z = -h(z)dz \cdot T' \cdot d\tau \cdot \varepsilon(z) \cdot z / (1 - T')$$

This is because we can write:

$$\delta z = \frac{\delta z}{d(1 - T')} \frac{1 - T'}{z} \frac{d(1 - T')}{1 - T'} z = -\varepsilon(z) \frac{d\tau}{1 - T'} z$$

At the optimum $dM + dW + dB = 0$
General Non-Linear Income Tax

- Optimal tax schedule satisfies:

\[
\frac{T'(z)}{1 - T'(z)} = \frac{1}{\epsilon(z)} \left( \frac{1 - H(z)}{zh(z)} \right) [1 - G(z)]
\]

- \(T'(z)\) decreasing in \(g(z')\) for \(z' > z\) [redistributive tastes]

- \(T'(z)\) decreasing in \(\epsilon(z)\) [efficiency]

- \(T'(z)\) decreasing in \(h(z)/(1 - H(z))\) [density]

- To see connection with top tax rate: consider \(z \to \infty\)

  - \(G(z) \to \bar{g}, (1 - H(z))/(zh(z)) \to 1/a\)

  - \(\epsilon(z) \to \bar{\epsilon} \Rightarrow T'(z) = (1 - \bar{g})/(1 - \bar{g} + a \cdot \bar{\epsilon}) = \tau_{\text{TOP}}\)
Negative Marginal Tax Rates Never Optimal

- Suppose $T' < 0$ in band $[z, z + dz]$

- Increase $T'$ by $d\tau > 0$ in band $[z, z + dz]$
  
  - $dM + dW > 0$ because $G(z) < 1$ for any $z > 0$ (with declining $g(z)$ and $G(0) = 1$)
  - $dB > 0$ because $T'(z) < 0$ [smaller efficiency cost]

- Therefore $T'(z) < 0$ cannot be optimal

- Marginal subsidies also distort local incentives to work

- Better to redistribute using lump sum grant
Numerical Simulations of Optimal Tax Schedule

- Formula above is a condition for optimality but not an explicit formula for optimal tax schedule. Distribution of incomes $H(z)$ endogenous to $T(.)$

- Therefore need to use structural approach (specification of primitives) to calculate optimal $T(.)$

- Saez (2001) uses two utility functions:

  $u(c, l) = \log(c - \frac{l}{1 + k}^{1+k})$

  $u(c, l) = \log(c) - \log\left(\frac{l}{1 + k}^{1+k}\right)$

- With the first, no income effects and $\varepsilon = 1/k$. With the second, $\varepsilon^C = 1/k$, and $\varepsilon^U$ goes to zero as $z$ goes to infinity.

- Calibrate the exogenous skill distribution $F(w)$ such that actual $T(.)$ yields empirical $H(z)$
Utilitarian criterion, utility type I

Source: Saez 2001
Results

- Dashed line: optimal **linear** tax rate

- Optimal tax rates are clearly U-shaped:
  - Decreasing from $0 to $75,000 and then increase until $200,000
  - Above $200,000, optimal rates close to asymptotic level
Diamond (1988)

- No income effects: \( u(c, l) = c + V(1 - l) \)
- Constant labor supply elasticity
- SWF \( G(u) \) increasing and strictly concave
- Ability distribution is Pareto above the mode \( (w_m) \)
- At least above \( w_m \), marginal tax rates should be increasing in \( w \)
- Not robust if utility function is \( u(c, l) = U(c) + V(1 - l) \), with \( U \) concave: Dahan and Strawczynski (2000)
Most governments tax both income and goods.

In Spain, differential commodity taxes with non-linear income tax

1. VAT has three rates: 18% (general), 8% and 4%. 8% applies to food, health, transportation. 4% to necessities, books,..

2. There are additional excise taxes (“impuestos especiales”) on some goods (alcohol, gasoline, cigarettes)

3. Imposes capital income taxes

What is the best combination of taxes?
$K$ goods $c = (c_1, \ldots, c_K)$ with pre-tax price $p = (p_1, \ldots, p_K)$

Individual $h$ has utility $u^h(c_1, \ldots, c_K, z)$

Suppose there is a nonlinear income tax on earnings $z$. Can government raise welfare using commodity taxes $t = (t_1, \ldots, t_K)$?

More instruments cannot hurt:

$$\max_{t, T(\cdot)} SWF \geq \max_{t=0, T(\cdot)} SWF$$
Atkinson and Stiglitz (1976) show that:

\[
\max_{t, T(.)} SWF = \max_{t=0, T(.)} SWF
\]

Commodity taxes not useful if \( u^h(c_1, ..., c_K, z) \) satisfy 2 conditions:

1. Weak separability between \((c_1, ..., c_K)\) and \(z\) in utility

2. All individuals have the same sub-utility function of consumption:

\[
u^h(c_1, ..., c_K, z) = U^h(\nu(c_1, ..., c_K), z)\]
If tax system is too progressive (relative to the optimum), a high ability individual will choose to work less to mimic income level \(z = wl\) of a lower ability individual.

Both pay same income tax and have the same after-tax income, but the high ability individual has more leisure. Government cannot use income tax to redistribute.

What about differentiated good taxes? Is it possible to reduce the incentive to mimic?

No. With separability and homogeneity, conditional on earnings \(z\), consumption choices \(c = (c_1, \ldots, c_K)\) do not provide information on ability. Two individuals with the same \(z\), but different abilities, choose the same consumption bundle.
Differentiated commodity taxes $t_1, \ldots, t_K$ create a tax distortion with no benefit.

It is optimal to have uniform taxation. In particular, zero taxation of goods.

It is better to do all the redistribution with the individual income tax.

With only linear income taxation (Diamond-Mirrlees 1971, Diamond 1975), differential commodity taxation can be useful to “non-linearize” the tax system.
A consequence of uniform taxation is that we should not subsidize particular goods (housing, health) for the purpose of redistribution.

Another implication is that even a government who would like to redistribute should not do so by taxing luxury goods more than basic goods.

A striking result, although it has limited practical relevance.
Atkinson-Stiglitz: Criticisms

- First, the result depends on a technical condition, weak separability. Without it, better to tax more heavily goods which are complementary to leisure.

- Second, the theorem requires that all individuals must have the same preferences. With differences in tastes, the result is not longer true (Saez (2002) in the Journal of Public Economics).

- Now the optimal non-linear income tax should be supplemented not only by taxes on commodities that are complementary to leisure, but also by taxes on commodities for which high-income people have a relatively strong taste.
In spite of the above, there are still some arguments in favor of uniform taxation:

- Government can lack information on compensated cross-price elasticities needed to compute optimal taxes.
- Changes in tastes and technologies may require frequent adjustments in taxes.
- A uniform VAT is easier to administer and less susceptible to fraud.
- Government would eliminate incentives for special interest groups to lobby for low tax rates on particular goods.
Should we tax capital income?

- The Atkinson-Stiglitz theorem has the implication that capital should not be taxed: since a capital tax is a tax on future consumption but not on current consumption, it violates the prescription of uniform taxation.
- Not only this, since a capital tax imposes an ever-increasing tax on consumption further in the future, its violation of the principle of uniform taxation is extreme.
- Consider a simple life-cycle model. Consumer supplies $L$ labor when young and enjoys consumption $C_1$ (when young) and $C_2$ (when old).
- Taxes on labor are $T(wL)$, where $w$ is pre-tax wage and the return to capital ($r$) is taxed at rate $t_r$.
- Finally, we denote savings when young by $S$. 

The constraints are:

\[ S = wL - T(wL) - C_1 \quad \text{and} \quad C_2 = [1 + r(1 - t_r)]S \]

We can combine them into:

\[ C_1 + \frac{C_2}{1 + r(1 - t_r)} = wL - T(wL) \]

The term on the left can be written as:

\[ C_1 + \left[ \frac{1}{1 + r} + \frac{rt_r}{(1 + r)(1 + r(1 - t_r))} \right] C_2 \]
• A tax on capital income is like an excise tax on future consumption, because it raises the relative price of future consumption above $1/(1+r)$, which is the price without taxes on capital income.

• If total time to 1, leisure when young is $1-L$.

• Consider that lifetime utility $U(1-L, C_1, C_2)$ can be written as:

$$U(1-L, C_1, C_2) = U(1-L, v(C_1, C_2))$$

• According to Atkinson-Stiglitz, if we interpret $C_1$ and $C_2$ as two different commodities, they should be taxed uniformly, provided that the government is using the income tax for redistributional purposes.
Interpretation

- Without precise information about the degree of complementarity or substitutability between leisure and $C_1$ and $C_2$, a possibility is to consider that both of them are equally complements with leisure.

- This implies that the marginal effective tax rate on capital income should be zero.

- However, if leisure and future consumption are complements, a positive capital income tax would be optimal. Empirically, leisure seems to be increasing with age. Then, tax rates on capital income should be positive.
In the standard neoclassical growth model with infinitely-lived individuals, an optimal income tax policy entails setting capital income taxes equal to zero in the long run. This is the classical Chamley-Judd result.

Intuition: assume that pre-tax return of capital in the steady state is $r$, and there is a tax rate on capital $\tau$.

Capital taxation changes the relative price of consumption at date $t$ and at date $t + T$ by a factor:

$$\left( \frac{1 + r}{1 + r(1 - \tau)} \right)^T$$

Without taxes, the relative price of $C_{t+T}$ in terms of $C_t$ is $\left( \frac{1}{1+r} \right)^T$. With taxes it is $\left( \frac{1}{1+r(1-\tau)} \right)^T$. 
If $\tau = 0$, that factor is 1

If $\tau \neq 0$, when $T$ goes to infinity the factor goes either to zero (when $\tau < 0$) or to $+\infty$ (when $\tau > 0$)

Taxing (or subsidizing) capital would distort in an explosive way the choice between current and future consumption

This result does not require any separability assumption

However, the result does not hold in OLG models: See Erosa and Gervais (2002) and Conesa, Kitao and Krueger (2009). The latter use a OLG models where households face borrowing constrains and are subject to idiosyncratic income shocks. They calibrate their model to the US economy and obtain an optimal tax rate on capital income of 36% (and a flat tax of 23% on labor income, together with a $7,200 deduction)
Key assumption in Chamley-Judd and Atkinson-Stiglitz results: people optimize their savings decisions

Recent evidence challenges this assumption

Madrian and Shea (2001) study employee 401(k) enrollment decisions and contribution rates at a U.S. corporation:

- Most people adhere to company defaults and do not make active savings choices
- Suggests that defaults may have much bigger impacts on savings decision than net-of-tax returns
Labor supply elasticity is estimated to be close to zero for males (higher for married women). This implies that efficiency cost of taxing labor income is expected to be low.

However, taxes trigger many other behavioral responses: changes in the form of compensation (e.g., employer-paid health insurance); moving away from taxed goods or activities to those that are taxed more lightly or not taxed at all; trying to avoid taxes by shifting income intertemporally (for instance, diverting income into a tax-deferred account); tax evasion.

The elasticity of taxable income (ETI) tries to capture all these: it measures how reported taxable income (TI) responds to tax changes. The bottom line is that the efficiency cost of taxation can be much higher than is implied if labor supply is the only margin of behavioral response.
Most empirical studies have found that the behavioral response to tax changes concentrates on top earners. We focus on them.

As above, we assume that all incomes above an income threshold $\bar{z}$ face a constant marginal tax rate $\tau$ and we assume the number of top earners is $N$.

The optimal tax on top earners (assuming the weight that government attaches to them is zero) is:

$$\tau_{\text{TOP}}^* = \frac{1}{1 + a \varepsilon},$$

where $a$ is the Pareto parameter and $\varepsilon$ is the elasticity of top earners.
The key ingredient is the elasticity $\varepsilon$ of top earners. If, for instance, $\varepsilon = 0.25$ and $a = 1.5$, $\tau^*_{\text{TOP}} = 0.73$, substantially higher than the current 42.5 percent top US marginal tax rate (all taxes combined).

The optimal value would be 42.5% if $\varepsilon$ is extremely large, around 0.9 (or if the marginal consumption of very high income earners is highly valued, $g = 0.72$).

A large literature has focused on the response of reported income, either “adjusted gross income” or “taxable income” to net-of-tax rates.

Found large and quick responses of reported incomes along the tax avoidance margin at the top of the distribution, smaller effects on real economic responses.
For example, in the US exercises of options surged in 1992 before the 1993 top rate increase took place.

Tax Reform Act of 1986 led to a shift from corporate to individual income because it became more advantageous to be organized as a business taxed solely at the individual level rather than as a corporation taxed first at the corporate level.

With tax avoidance or evasion opportunities, the tax base in a given year is very sensitive to tax rates. The optimal tax rate is low. But...

If avoidance channels is re-timing or income shifting, there are “tax externalities”: changes in tax revenue in other periods or other tax bases, optimal taxes must not decrease.

Also the tax avoidance or evasion component of $\varepsilon$ can be reduced through base broadening and tax enforcement.
Gruber and Saez (2002) find a substantial taxable income elasticity estimate ($\varepsilon = 0.57$) at the top. However, they also find a small elasticity ($\varepsilon = 0.17$) for income before any deductions.

With the first estimate, $\tau^*_{TOP} = 0.54$ but with the second $\tau^*_{TOP} = 0.80$.

Taking as fixed state and payroll taxes, such rates correspond to top federal rates of 0.48 and 0.76, respectively.

In any case, there is some agreement that $\varepsilon = 0.57$ is a conservative upper bound.
Example: Income shifting

- Taxable income can be also reduced because individuals shift from taxable compensation toward untaxed fringe benefits. Or from individual income toward corporate income; deferred compensation, etc.
- Assume that a fraction $s < 1$ of incomes that disappear from TI after an increase in taxes $d\tau$ are shifted to corporate income that is taxed on average at rate $t < \tau$
- The tax that maximizes revenue is:

$$\tau_s^* = \frac{1 + a\varepsilon ts}{1 + a\varepsilon} > \tau_{TOP}^* = \frac{1}{1 + a\varepsilon}$$

- If $a = 1.5$, $\varepsilon = 0.5$, $\tau = 0.35$, $s = 1/2$ and $t = 0.3$, $\tau_s^* = 0.636 > \tau_{TOP}^* = 0.57$
Until now, we assume agents respond to tax changes in the same way they react to price changes.

Suppose that for some commodity the producer price is 10 euros and there is a tax of 1 euro per unit, so final consumer price is 11 euros. An increase in producer price of 50 cents has the same effect on purchases than an increase in 50 cents in the tax.

Chetty-Looney-Kroft challenge this. They find that “salience” (or visibility) of taxes affect consumers’ purchase decisions, because consumers are inattentive to taxes.

For example, sometimes taxes on commodities are not displayed in posted prices.

They find that commodity taxes that are included in posted prices that consumers see when shopping have larger effects on demand.
Two goods, called $x$ and $y$, are supplied perfectly elastically. Price of $y$ is 1 and we call $p$ the pre-tax price of $x$.

Good $y$ is untaxed and $x$ pays sales tax at rate $\tau^S$ so total price is $q = (1 + \tau^S)p$.

Consumers see $p$ when shopping because sales tax is not included in the posted price. They have to calculate $q$, making $q$ less “salient” than $p$.

Call $x(p, \tau^S)$ the demand as a function of posted price and sales tax. With rational shoppers, demand depends only on total price: $x(p, \tau^S) = x((1 + \tau^S)p, 0)$.

With rational consumers, a 1% increase in $p$ and a 1% increase in $(1 + \tau^S)$ reduce demand by exactly the same amount. If $\varepsilon_{x,p}$ and $\varepsilon_{x,1+\tau^S}$ are the elasticities with respect to $p$ and $(1 + \tau^S)$:

$\varepsilon_{x,p} = \varepsilon_{x,1+\tau^S}$.
In contrast, ChLK propose consumers underreact to the tax $\tau^S$ because it is less salient: $\varepsilon_{x,p} > \varepsilon_{x,1+\tau^S}$.

Log-linearizing demand $x(p, \tau^S)$:

$$\log x(p, \tau^S) = \alpha + \beta \log p + \theta_r \beta \log (1 + \tau^S)$$

The parameter $\theta_r$ captures the degree to which consumers underreact to the tax. In particular:

$$\theta_r = \frac{\partial \log x / \partial \log (1 + \tau^S)}{\partial \log x / \partial \log p} = \frac{\varepsilon_{x,1+\tau^S}}{\varepsilon_{x,p}}$$

The null hypothesis is $\theta_r = 1$. They propose two strategies to estimate $\theta_r$. 


They conducted an experiment in a supermarket chain for three weeks. Stores do not include the sales tax in the price tag on the shelves for taxable goods like cosmetics. Taxes only appear on the sales slip after paying at the cash register. ChLK adjusted the price tags to display prices including the 7.375% sales tax. The result was a decline in sales of those items by about 8% relative to the control groups. Making visible the sales tax reduce purchases, suggesting that most consumers do not normally take into account the sales tax on such products. They estimate a value of $\hat{\theta}_r \approx 0.35$. That is, there is substantial underreaction.
In a second test they used data over a long period of time of alcohol purchases to compare the effect of price changes with tax changes.

Alcohol is subject to an excise tax (salient) and the less-salient sales tax that appears only at the cash register.

They found a much larger effect of increases in the excise tax than to increases in the sales tax. They obtain an estimate of $\hat{\theta}_r \approx 0.06$.

The explanation by ChLK is not that consumers are not aware of the existence of taxes, but that they simply do not bother to compute tax-inclusive prices.
Policy implications

- Standard theory predicts that taxes create an efficiency cost only to the extent that it reduces demand for the taxed good.
- Here, even if the demand for the taxed good does not change, taxes may have a substantial efficiency cost. The reason is that the tax can distort consumption of other goods. A consumer that is not attentive to taxes on some good may end up spending too much on it and, thus, will have less money to spend on other goods.
- For instance, an individual who does not take into account taxes on cars, would over-spend on the car and end up with less money left than he would like for food, reducing economic welfare.
Another implication is that now statutory incidence matters: suppose a phone has $39.99 in the price tag, but adding taxes and costumer fees, actual price is $47.00. If taxes were levied by the firm, the only way the firm can pass taxes to costumers is by raising prices in the price tags, reducing demand.

Firms will bear a higher burden if the tax is levied on them rather than on consumers.