Labor income Taxation with labor supply responses along intensive and/or extensive margin

Laurence Jacquet

THEMA, Oslo Fiscal Studies

September 2016

1. Actual tax systems
2. Optimal nonlinear taxation with intensive margin
   - Labor supply responses to change in the marginal tax rate(s)
   - Mirrlees (ReStud 1971): setting
   - First-best allocations
   - Second-best allocations
   - Derivation of optimal tax rates: 2 skill levels
   - Derivation of optimal tax rates: continuum of skills
   - Derivation of optimal tax rates using a tax perturbation
   - Marginal tax rates for interior skill levels
   - Marginal tax rates at the top
   - Migration effects on marginal tax rates at the top
   - Marginal tax rates at the bottom
3. Optimal nonlinear taxation with extensive margin
4. Optimal nonlinear taxation with both margins
   - Derivation of tax rates
   - Tax perturbation method
   - Method to sign marginal tax rates

- Optimal tax profiles with US data

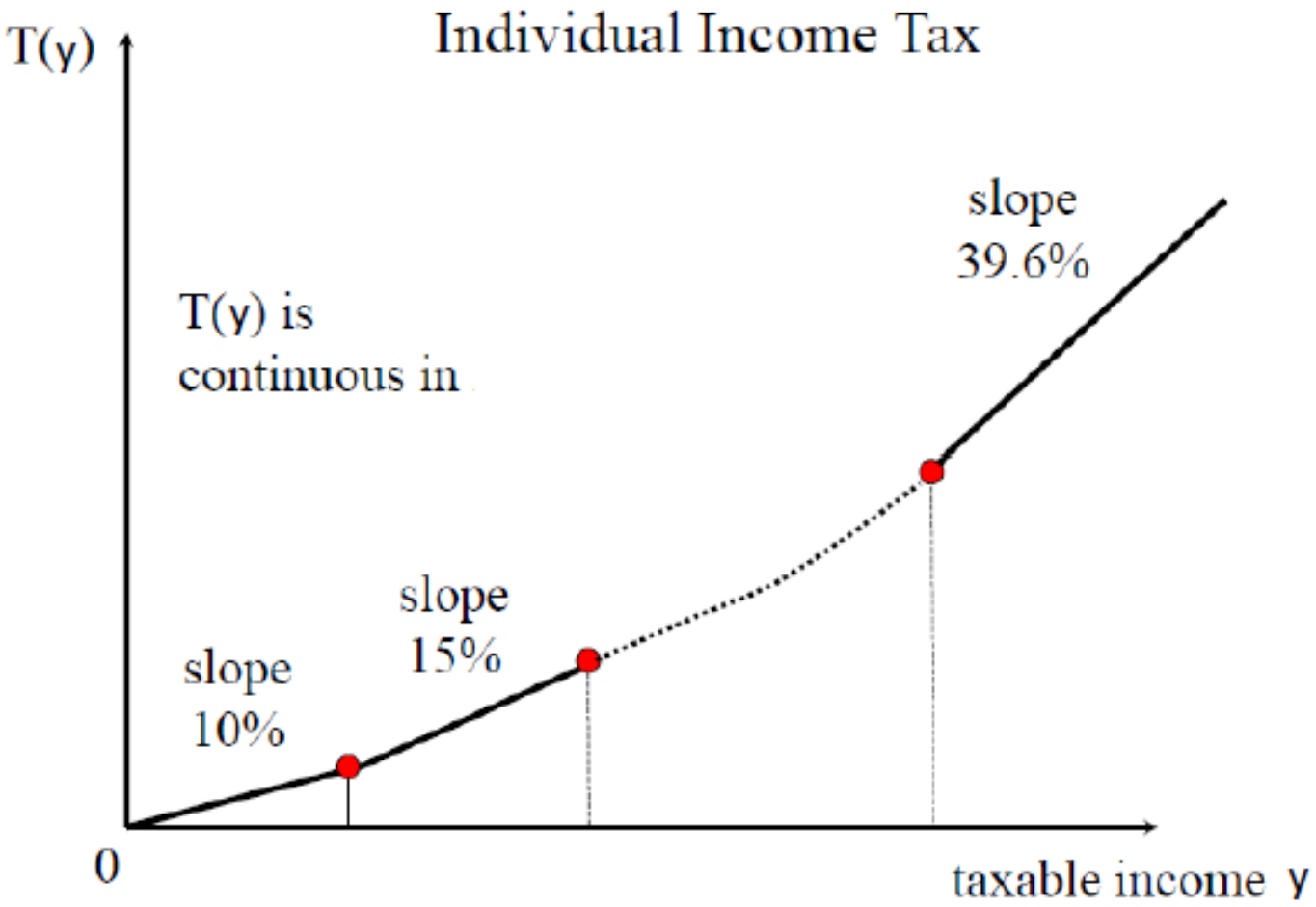
5 Reference

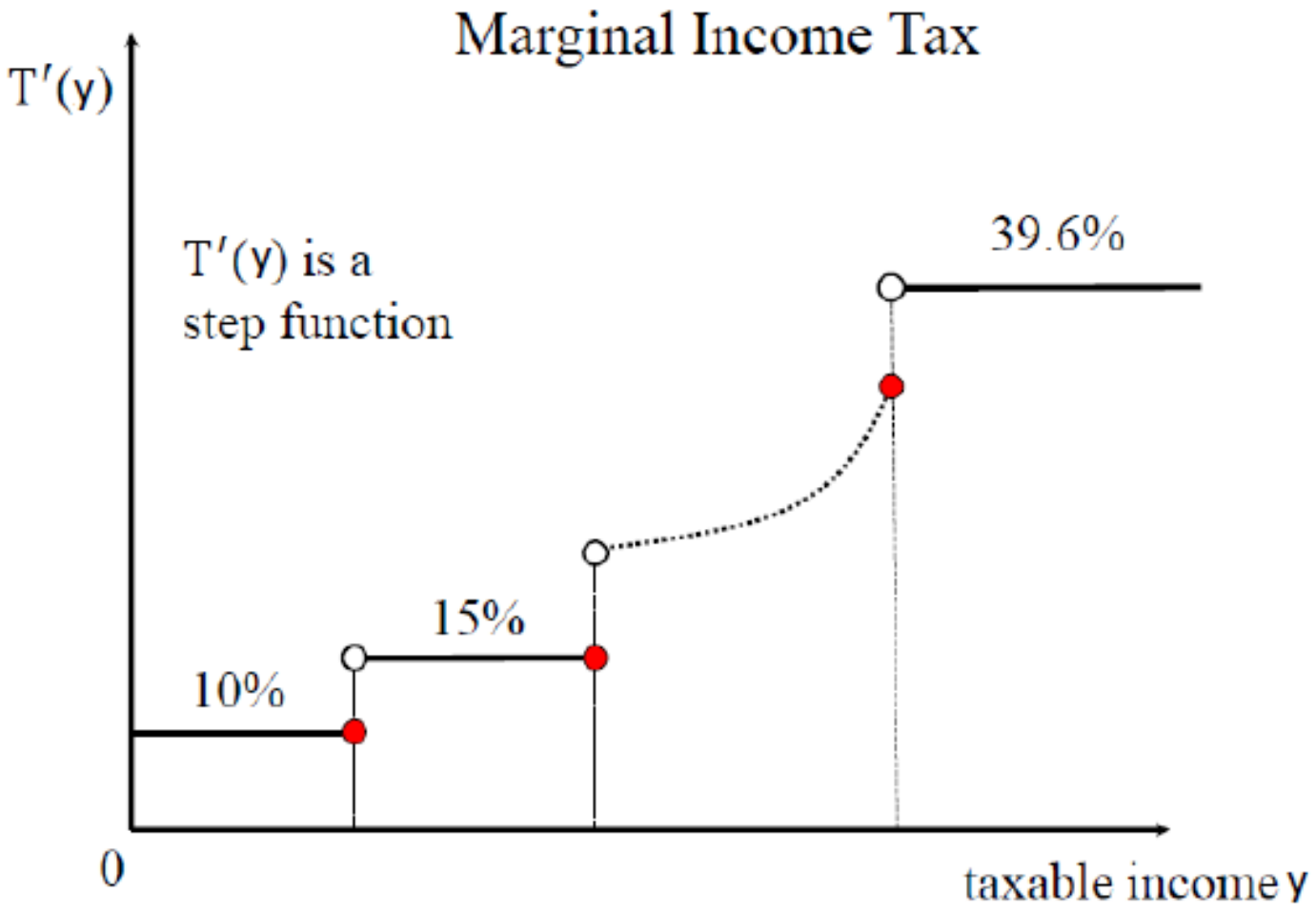
Outline

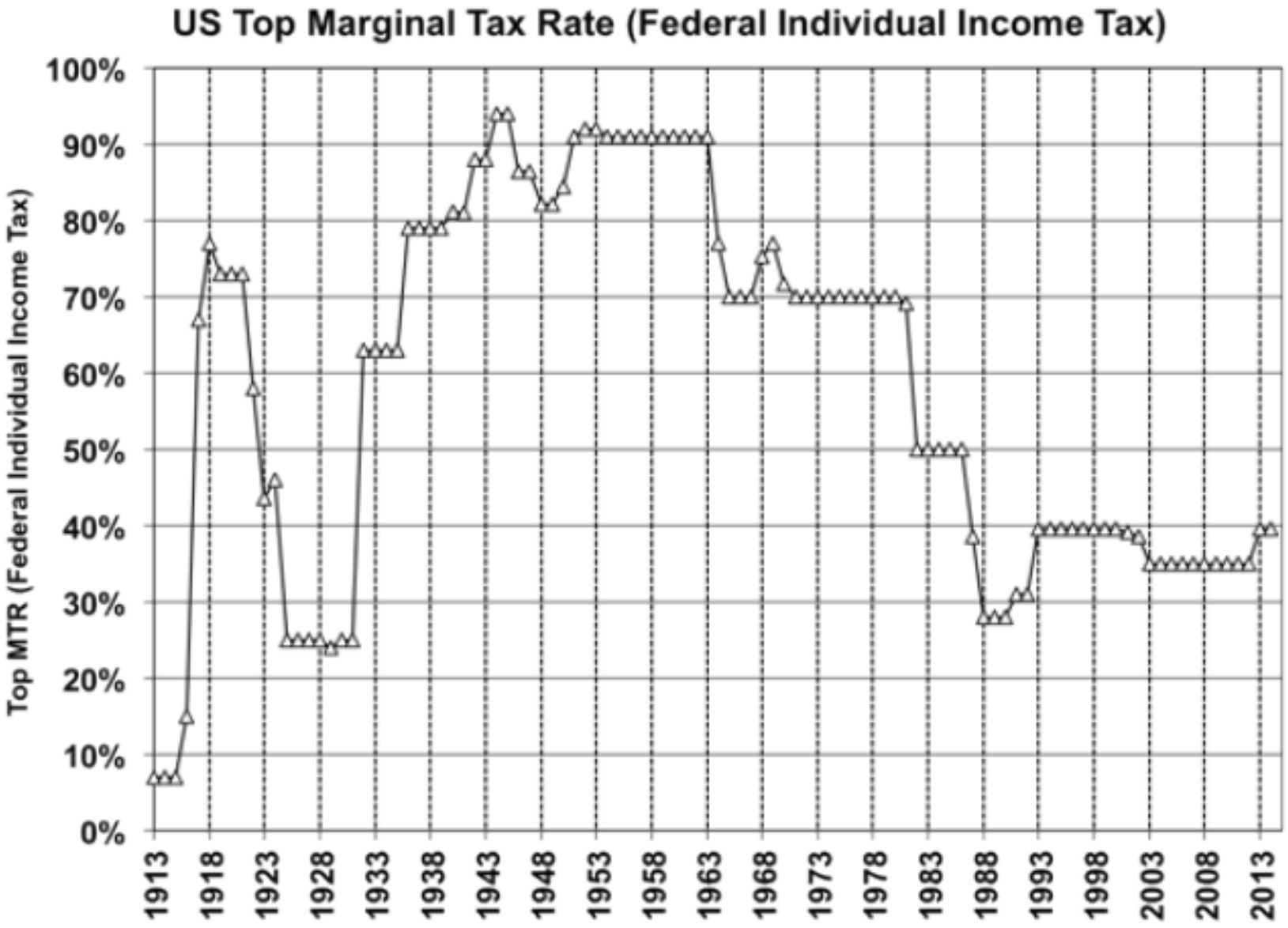
Tax brackets in the (federal) US income tax:

- Tax $T(y)$ is piecewise linear and continuous function of taxable income y with constant marginal tax rates (MTR) $T'(y)$ by brackets.
- In 2013+, 6 brackets with MTR 10%, 15%, 25%, 28%, 33%, 35%, 39.6% (top bracket for y above \$470K), indexed on price inflation.
- Tax rates change frequently over time. Top MTRs have declined drastically since 1970s, in the US as in many OECD countries.
- For example, the US top marginal federal individual tax rate stood at 91% in the 1950–1960s but is only 35% today.

Main reference for this section: Piketty and Saez (2013).


Individual Income Tax
T(y)
T(y) is
continuous in
slope
39.6%
slope
15%
slope
10%
0
taxable income y


Marginal Income Tax
T′(y)
T′(y) is a
step function
39.6%
15%
10%
0
taxable income y



Source: IRS, Statistics of Income Division, Historical Table 23.

Top marginal tax rates

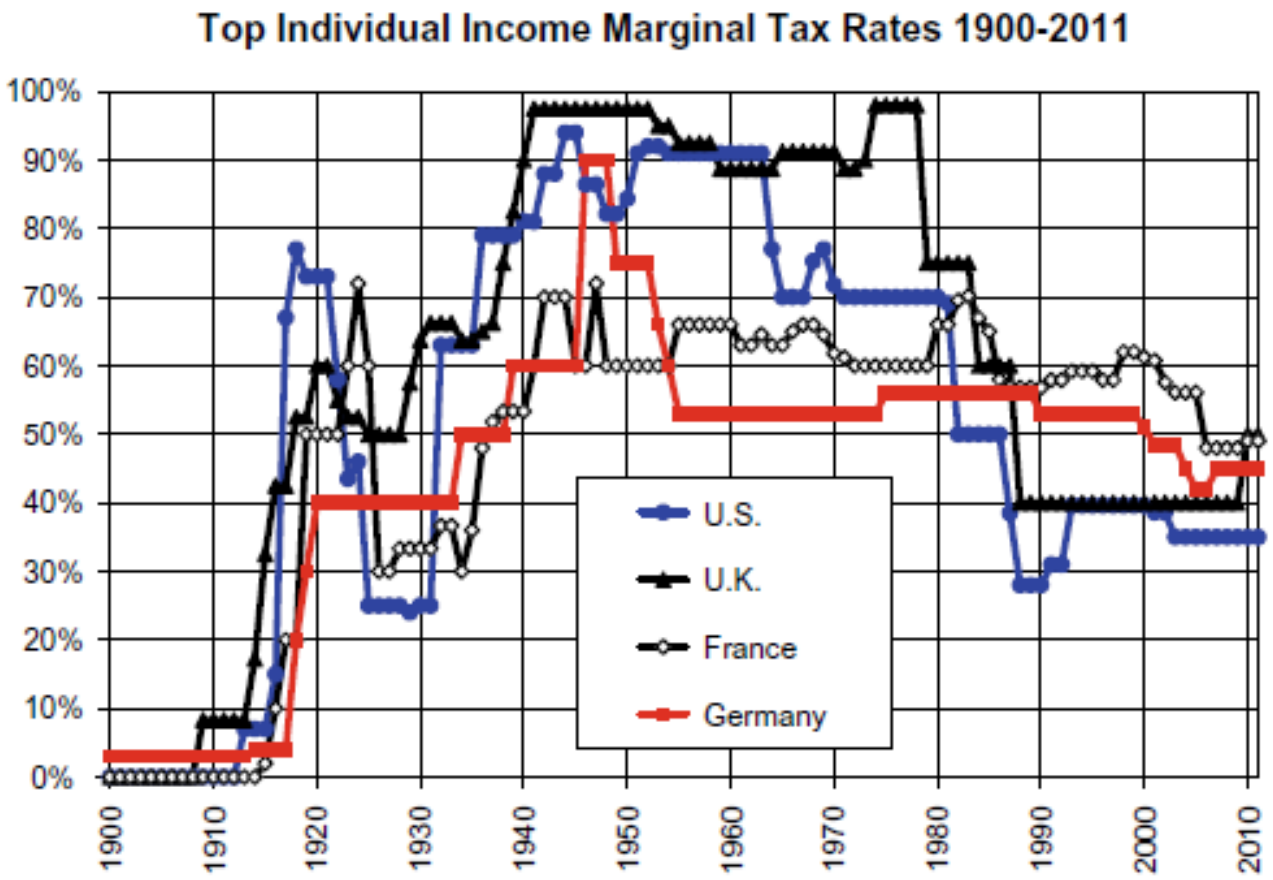


Figure 1 Top Marginal income tax rates in the US, UK, France, Germany. This figure, taken from Piketty et al. (2011), depicts the top marginal individual income tax rate in the US, UK, France, Germany since 1900. The tax rate includes only the top statutory individual income tax rate applying to ordinary income with no tax preference. State income taxes are not included in the case of the United States. For France, we include both the progressive individual income tax and the flat rate tax "Contribution Sociale Généralisée."

- When progressive income taxes were instituted (around 1900–1920) in most developed countries, top rates were very small (typically less than 10%). They rose very sharply in the 1920–1940s, particularly in the US and in the UK.
- Since the late 1970s: top tax rates on upper income earners have declined significantly in many OECD countries.

Transfers and public goods

The rise in taxes has been used to fund growing public goods and social transfers in:

- education,
- health care,
- retirement and disability,
- income security

Transfers and public goods

Table 1 Public Spending in OECD Countries (2000–2010, Percent of GDP)

	US *(1)*	Germany *(2)*	France *(3)*	UK *(4)*	Total OECD *(5)*
Total public spending	**35.4%**	**44.1%**	**51.0%**	**42.1%**	**38.7%**
Social public spending	**22.4%**	**30.6%**	**34.3%**	**26.2%**	**25.1%**
Education	4.7%	4.4%	5.2%	4.8%	4.9%
Health	7.7%	7.8%	7.1%	6.1%	5.6%
Pensions	6.0%	10.1%	12.2%	4.8%	6.5%
Income support to working age	2.7%	3.9%	4.8%	4.9%	4.4%
Other social public spending	1.3%	4.4%	5.1%	5.7%	3.7%
Other public spending	**13.0%**	**13.5%**	**16.7%**	**15.9%**	**13.6%**

Notes and sources: OECD Economic Outlook 2012, Annex Tables 25–31; Adema et al., 2011, Table 1.2; Education at a Glance, OECD 2011, Table B4.1. Total public spending includes all government outlays (except net debt interest payments). Other social public spending includes social services to the elderly and the disabled, family services, housing and other social policy areas (see Adema et al., 2011, p.21). We report 2000–2010 averages so as to smooth business cycle variations. Note that tax to GDP ratios are a little bit lower than spending to GDP ratios for two reasons: (a) governments typically run budget deficits (which can be large, around 5–8 GDP points during recessions), (b) governments get revenue from non-tax sources (such as user fees, profits from government owned firms, etc.).

Income security programs

- Income security programs include unemployment benefits + an array of means-tested transfers (both cash and in-kind).
- They are a relatively small fraction of total transfers. Typically $< 5\%$ of GDP (see other public spending in the previous table), out of a total around $25 - 35\%$ of GDP for social spending as a whole.

Income security programs

- Targeted to lower income individuals.
- The most redistributive component of the transfer system.
- Often, they are in-kind benefits such as subsidized housing, subsidized food purchases (e.g., food stamps and free lunches at school in the US).
- They often target special groups such as the unemployed (unemployment insurance), the elderly or disabled with no resources (for example Supplemental Security Income in the U.S.).
- Means-tested cash transfer programs for "able bodied" individuals are only a small fraction of total transfers.

To a large extent, the rise of the modern welfare state is the rise of universal access to "basic goods" (education, health, retirement and social insurance), and not the rise of cash transfers (see e.g., Lindert, 2004).

Bottom line on actual taxes/transfers

- Based on current income, family situation, and disability (retirement) status $\Rightarrow$ Strong link with current ability to pay.
- Provisions pile up overtime making tax/transfer system more and more complex until significant simplifying reform happens (e.g., US Tax Reform Act of 1986).
- Some allowances made to reward / encourage certain behaviors: charitable giving, home ownership, savings, energy conservation, **and more recently work** (refundable tax credits such as EITC).

Mean-tested transfers versus in-work transfers

- In recent years, traditional means-tested cash welfare programs have been partly replaced by in-work benefits.
- The shift has been particularly large in the US and the UK.

Mean-tested transfers

Means-tested programs are **L-shaped with income**.

i.e. the largest benefits to those with no income and those benefits are then phased-out at high rates for those with low earnings.

- Such a structure concentrates benefits among those who need them most.
- These phase-outs discourage work as they create large implicit taxes for low earners.

In-work transfers

In-work benefits are **inversely U-shaped.**

i.e. they first rise and then decline with earnings.

- Benefits are nil for those with no earnings and concentrated among low earners before being phased-out.
- Such a structure encourages work but fails to provide support to those with no earnings, arguably those most in need of support.

Budget set

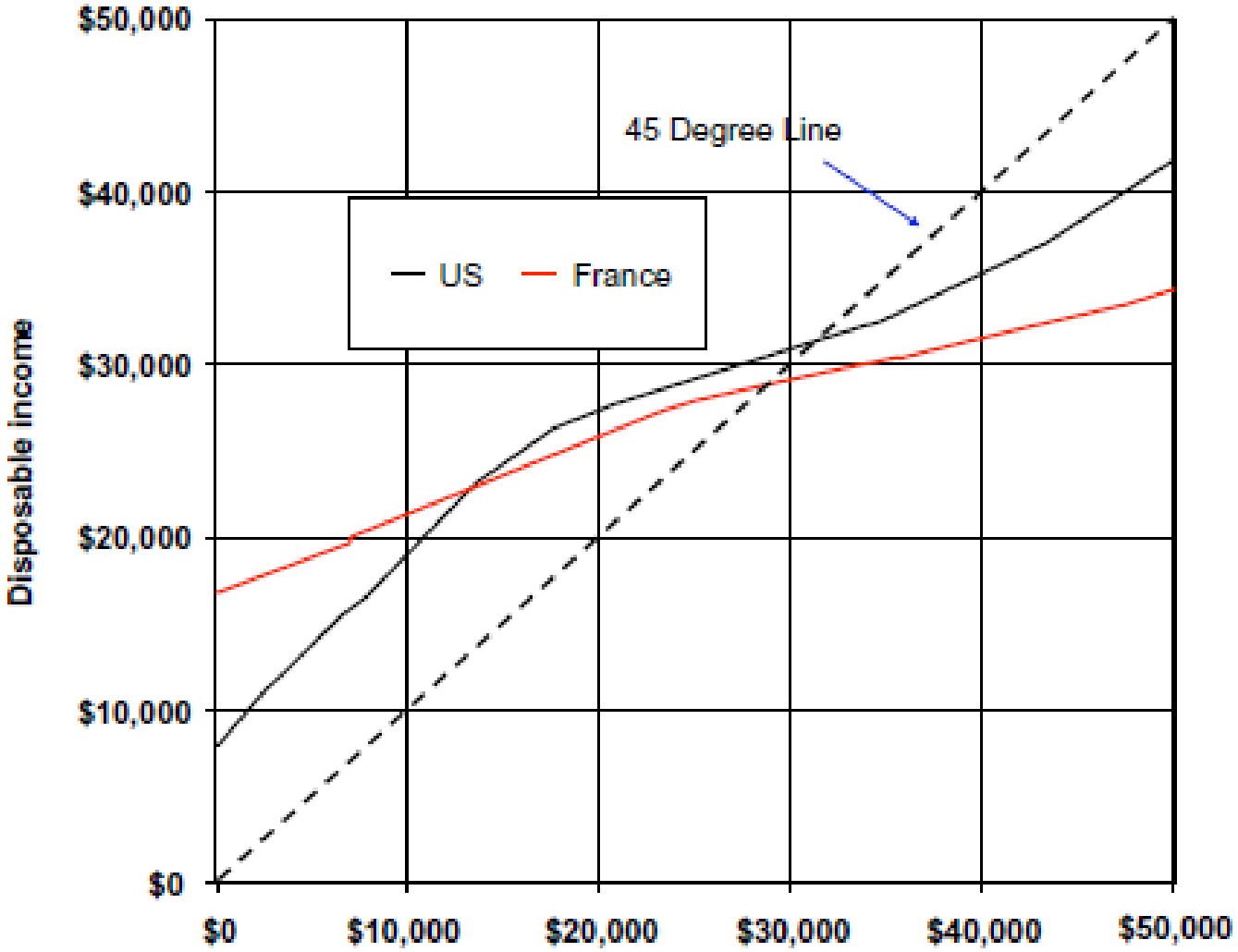
The budget set relating pre-tax and pre-transfers earnings to post-tax post-transfer disposable income summarizes the net impact of the tax and transfer system.

The next figure of **budget sets in France and US** includes:

- all payroll taxes and the income tax.
- means-tested transfer programs (Temporary Assistance for Needy Family and Food Stamps in the US, and the minimum income—RSA for France)
- tax credits (the Earned IncomeTax Credit and the Child Tax Credit in the United States, in-work benefit Prime pour l'Emploi and cash family benefits in France).

Budget set

Budget set for a single parents with 2 children in France and the US:



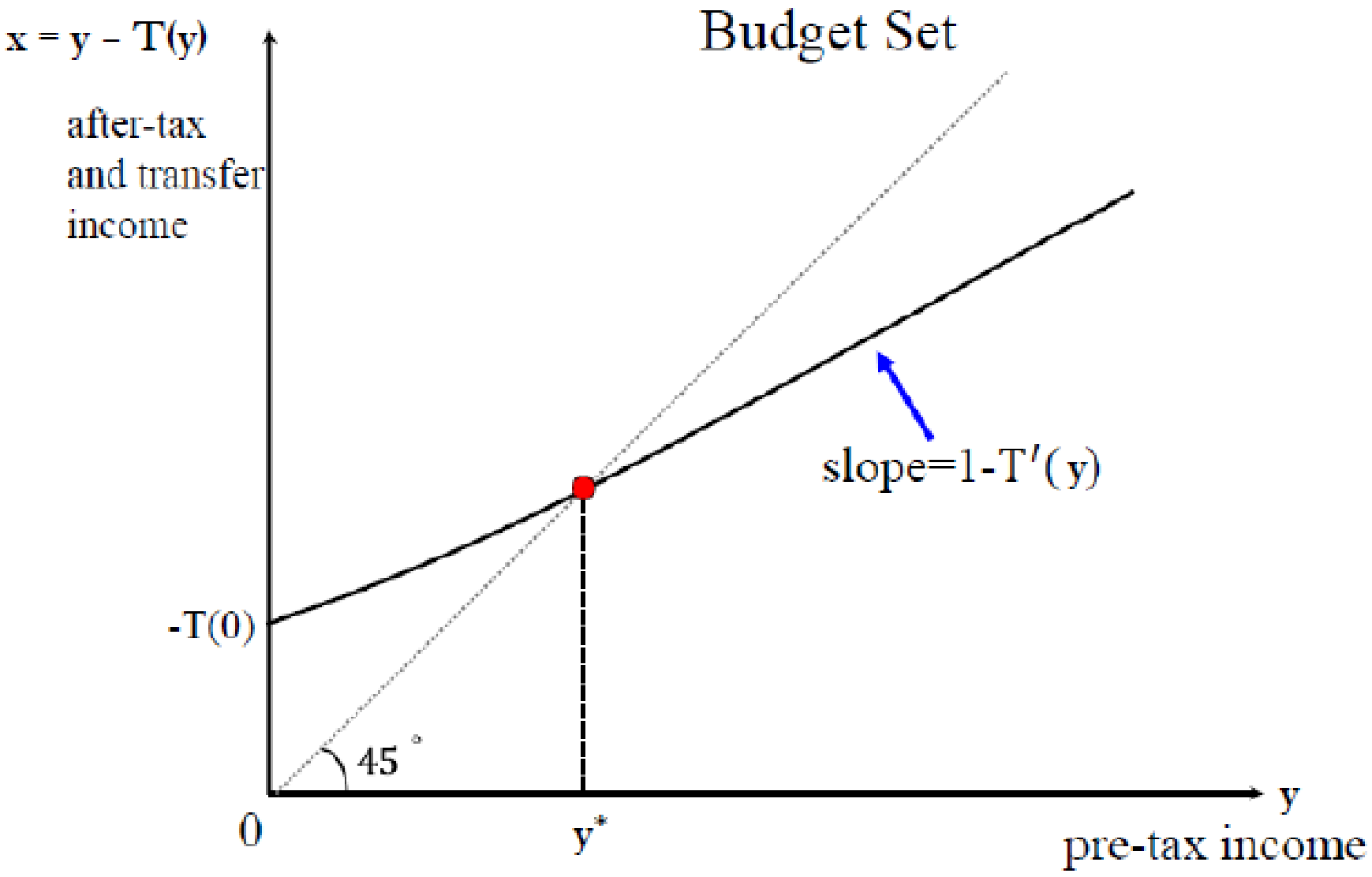
$\Rightarrow$ France offers more generous support to single parents with no earnings but the French tax and transfer system imposes higher implicit taxes on work.

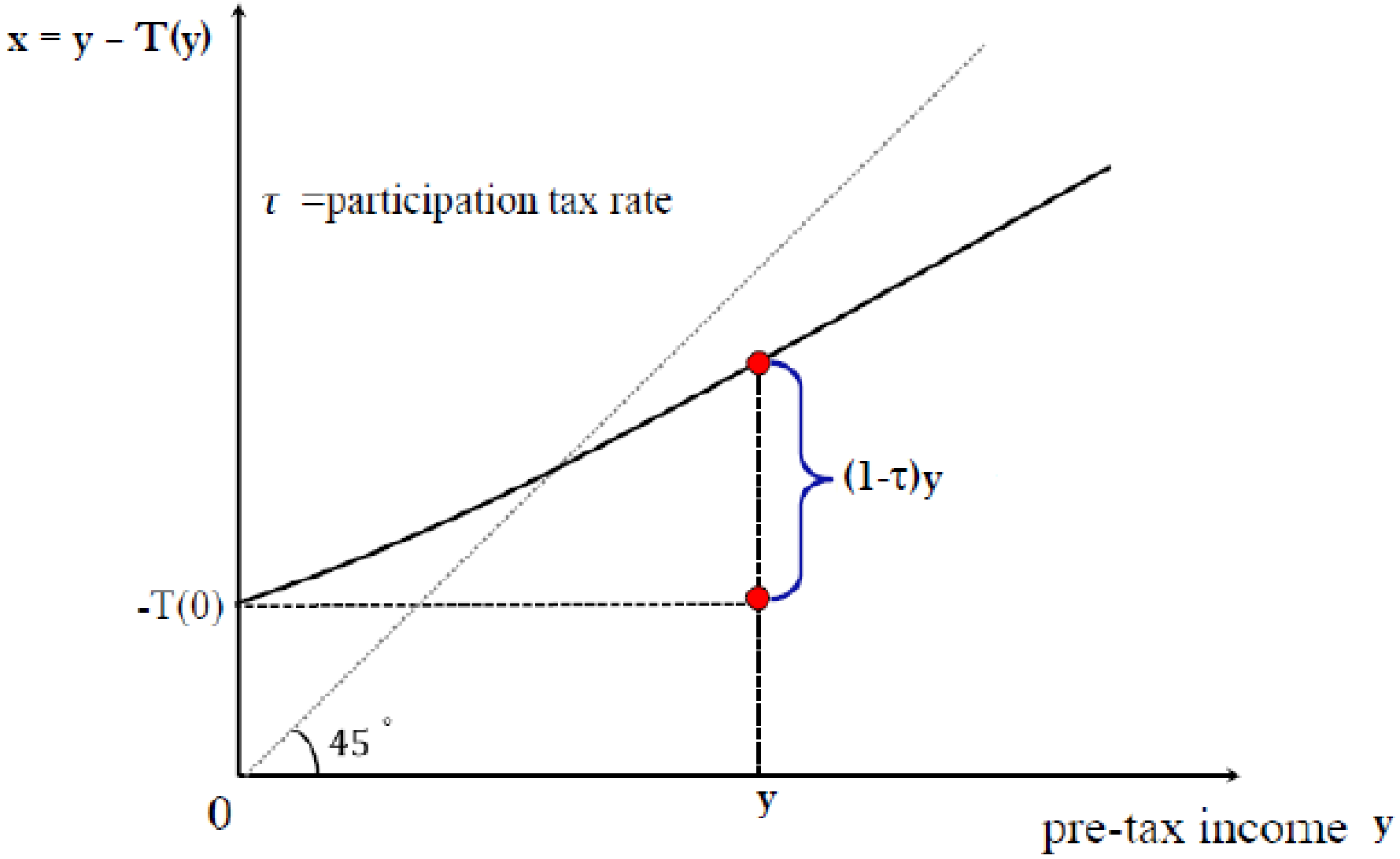
Key concepts for taxes/transfers

- Marginal tax rate (or phasing-out rate) $T'(y)$: individual keeps $1 - T'(y)$ for an additional \$1 of earnings (intensive labor supply response)
- Transfer benefit with zero earnings $-T(0)$ sometimes called basic income, demogrant or lumpsum grant.
- Participation tax rate $\tau = [T(y) - T(0)]/y$: individual keeps fraction $1 - \tau$ of earnings when moving from zero earnings to earnings y (extensive labor supply response);

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

- Break-even earnings point y^*: point at which $T'(y^*) = 0$.


Budget Set
x = y – T(y)
after-tax
and transfer
income
slope=1-T′(y)
-T(0)
45 °
0
y*
y
pre-tax income


x = y – T(y)
τ =participation tax rate
(1-τ)y
-T(0)
45 °
0
y
pre-tax income y

Optimal nonlinear labor income taxation: How best to solve the trade-off between equity and efficiency

- What is the best way to design **nonlinear** labor income taxes given equity and efficiency concerns?
- Seminal paper: Mirrlees (ReStud 1971)
- Information constraints (no way to make people reveal their skill at no cost)

  $\Rightarrow$ force to move from the 1st best world of the Second Welfare Theorem (where lump sum taxes meet distributional goals given revenue requirement) to the 2nd best world with inefficient (distortionary) taxation

Outline

1. Actual tax systems

2. Optimal nonlinear taxation with intensive margin
   - Labor supply responses to change in the marginal tax rate(s)
   - Mirrlees (ReStud 1971): setting
   - First-best allocations
   - Second-best allocations
   - Derivation of optimal tax rates: 2 skill levels
   - Derivation of optimal tax rates: continuum of skills
   - Derivation of optimal tax rates using a tax perturbation
   - Marginal tax rates for interior skill levels
   - Marginal tax rates at the top
   - Migration effects on marginal tax rates at the top
   - Marginal tax rates at the bottom

Labor supply responses to taxation

Disincentive effects of the tax and benefit system using a labor supply model.

Consider a decrease in T' at income level $\widehat{y}$

For people with earnings $y < \widehat{y}$
$\Rightarrow$ No impact on their labor supply

For people with earnings $\widehat{y}$
$\Rightarrow$ Substitution and income effects

Very often, no income effects are assumed,
e.g. Diamond (AER 1998), Saez (ReStud 2001)

Three effects when modifying a marginal tax rate

Consider a small (uniform) increase of T'
in the neighborhood of $y = \widehat{y} \Rightarrow$

- **Behavioral effect**: Decrease of labor supply hence earnings for taxpayers with earnings $\widehat{y} \Rightarrow$ Decrease in tax revenue
- **Mechanical effect**: Increase of the tax paid by taxpayers with $y > \widehat{y}$ $\Rightarrow$ Increase in tax revenue
- **Welfare cost**: Decrease of the net income of every taxpayer with $y > \widehat{y} \Rightarrow$ Welfare cost whose size depends upon the redistributive preferences of the Government

The optimal T' at any income level balances the gains and costs from changing its level.

Mirrlees' model (1971)

- Agents have identical preferences $u(x, \ell)$, x : consumption, ℓ : labor hours/effort
- $u(.,.)$ twice continuously differentiable, $u_x > 0 > u_\ell$
- Agents differ only by their productivity, $w \sim F(.)$ over a support $\Omega \subset \mathbb{R}^+$

  $F'(w) = f(w) > 0$
- Agents freely choose their labor supply (no fixed costs, no externalities...)
- Perfectly competitive and frictionless labor market

Mirrlees' model (1971)

First-best:
Taxation can be conditioned on (labor income and) skill w

$$T(w)$$

Second-best:
Taxation can be conditioned on labor income only

$$T(y)$$

Geometry of individual preferences

In the (y, x) space:

Utility increases in the North-West direction

Slope of an indifference curve. Define

$$u\left(x, y/w\right) = V\left(x, y; w\right)$$

Hence,

$$\left.\frac{\partial x}{\partial y}\right|_{V=\text{cst}} = -\frac{u_y\left(x, \frac{y}{w}\right)}{w u_x\left(x, \frac{y}{w}\right)} \text{ : slope, i.e. } \text{MRS}_{y,x}$$

$\text{MRS}_{y,x}$ is the marginal cost of labor income in terms of consumption.

Along a given indifference curve, the MRS increases with y
$\Rightarrow$ Indifference curves are convex

Geometry of individual preferences

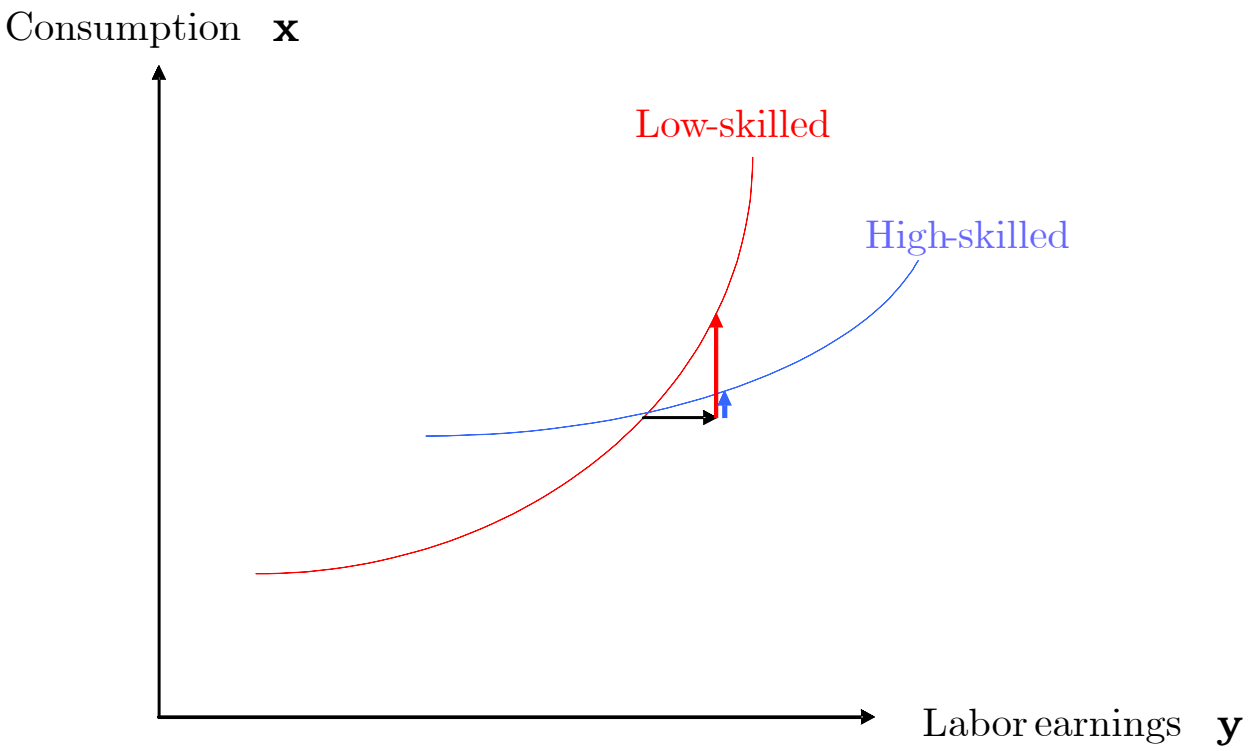
Strict single-crossing condition or Spence-Mirrlees condition
One expects that getting one more euro of labor income requires a smaller additional effort for more skilled workers, hence:

$$\text{For all } x, y, w \quad \frac{\partial MRS(x, y, w)}{\partial w} < 0$$

$w_H > w_L$, take a bundle (x, y, w), the MRS of w_H-agents is lower than the MRS of w_L-agents

$\Leftrightarrow$ The indifference curve of a low-skilled agent is always steeper than the one of a higher skilled agent.

Geometric representation of individuals' preferences



Preferences that (automatically) satisfy Spence-Mirrlees condition

- Additively separable utility:

Take $u(x, y/w)$ additively separable; $u(x, \ell) = v(x) - h(\ell)$, with $v'(x) > 0 \geq 0$, $h'(\ell) > 0$ and $h''(\ell) > 0$. The $\text{MRS}_{y,x}$ becomes

$$MRS_{y,x} = \frac{h'\left(\frac{y}{w}\right)}{wv'(x)}$$

The convexity of $h(.)$ (i.e. $h'(\ell) > 0$ and $h''(\ell) > 0$) ensures that

$$\text{for all } x, y, w \quad \frac{\partial MRS(x, y, w)}{\partial w} = \frac{\frac{-y}{w^2} h''\left(\frac{y}{w}\right) w - h'\left(\frac{y}{w}\right)}{w^2 v'(x)} < 0.$$

- When consumption is a normal good.

The Spence-Mirrlees condition does not seem to be a strong assumption.

Individual's optimization problem

An individual of type w solves:

$$\begin{aligned} & Max_{\ell} \quad u(w \cdot \ell(w) - T(w \cdot \ell(w)), \ell(w)) \\ & \Leftrightarrow Max_{y} \quad u\left(y(w) - T(y(w)), \frac{y(w)}{w}\right) \end{aligned}$$

Let $y(w)$ be the solution to this pgm. If $T(.)$ is differentiable in y, FOC (y):

$$1 - T'(y(w)) = \frac{-1}{w}\frac{u_{\ell}}{u_x}\left(x(w), \frac{y(w)}{w}\right) : \text{MRS}_{y,x} \tag{1}$$

Government's preferences

$$\int_{\Omega} \Phi[u(x(w), \ell(w))] dF(w)$$

with $\Phi'(.) > 0$ and $\Phi''(.) < 0$

Aversion to inequality captured by concavity of Φ

Government's budget constraint

$$\int_{\Omega} \left(w\ell\left(w\right) - x\left(w\right)\right) dF\left(w\right) \geq R$$

or

$$R + \int_{\Omega} x\left(w\right) dF(w) \leq \int_{\Omega} y(w) dF(w)$$

First-best problem

The government chooses bundles $w \mapsto (y(w), x(w))$ to maximize social welfare s.t. budget constraint

$$\underset{\{y(w),x(w)\}_{w\in\Omega}}{Max} \int_{\Omega} \Phi\left[u\left(x(w), \frac{y(w)}{w}\right)\right] dF(w)$$

s.t.

$$\int_{\Omega} \{y(w) - x(w)\} dF(w) \geq R \qquad (\lambda)$$

First-best problem: necessary conditions

FOC:
Denote $\left\{\left(x^1(w), y^1(w)\right)\right\}_{w\in\Omega}$ the solution,

$$\frac{\lambda}{\Phi'(u^1(w))} = u_x\left(x^1(w), \frac{y^1(w)}{w}\right) = -\frac{u'_\ell\left(x^1(w), \frac{y^1(w)}{w}\right)}{w}$$

: First-best optimal allocation

First-best problem

The equilibrium (from the agent's optimization problem) coincides with the first-best allocation **iff the FOC of the agent's maximization problem**

$$1 - T'(y(w)) = \frac{-1}{w}\frac{u_\ell}{u_x}\left(x(w), \frac{y(w)}{w}\right)$$

coincides with the FOC of the government's maximization problem:

$$1 = \frac{-1}{w}\frac{u_\ell}{u_x}\left(x(w), \frac{y(w)}{w}\right)$$

First-best problem

$\Rightarrow$

$$T'(y(w)) = 0$$

The first-best is decentralized by a set of transfers

- that vary with the exogenous w
- that do not vary with the endogenous y.

$\Rightarrow$ **differentiated lump-sum transfers**
(Second theorem of welfare economics)

First-best solution has a straightforward intuition

$\Rightarrow$ The transfer does not depend on earnings
$\Rightarrow$ It does not affect the marginal return of effort
$\Leftrightarrow$ <u>No distortion in the labor supply choice</u>

Second-best problem

If you want to redistribute from rich to poor:
$w \mapsto T\left(y(w)\right)$ has to be increasing

$\Rightarrow$ Marginal tax rates $T'(y(w)) > 0$

$\Rightarrow$ This reduces the marginal return of labor income in terms of consumption

$\Rightarrow$ downward distortion on the labor supply (substitution effect).
$+$ income effects.

Second-best problem

- The government does not observe w but

does observe gross income $y(w) = w\ell(w)$ hence $T(y)$. In other words, taxation can only be conditioned on earnings y and not on skill levels w. $\Rightarrow$ Redistributive taxation is distortionary.

- The w-individual optimization problem becomes:

$$\underset{y}{Max} \quad u\left(y - T(y), \frac{y}{w}\right)$$

Second-best problem

The **revelation principle** (see e.g., Salanié 1997, Chapter 2) implies that

- the planner's problem is to find the best allocation implementable by a direct truthful/revealing mechanism.
- or the gvt cannot do better than by choosing a **direct revealing mechanism**, that is, a pair of functions $(x(w), y(w))$ such that each taxpayer finds it best to announce his own productivity:

$$\forall (w,s) \in \Omega^2 \quad u\left(x(w), \frac{y(w)}{w}\right) \geq u\left(x(s), \frac{y(s)}{w}\right)$$

We have a double infinity of **incentive compatibility constraints**.

Second-best problem

Therefore,

- A feasible allocation must be **compatible with the incentive compatibility constraints**:

$$\forall (w,s) \in \Omega^2 \quad u\left(x(w), \frac{y(w)}{w}\right) \geq u\left(x(s), \frac{y(s)}{w}\right) \tag{2}$$

- Allocations $w \mapsto (x(w), y(w))$ that can be reached by the government have to be consistent with (2) (cf. revelation principle above).

Second-best problem

- Is it **sufficient** to include these IC restrictions **to fully characterize the set of allocations that can be obtained by a government**?

  Yes, this is the **Taxation principle**!
  (Hammond ReStud 1979, Rochet JMathE 1985 and Guesnerie 1995)

Taxation Principle

The **taxation principle** ensures that implementing income taxation $y \mapsto T(y)$ gives the same possibilities as implementing a direct truthful/revealing mechanism $w \mapsto (x(w), y(w))$ s.to the ICC.

Formally:
Let $w \mapsto (x(w), y(w))$ an allocation **that verifies ICC**. **One can "build" an income tax schedule** $y \mapsto T(y)$, such that any agent w chooses the bundle $(x(w), y(w))$ **intended for him/her**.

Taxation Principle

Proof.
Let $w \mapsto (x(w), y(w))$ an allocation that verifies ICC and let $Y = \{y$ such that there exists $w \in \Omega$ for which $y = y(w)\}$.

Construct a tax function $T : y \mapsto T(y)$ such that for all $w \in \Omega$, $x(w) = y(w) - T(y(w))$ and $y(w)$ solves $w-$agent's maximization problem.

- Let's construct this tax function $T(y)$:
  (1) Let $y \in Y$:

(1.a) If there exists a single skill level $w \in \Omega$ for which $y = y(w)$, then one must simply have $T(y(w)) = y(w) - x(w)$.

(1.b) Consider there exist two skill levels $(w, s) \in \Omega^2$ with $w \neq s$ such that $y = y(w) = y(s)$. Then applying ICC (2), one has

$$u\left(x(w), \frac{y(w)}{w}\right) \geq u\left(x(s), \frac{y(s)}{w}\right) \quad \text{then } x(w) \geq x(s).$$

Taxation Principle

Proof. (cont'd)
Symmetrically, inverting the roles of w and s in ICC (2), one has

$$u\left(x(s), \frac{y}{s}\right) \geq u\left(x(w), \frac{y}{s}\right) \quad \text{then } x(s) \geq x(w)$$

$\Rightarrow$ one must have $x(w) = x(s)$ if $y(w) = y(s)$. **We can therefore define** $T\left(y\left(w\right)\right) = y - x\left(w\right)$ **without amibiguity**.

(2) If $y \notin Y$, we define $T(y) = +\infty$.

Taxation Principle

Proof. (cont'd)
Given such a tax function, we have now **to verify that for any skill level** $w \in \Omega$, $y(w)$ **solves** $w-$**agent's maximization program**.

- If we choose an earnings level $y \notin Y$, this implies $T(y) = +\infty \Rightarrow$ This choice of y is suboptimal.
- If now we choose an earning level $y \in Y$ with $y \neq y(w)$, this is equivalent to choosing a skill level s such that $y = y(s)$. However, from ICC (2) and $x(t) = y(t) - T(y(t))$ for $t = w, s$ we have:

$$u\left(y(w) - T(y(w)), \frac{y(w)}{w}\right) \geq u\left(y - T(y), \frac{y}{w}\right) = u\left(x(s), \frac{y(s)}{w}\right)$$

(i.e. the mimicker, agent s, ends up with a lower utility level than agent-w who takes the bundle intended for her), which ends the proof that $y(w)$ is a solution to the individual maximization program.

Remark: The restriction of the function $y \mapsto y - T(y)$ over Y has to be **increasing**.

Proof.

Let s and w two skill levels such that $y(w) > y(s)$. Assume $x(s) \geq x(w)$ then

$$u\left(x(s), \frac{y(s)}{w}\right) \geq u\left(x(w), \frac{y(s)}{w}\right) \geq u\left(x(w), \frac{y(w)}{w}\right)$$

which violates ICC (2). Therefore, $y(s) < y(w) \Rightarrow x(s) < x(w)$.

Remarks:

- $y \mapsto y - T(y)$ (over Y) increasing implies that $T'(y)$ **has to be lower than one** for any chosen y.
- $T'(y) = 1$ only pointwise.
- If $T'(y) \geq 1$ for some y then y will be not chosen by any worker, whatever her skill level.
- $T'(y) < 1$ has nothing to do with optimized tax schedules. It is a property of any incentive compatible allocation!

Second-best government's problem

$$Max_{\{y(w),x(w)\}} \quad \int_{\Omega} \Phi\left[U\left(x(w), \frac{y(w)}{w}\right)\right] dF(w)$$

$$\int_{\Omega} \{y(w) - x(w)\} \, dF(w) \geq R$$

subject to:

$$\forall (w,s) \quad u\left(x(w), \frac{y(w)}{w}\right) \geq u\left(x(s), \frac{y(s)}{w}\right)$$

Solution to this program: optimal second-best allocation $\{x(w), y(w)\}$.
Then optimum tax schedule: $T(y(w)) = y(w) - x(w)$.

Optimal income taxation with two levels of productivity

Stiglitz (NBER or JPubE, 1982)

Two skill levels $0 < w_L < w_H$ (i.e. $F(w)$ reduced to two mass points)

π_L: the mass of workers of skill L

π_H: the mass of workers of skill H

$\pi_L + \pi_H = 1$ (and $\pi_L > 0$ and $\pi_H < 1$)

Government's constrained optimization problem

$$\underset{y_L, x_L, y_H, x_H}{Max}$$

$$\pi_L \Phi\left[u\left(x_L, \frac{y_L}{w_L}\right)\right] + \pi_H \Phi\left[u\left(x_H, \frac{y_H}{w_H}\right)\right]$$

$$u\left(x_H, \frac{y_H}{w_H}\right) \geq u\left(x_L, \frac{y_L}{w_H}\right) \qquad \text{(mult. } \lambda_H\text{)} \qquad (\text{IC}_H)$$

$$u\left(x_L, \frac{y_L}{w_L}\right) \geq u\left(x_H, \frac{y_H}{w_L}\right) \qquad \text{(mult. } \lambda_L\text{)} \qquad (\text{IC}_L)$$

$$\pi_L\left(y_L - x_L\right) + \pi_H(y_H - x_H) \geq R \qquad \text{(mult. } \gamma\text{)}$$

Necessary conditions

$$\pi_H \left\{ \Phi_{u_H} u_x \left(x_H, \frac{y_H}{w_H} \right) - \gamma \right\} + \lambda_H u_x \left(x_H, \frac{y_H}{w_H} \right) - \lambda_L u_x \left(x_H, \frac{y_H}{w_L} \right) = 0$$

$$\pi_H \left\{ \Phi_{u_H} \frac{u_\ell \left(x_H, \frac{y_H}{w_H} \right)}{w_H} + \gamma \right\} + \lambda_H \frac{u_\ell \left(x_H, \frac{y_H}{w_H} \right)}{w_H} - \lambda_L \frac{u_\ell \left(x_H, \frac{y_H}{w_L} \right)}{w_L} = 0$$

$$\pi_L \left\{ \Phi_{u_L} u_x \left(x_L, \frac{y_L}{w_L} \right) - \gamma \right\} - \lambda_H u_x \left(x_L, \frac{y_L}{w_H} \right) + \lambda_L u_x \left(x_L, \frac{y_L}{w_L} \right) = 0$$

$$\pi_H \left\{ \Phi_{u_L} \frac{u_\ell \left(x_L, \frac{y_L}{w_L} \right)}{w_L} + \gamma \right\} - \lambda_H \frac{u_\ell \left(x_L, \frac{y_L}{w_H} \right)}{w_H} + \lambda_L \frac{u_\ell \left(x_L, \frac{y_L}{w_L} \right)}{w_L} = 0$$

Solving of the second-best

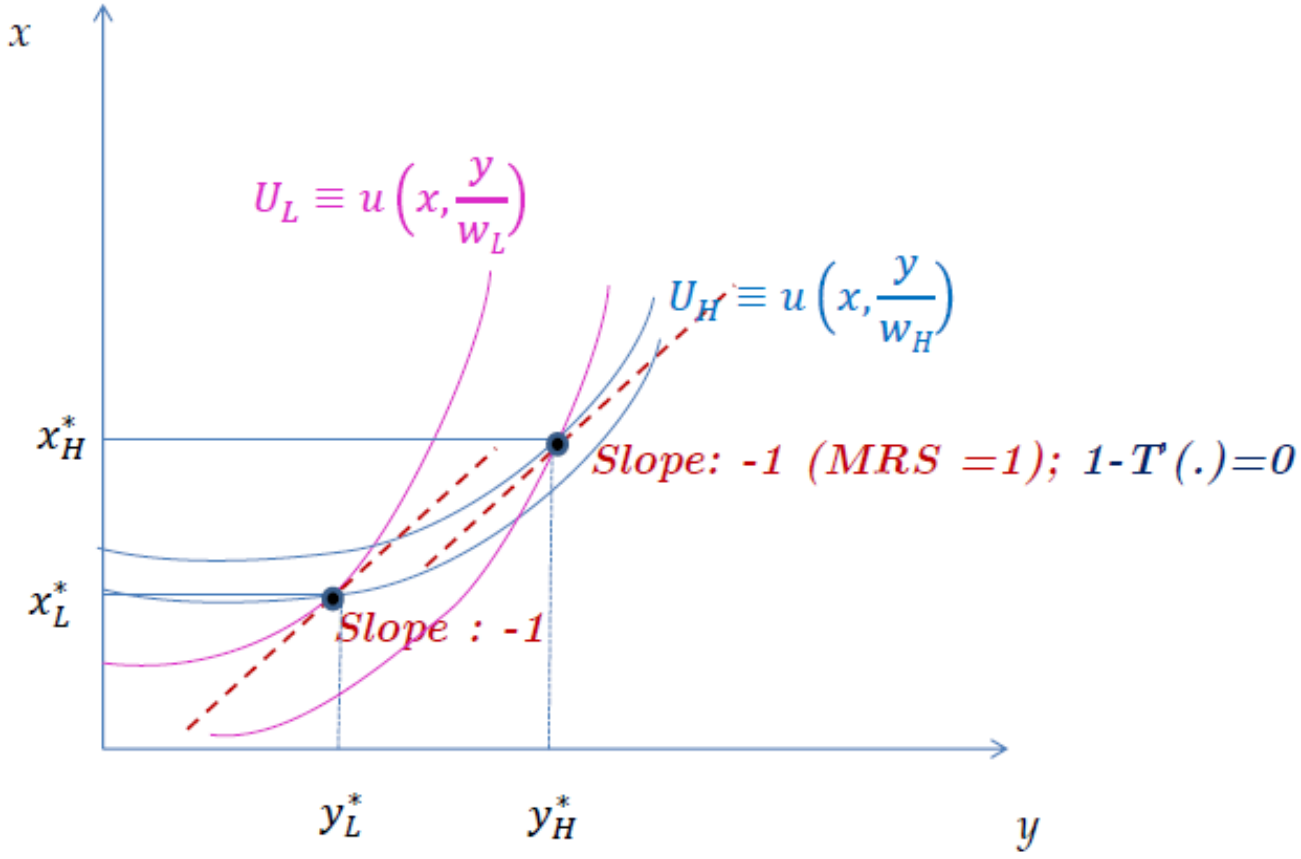
In principle three cases have to be considered:

- $\lambda_L = \lambda_H = 0$: no constraint
- $\lambda_H > \lambda_L = 0$: the normal case
- $\lambda_L > \lambda_H = 0$: the non-normal (anti-redistributive) case

These cases depend on the labor income level where the two indifference curves intersect compared to the optimal labor income levels without ICC (i.e. first-best).

The "no constraint" case

The indifference curves intersect at a labor income levels between the first-best optimal income levels: fully revealing taxation



The "no constraint case"

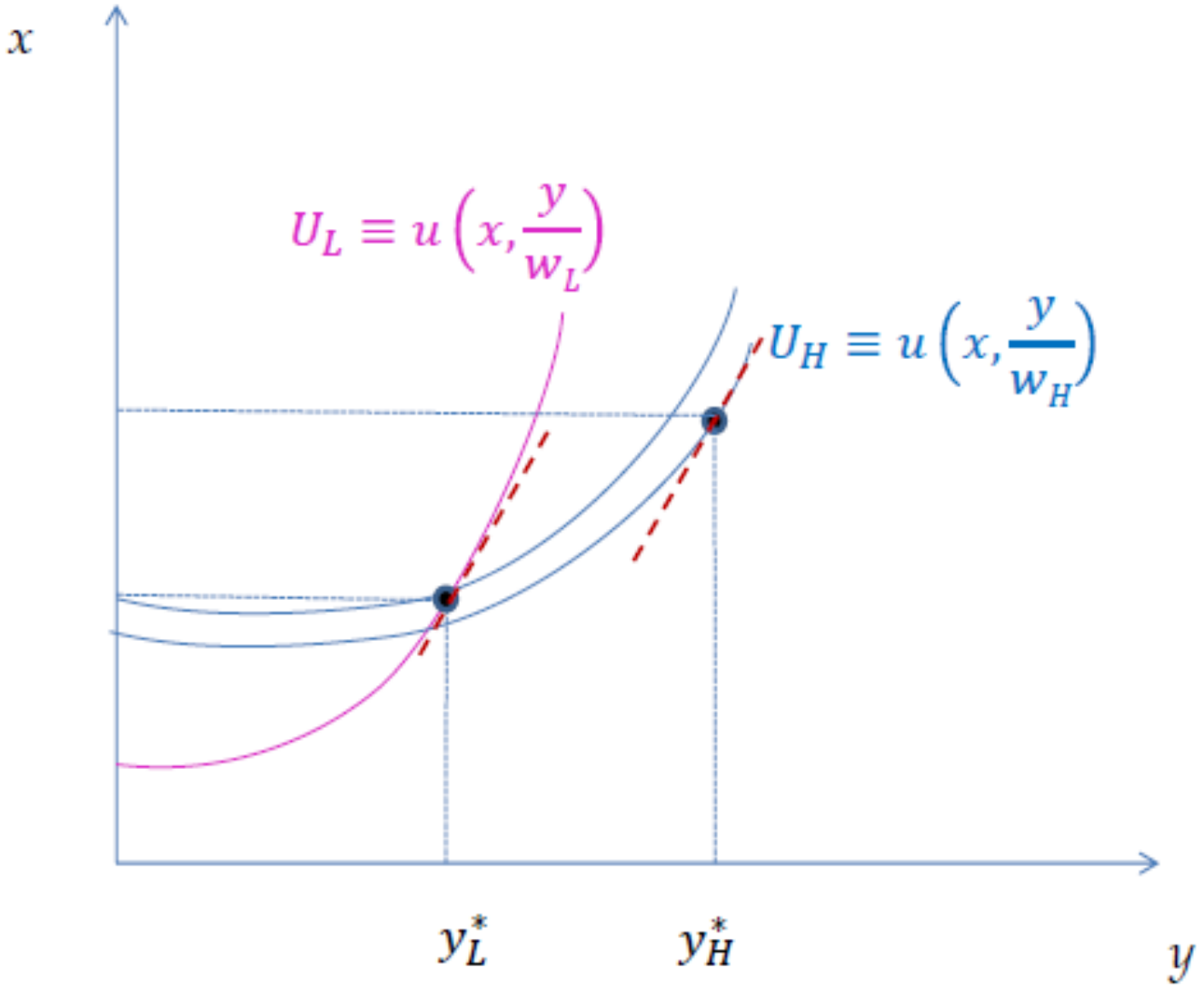
Absence of IC constraints $\lambda_H = \lambda_L = 0$

$$\frac{1}{w_H}\frac{u_\ell\left(x_H, \frac{y_H}{w_H}\right)}{u_x\left(x_H, \frac{y_H}{w_H}\right)} = \frac{1}{w_L}\frac{u_\ell\left(x_L, \frac{y_L}{w_L}\right)}{u_x\left(x_L, \frac{y_L}{w_L}\right)} = -1$$

- Hence: $T'(y_H) = T'(y_L) = 0$
- Redistribution happens through lump-sum transfers
- Fully revealing taxation

Towards the "normal" case

The indifference curves intersect at a labor income level lower than y_L^*



$\Rightarrow$ High-skilled people mimic low-skilled people.

Towards the "normal" case

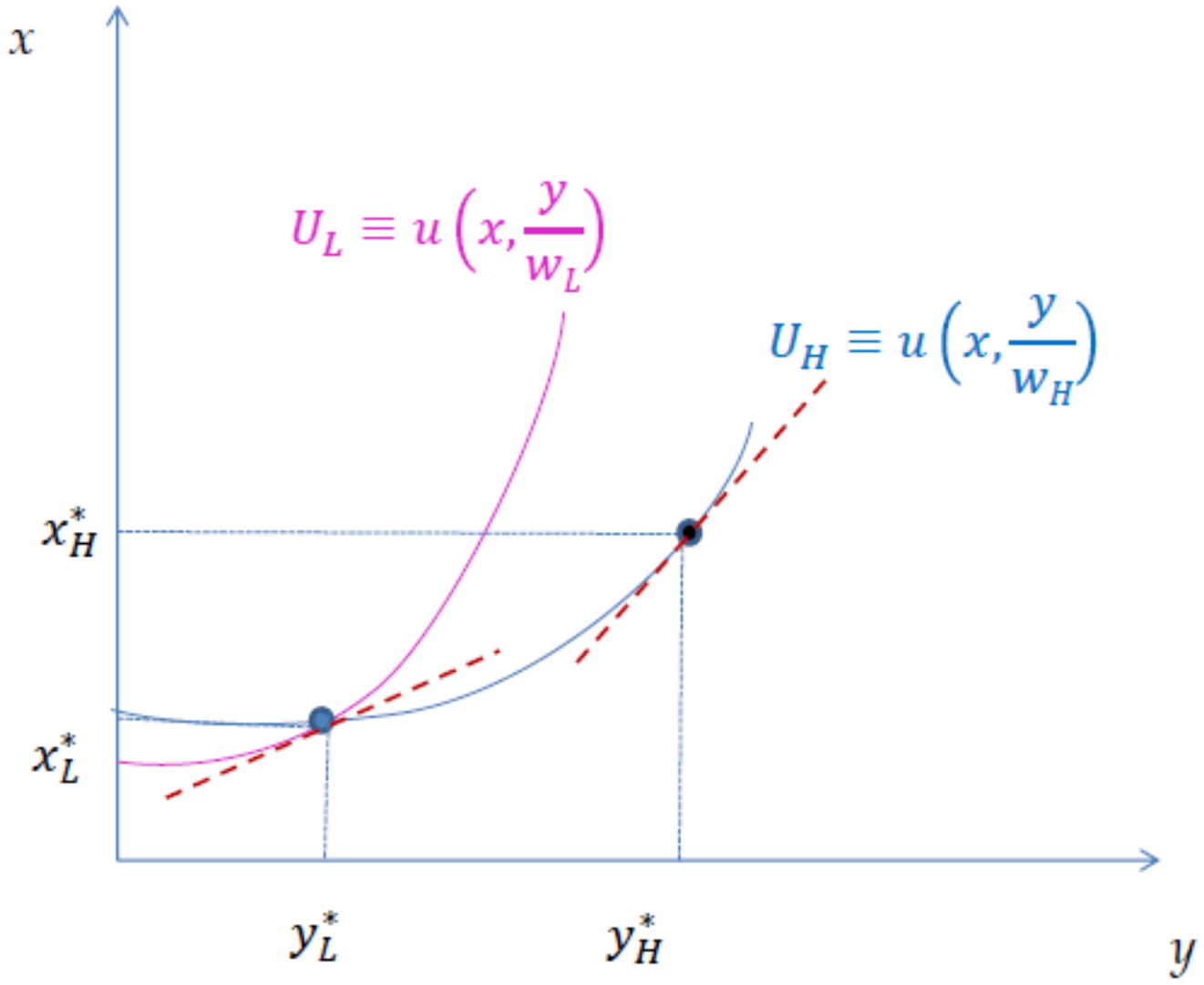
To prevent mimicking of low-skilled workers by high-skilled workers, the government reduces both earnings (hence labor supply) and consumption designed for low-skilled workers along the low-skilled workers' indifference curve

$\Rightarrow$ High-skilled workers have less and less incentives to mimic low-skilled workers

$\Rightarrow$ The tax $y_L - x_L$ paid by low-skilled workers decreases

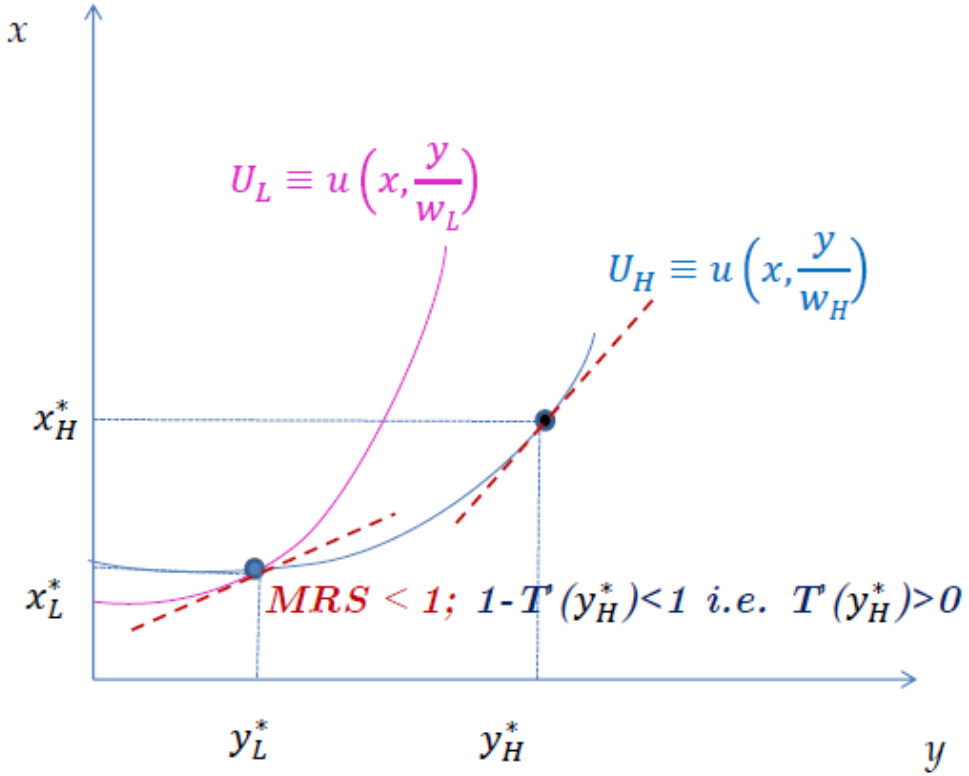
Towards the "normal" case

Mimicking stops as soon as the optimum corresponds to the intersection of the 2 indifference curves



The "normal" case

The new tax schedule satisfies IC_H



The ICC is binding on the w_H-types. The ICC is slack on the wL-types, i.e. $\lambda_H > \lambda_L = 0$. This is the normal case.

The ICC is binding on the w_H-types. The ICC is slack on the w_L-types. i.e. $\lambda_H > \lambda_L = 0$. This is the normal case

The "normal" case

$\lambda_H > \lambda_L = 0$: the normal case

$$\gamma \pi_H = (\pi_H \Phi_{u_H} + \lambda_H)\, u_x \left(x_H, \frac{y_H}{w_H} \right)$$

$$\gamma \pi_H w_H = (\pi_H \Phi_{u_H} + \lambda_H)\, u_\ell \left(x_H, \frac{y_H}{w_H} \right)$$

$$\gamma \pi_L = \pi_\ell \Phi_\ell u_x \left(x_L, \frac{y_L}{w_L} \right) - \lambda_H u_x \left(x_L, \frac{y_L}{w_H} \right)$$

$$\gamma \pi_L w_L = \pi_\ell \Phi_\ell u_\ell \left(x_L, \frac{y_L}{w_L} \right) - \lambda_H \frac{y_L}{w_H} u_x \left(x_L, \frac{y_L}{w_H} \right)$$

The "normal" case

Hence,

$$1 - T'(y_H) = \frac{1}{w_H} \frac{u_\ell\left(x_H, \frac{y_H}{w_H}\right)}{u_x\left(x_H, \frac{y_H}{w_H}\right)} = 1 \Rightarrow T'(y_H) = 0$$

$$1 - T'(y_L) = \frac{1}{w_L} \frac{u_\ell\left(x_L, \frac{y_L}{w_L}\right)}{u_x\left(x_H, \frac{y_L}{w_L}\right)} < 1 \Rightarrow T'(y_L) > 0$$

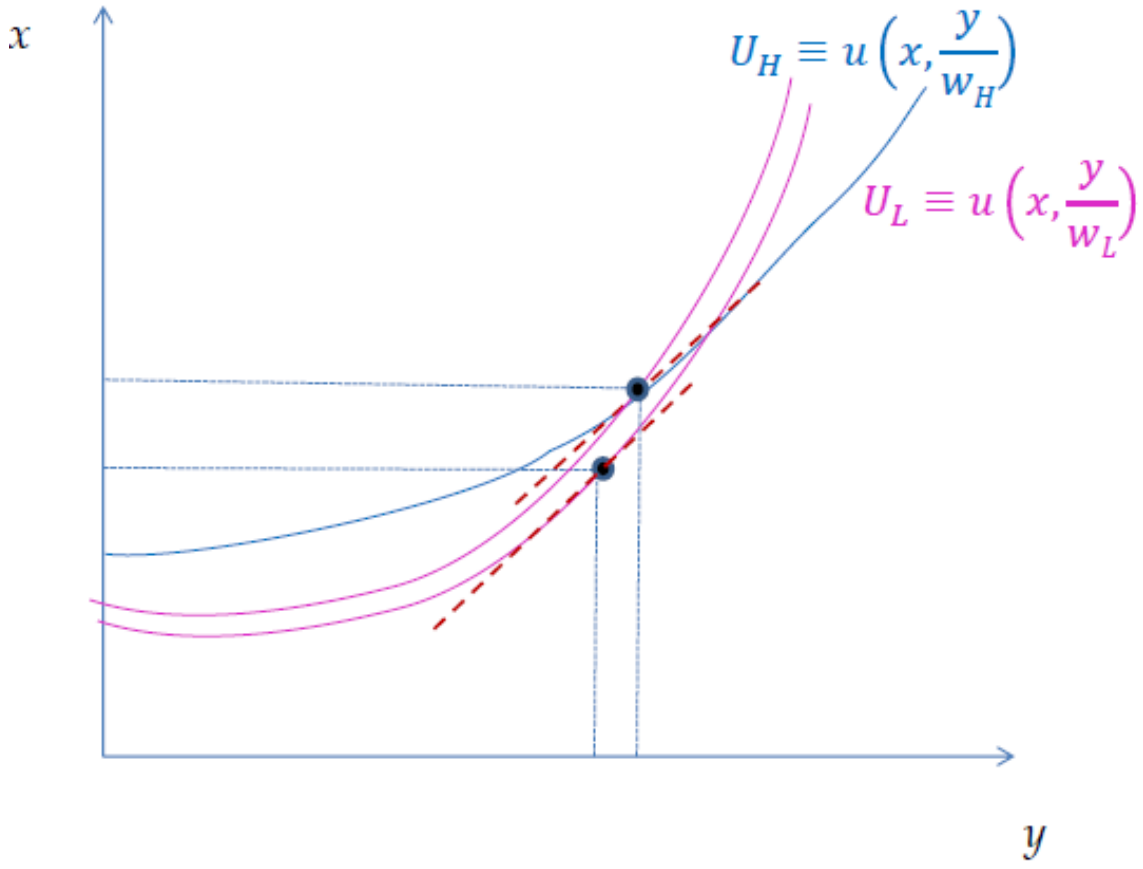
Interpretations in the "normal" case

High-skilled workers' labor supply is undistorted

Low-skilled workers' labor supply is distorted downwards

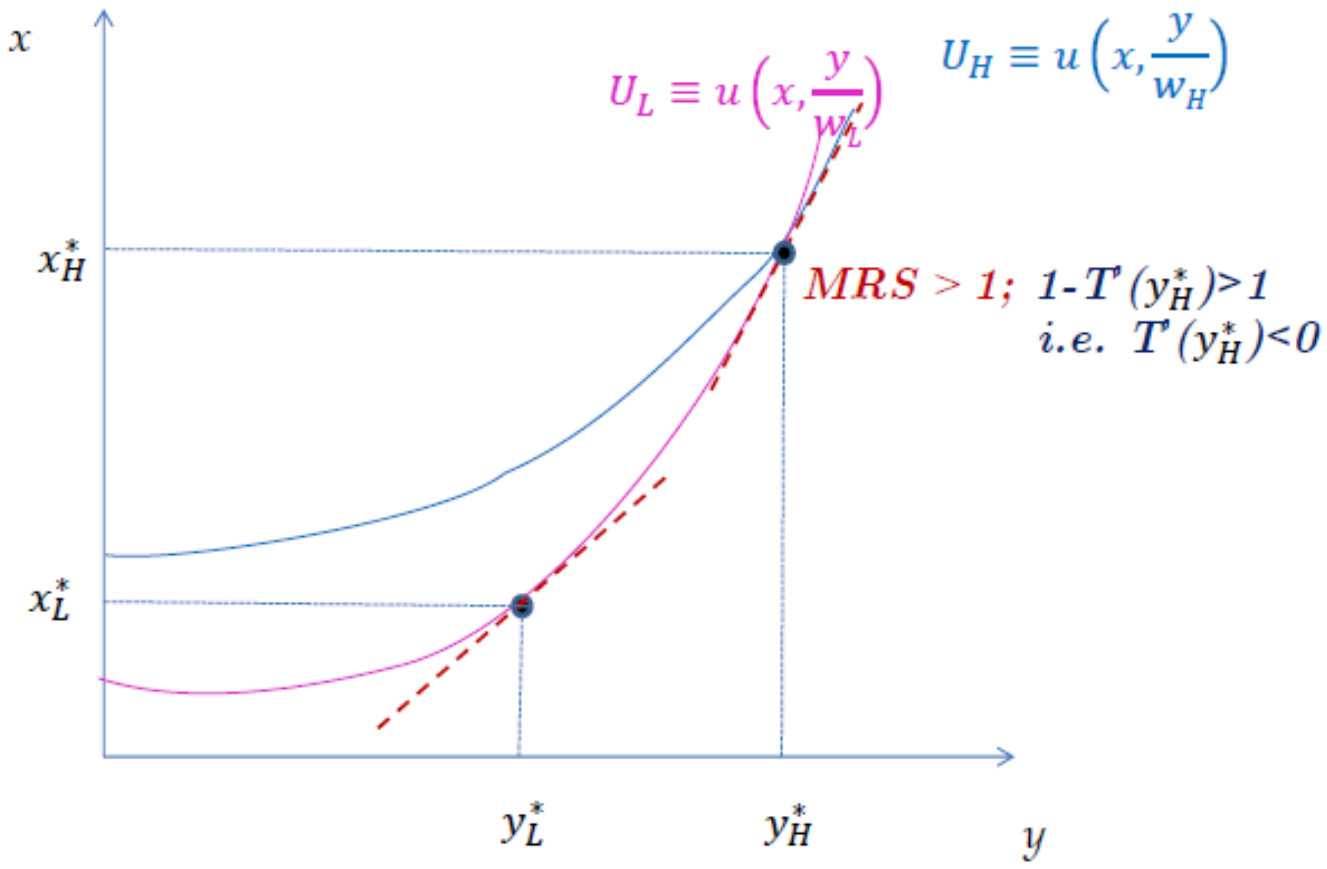
Towards the "anti-redistributive" case

i.e. when the two indifference curves intersect at a labor income larger than y_H^*



$\Rightarrow$ Low-skilled people mimic high-skilled people.

The "anti-redistributive" case



The ICC is binding on the w_L-types. The ICC is slack on the w_H-types. i.e. $\lambda_L > \lambda_H = 0$. This is the anti-redistributive case (or.non-normal case).

The "anti-redistributive" case

Hence,

$$T'(y_H) < 0$$
$$T'(y_L) = 0$$

Optimal labor income tax, continuum of skills

Important Assumptions:

- Continuum of skills: $\Omega \equiv [w_0, w_1]$
- No mass point and continuous density $f(w)$

Incentive compatibility constraints (ICC) with a continuum of skills

$$\forall (w, s) \quad u\left(x(w), \frac{y(w)}{w}\right) \geq u\left(x(s), \frac{y(s)}{w}\right) \tag{3}$$

= a double continuum of inequalities

How to deal with constraints (3), that is with a double continuum of inequalities? Mirrlees (1971) has shown that under the Spence-Mirrlees condition, these constraints are equivalent to a differential equation and a monotonicity constraints. To ease the presentation, I restrict the utility function to be **additively separable**:

$$u(x,\ell) \equiv v(x) - h(\ell) \text{ where } v'(x) > 0 \geq v''(x),\ h'(\ell) > 0 \text{ and } h''(\ell) > 0$$

For general preferences, see e.g. Jacquet, Lehmann and Van der Linden (JET 2013)

Let $u(w)$ be the value function associated to the optimization program of workers of skill w, that is

$$\begin{aligned} u(w) &= \underset{y}{Max}\ v(y - T(y)) - h\left(\frac{y}{w}\right) \\ &= v(x(w)) - h\left(\frac{y(w)}{w}\right) \end{aligned}$$

These ICC are equivalent to a differential equation and a monotonicity constraint

Mirrlees (1971): Under the Spence-Mirrlees condition, these ICC are equivalent to the requirements that:

- $w \longmapsto u(w)$ is continuous and

$$u'(w) = \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) = \frac{\ell(w)h'(\ell(w))}{w} \text{ almost everywhere} \tag{IC$_1$}$$

- $w \longmapsto y(w)$ is non decreasing (Monotonicity of $y(w)$), i.e.

$$y'(w) \geq 0 \quad \Leftrightarrow \ell'(w) + w\ell'(w) \geq 0 \tag{IC$_2$}$$

More formally:

Proposition

Under Spence-Mirrlees condition, the set of allocations $w \longmapsto (x(w), y(w), u(w))$ *that verify* $u(w) = u\left(x(w), \frac{y(w)}{w}\right)$ *and ICC (2) is the same as the set of allocations* $w \longmapsto (x(w), y(w), u(w))$ *that verify* $u(w) = u\left(x(w), \frac{y(w)}{w}\right)$ *and* $w \longmapsto u(w)$ *is continuous,* $u'(w) = \frac{y(w)}{w^2} v'\left(\frac{y(w)}{w}\right)$ *almost everywhere (a.o.) and* $w \longmapsto y(w)$ *is non decreasing (monotonicity).*

Proof in 2 steps [Skip]

1. Under Spence-Mirrlees condition, ICC (2) $\Rightarrow w \longmapsto u(w)$ is continuous, $u'(w) = \frac{y(w)}{w^2} v'\left(\frac{y(w)}{w}\right)$ almost everywhere (a.o.) and $w \longmapsto y(w)$ is non decreasing.
2. Under Spence-Mirrlees condition, $w \longmapsto u(w)$ is continuous, $u'(w) = \frac{y(w)}{w^2} v'\left(\frac{y(w)}{w}\right)$ almost everywhere (a.o.) and $w \longmapsto y(w)$ is non decreasing $\Rightarrow$ ICC (2) satisfied

Proof: Step 1.

Proof that ICC implies monotonicity

- On Figure 1, indifference curves of w_L-agents and w_H-agents ($w_L < w_H$).
- These indifference curves intersect at bundle designed for w_L-individuals.
- Single-crossing $\Rightarrow$ indifference curve of the w_L-workers steeper than the one of the w_H-workers.

Proof that ICC implies monotonicity (cont'd)

- To respect ICC, the government needs to assign a bundle $(x(w_H), y(w_H))$ to the w_H-workers that is
  - above the indifference curve of the high-skilled workers (otherwise, w_H-individuals would prefer$(x(w_L), y(w_L))$ to $(x(w_H), y(w_H))$ designed for them)
  - and below the indifference curve of the w_L-workers (otherwise, w_L-individuals would prefer $(x(w_H), y(w_L))$ to $(x(w_L), y(w_L))$ designed for them).

$\Rightarrow$ The bundle $(x(w_H), y(w_H))$ designed for the w_H-workers must be in the non-shaded area in Figure 1. This bundle implies that

$$y(w_L) \leq y(w_H) \text{ and } x(w_L) \leq x(w_H)$$

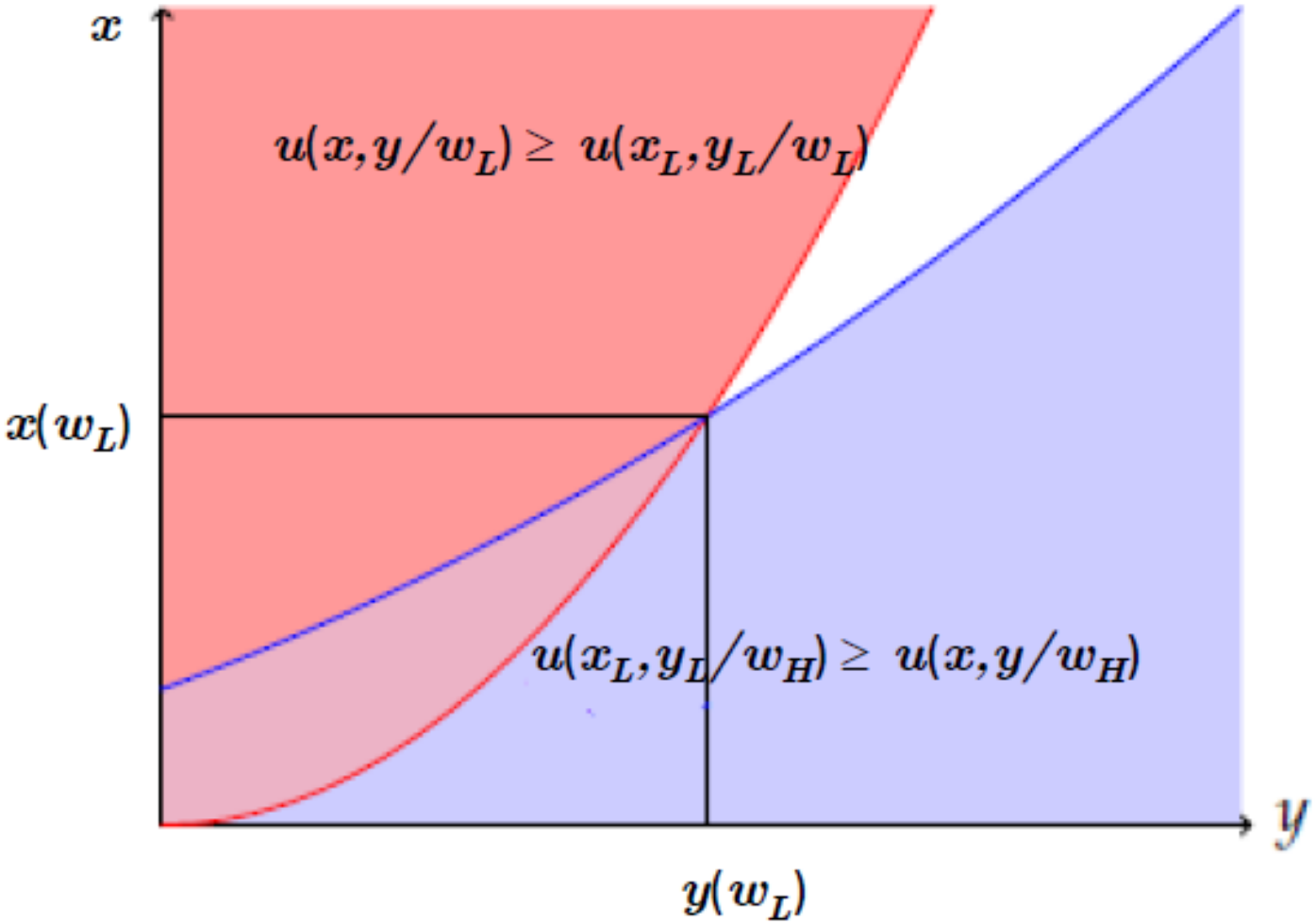


Figure: Fig. 1: Under Spence-Mirrlees condition, $y(w)$ and $x(w)$ non-decreasing in w

Proof that ICC implies continuity of u(w)

Since for any s, one has $u(s) = v(x(s)) - h(y(s)/s)$, the ICC incentive (2) implies

$$u(w) \geq v(x(s)) - h\left(\frac{y(s)}{w}\right) = u(s) + h\left(\frac{y(s)}{s}\right) - h\left(\frac{y(s)}{w}\right).$$

Therefore,

$$h\left(\frac{y(s)}{s}\right) - h\left(\frac{y(s)}{w}\right) \leq u(w) - u(s)$$

Proof that ICC implies continuity of u(w) (cont'd)

Inverting the role of s and w, we also have

$$h\left(\frac{y(w)}{w}\right) - h\left(\frac{y(w)}{s}\right) \leq u(s) - u(w)$$

$$\Leftrightarrow u(w) - u(s) \leq h\left(\frac{y(w)}{s}\right) - h\left(\frac{y(w)}{w}\right)$$

$$\Rightarrow \; h\left(\frac{y(s)}{s}\right) - h\left(\frac{y(s)}{w}\right) \leq u(w) - u(s) \leq h\left(\frac{y(w)}{s}\right) - h\left(\frac{y(w)}{w}\right) \tag{4}$$

Upper and lower bounds of (4)$\longrightarrow 0$ as $s \longrightarrow w$.
Therefore $u(w) - u(s) \longrightarrow 0$ as $s \longrightarrow w$.
This ensures the continuity of $u(w)$ at w.

Proof: u'(w) a.o.

From

$$u(w) = \underset{y}{Max} v(y - T(y)) - h\left(\frac{y}{w}\right) = v(x(w)) - h\left(\frac{y(w)}{w}\right),$$

using the Envelope theorem, we have:

$$u'(w) = \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) > 0. \tag{5}$$

Proof: u'(w) a.o. (cont'd)

We can use the mathematical results that any nondecreasing function is continuous everywhere except on a set that is at worst countable. Therefore, Accordingly, $w \longmapsto y(w)$ is **a.o. continuous**.

We now show that (5) holds a.o.:

Let w be a skill level at which $w \longmapsto y(w)$ is continuous.
For $s < w$, from (4), we have

$$\frac{h\left(\frac{y(s)}{s}\right) - h\left(\frac{y(s)}{w}\right)}{w - s} \leq \frac{u(w) - u(s)}{w - s} \leq \frac{h\left(\frac{y(w)}{s}\right) - h\left(\frac{y(w)}{w}\right)}{w - s}.$$

Proof: u'(w) a.o. (cont'd)

By continuity at w of $w \longmapsto y(w)$, from (5), the lower and upper bounds of the latter equation $\longrightarrow \frac{y(w)h'\left(\frac{y(w)}{w}\right)}{w^2}$ as $s \longrightarrow w$.
Then, $w \longmapsto u(w)$ admits a **left**-derivative at w that is equal to $\frac{y(w)h'\left(\frac{y(w)}{w}\right)}{w^2}$.

Proof: u'(w) a.o. (cont'd)

For $s > w$, a symmetric reasoning applies to show that $w \longmapsto u(w)$ admits a **right**-derivative at w that is aslo equal to $\frac{y(w)h'\left(\frac{y(w)}{w}\right)}{w^2}$

We can conclude that

$$u'(w) = \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) > 0$$

holds a.o.

Proof: Step 2.

We now show the reciprocal, i.e. Under Spence-Mirrlees condition, $w \longmapsto u(w)$ is continuous, $u'(w) = \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right)$ almost everywhere (a.o.) and $w \longmapsto y(w)$ is non decreasing $\Rightarrow$ ICC (2) satisfied

Proof: sufficiency

Let two skill levels w and s. By the continuity of $w \longmapsto u(w)$ and the fact that $u'(w) = \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right)$ holds a.o., we have

$$u(w) - u(s) = \int_s^w \left(\frac{y(t)}{t^2}\right) h'\left(\frac{y(t)}{t}\right) dt \tag{6}$$

Since $h(.)$ is increasing and convex, for any t one has $y \longmapsto \left(\frac{y}{t^2}\right) h'\left(\frac{y}{t}\right)$ is increasing.
If $s < w$, (resp. $w < s$), then for all $t \in (s, w)$ (resp. $t \in (w, s)$) one has $\left(\frac{y(t)}{t^2}\right) h'\left(\frac{y(t)}{t}\right) \geq \left(\frac{y(s)}{t^2}\right) h'\left(\frac{y(s)}{t}\right)$ (resp. $\leq$) since $w \longmapsto y(w)$ is nondecreasing.

Proof: sufficiency (cont'd)

Hence we have

$$\int_s^w \left(\frac{y(t)}{t^2}\right) h'\left(\frac{y(t)}{t}\right) dt \geq \int_s^w \left(\frac{y(s)}{t^2}\right) h'\left(\frac{y(s)}{t}\right) dt$$

since $w > s$ (resp. $s > w$), we obtain, using (6):

$$u(w) - u(s) \geq \int_s^w \left(\frac{y(s)}{t^2}\right) h'\left(\frac{y(s)}{t}\right) dt$$

Using $u(s) = v(x(s)) - h\left(\frac{y(s)}{s}\right)$ and integrating we obtain the ICC.

Using $x(w) \equiv \chi\left(u(w), \frac{y(w)}{w}\right)$, the **government's optimization problem** can then be rewritten in terms of $y(w)$ and $u(w)$ only:

$$\underset{\{y(w),u(w)\}\in\Omega}{Max} \quad \int_{\Omega} \Phi[u(w)]f(w)dw$$

s.to:

$$u'(w) = \frac{y(w)}{w^2} h'\left(\frac{y(w)}{w}\right) \qquad (\mathrm{IC}_1 \quad (q(w)))$$

$$w \longmapsto y(w) \text{ nondecreasing} \qquad (\mathrm{IC}_2)$$

$$\int_{\Omega} \left[y(w) - \chi\left(u(w), \frac{y(w)}{w}\right)\right] f(w)dw \geq R$$

Shadow price of the public funds: λ

First-order approach: we solve this problem without IC_2 with optimal control theory and in a second step, we check that $y(w)$ is well increasing with w, i.e. we check if IC_2 is satisfied.

Solving:
Write down the Hamiltonian

Take:
$y(w)$ as the control variable
$u(w)$ as the state variable
and
Co-state or adjoint variable: $q(w)$

Hamiltonian

$$H\left(y,u,w,q,\lambda\right)\equiv\left\{\Phi\left(u\right)+\lambda\left[y-\chi\left(u,\frac{y}{w}\right)f\right]f\left(w\right)+q\frac{y}{w^{2}}h'\left(\frac{y}{w}\right)\right\}$$

where we have assumed that the distribution of skill has **no mass points** and admits a **continuous density f (w)**.

- Apply the optimal control tools to get necessary conditions that the second-best optimal allocation has to verify.
- Several difficulties may however appear: (1) due to monotonicity constraint, (2) discontinuity of $y(w)$.

First difficulty: binding monotonicity constraint; bunching

Monotonicity constraint on $w \longmapsto y(w)$ makes difficult $y(w)$ as "true" control variable.

Monotonicity constraint can be binding in an interval $[w_1, w_2] \Rightarrow y(w)$ and $x(w)$ constant in $[w_1, w_2]$.
$\Rightarrow$ bunching (due to violation of monotonicity); individuals with $w \in [w_1, w_2]$ choose the same bundle $(x(w), y(w))$.

Figure 2: Tax function $y \longmapsto T(y)$ has a **kink** at earning level y with a sudden increase in $T'(.)$.
$\Rightarrow$ Function $y \longmapsto y - T(y)$ has a **"downward" kink** at y.

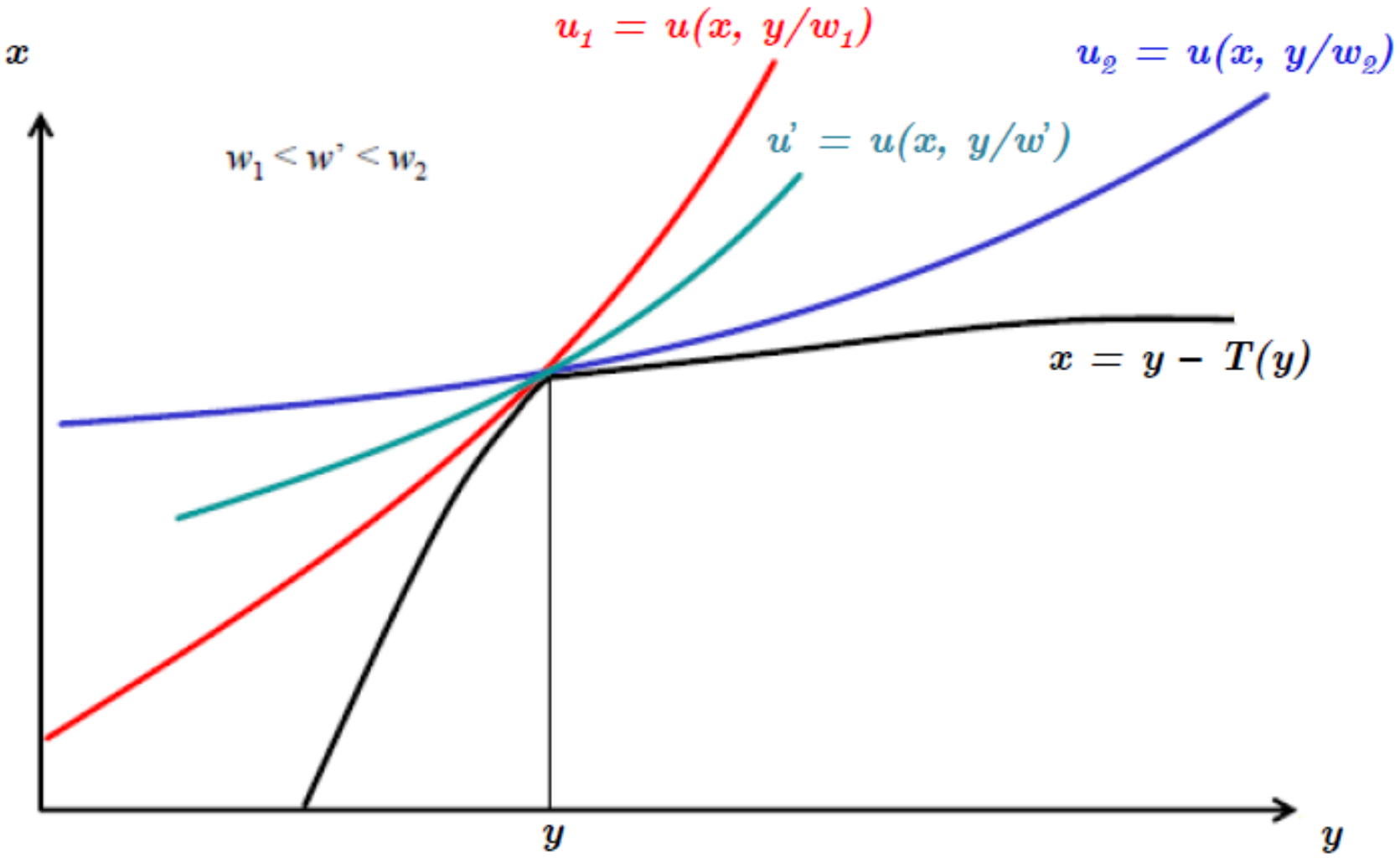


Figure: Fig.2: Bunching due to violation of monotonicity

First difficulty: binding monotonicity constraint; bunching (cont'd)

w_1-workers: indifference curve tangent to $y \longmapsto y - T(y)$ just before the kink.

w_2-workers: indifference curve tangent to $y \longmapsto y - T(y)$ just after the kink.

$w'-$workers, with $w' \in (w_1, w_2)$:

face a too low $T'(.)$ just before $y \Rightarrow$ incentives to work more than y,

face a too high $T'(.)$ just after $y \Rightarrow$ incentives to work less than y,

$\Rightarrow$ they earn exactly y

$\Rightarrow$ **bunching** for individuals with $w \in [w_1, w_2]$.

First difficulty: binding monotonicity constraint; bunching (cont'd)

To deal with bunching:

1. Solve problem without monotonicity constraint
2. Check (analytically or numerically) that the obtained solution satisfies monotonicity constraint

= **First-order approach**

Other approach: e.g. Guesnerie and Laffont (JPubE 1984) (integrate one of the below FOC over the bunching interval)

Second difficulty: discontinuity of y(w)

$w \longmapsto y(w)$ discontinuous at earning level y.

Figure 3: $y \longmapsto y - T(y)$ becomes locally more convex than the indifference curve of workers of skill w_2.
$\Rightarrow$ 2 tangency points at y_2^L and y_2^H, with $y_2^L < y_2^H$.
$\Rightarrow$ Function $w \longmapsto y(w)$ **jumps** from y_2^L to y_2^H at skill level w_2.

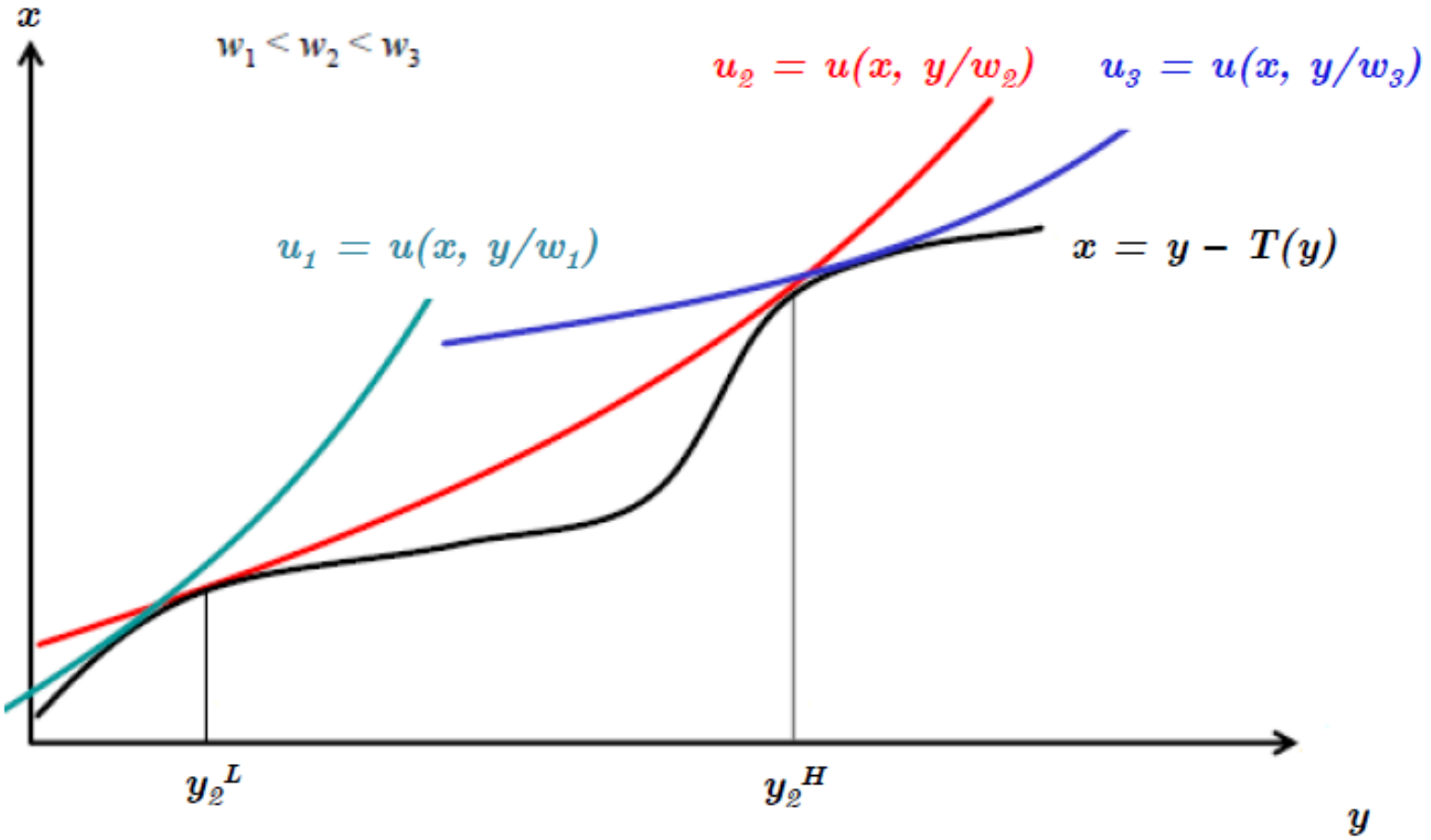


Figure: Fig.3: A discontinuous allocation

Second difficulty: discontinuity of y(w)

To deal with discontinuity of $y(w)$:

- Literature typically ignores this.
- From a technical viewpoint, the **necessary conditions derived from the control technics** are only available if there is a **finite number of discontinuity points**. However, since $w \longmapsto y(w)$ is nondecreasing, the **set of points of discontinuity** is at worst **countable**, thereby of zero measure (since we have assume no mass points in the skill distribution). We therefore **assume** that this set is **finite**.
  - Note: within the set of real numbers, the subset of rational number is countable but dense (i.e. no finite)

Solving

$$H\left(y,u,w,q,\lambda\right)\equiv\left\{\Phi\left(u\right)+\lambda\left[y-\chi\left(u,\frac{y}{w}\right)\right]f\left(w\right)+q\frac{y}{w^{2}}h'\left(\frac{y}{w}\right)\right\}$$

$$\frac{\partial H}{\partial y}=0\Leftrightarrow\lambda\left[1-\frac{h'\left(\frac{y(w)}{w}\right)}{wv'\left(x\left(w\right)\right)}\right]f\left(w\right) \tag{7}$$

$$+\frac{q\left(w\right)}{w^{2}}\left[h'\left(\frac{y}{w}\right)+\frac{y}{w}h''\left(\frac{y}{w}\right)\right]\overset{a.e.}{=}0 \tag{8}$$

$$\frac{\partial H}{\partial u}=-q'\left(w\right)\Leftrightarrow-q'(w)\overset{a.e.}{=}\left\{\Phi'\left(u(w)\right)-\frac{\lambda}{v'\left(x\left(w\right)\right)}\right\}f(w)$$

$$q(\underline{w})=q(\overline{w})=0$$

where we have used:

$$u = u\left(x, y/w\right) \Leftrightarrow x = \chi\left(u, y/w\right)$$

$\Rightarrow$

$$\chi_u\left(u, y/w\right) = \frac{1}{u_x\left(\chi\left(u, y/w\right), y/w\right)} = \frac{1}{v'}$$

and

$$\chi_y\left(u, y/w\right) = \frac{-u_y}{wu_x}\left(\chi\left(u, y/w\right), y/w\right) = \frac{-h'}{wv'}$$

Integrate the derivative of the costate variable to get

$$q(w) = q(\overline{w}) - \int_{w}^{\overline{w}} q'(n)dn = \int_{w}^{\overline{w}} \left\{\Phi'(u(n)) - \lambda\right\} f(n)dn.$$

Then the transversality condition yields

$$q(\underline{w}) = 0 \Leftrightarrow \lambda \int_{\underline{w}}^{\overline{w}} \frac{1}{v'(x(n))} f(n)\, dn = \int_{\underline{w}}^{\overline{w}} \Phi'(u(n))\, f(n)dn$$

Hence, the **optimum** is given by

$$1 - \frac{h'\left(\frac{y(w)}{w}\right)}{wv'(x(w))} \overset{a.e.}{=} \frac{h'\left(\frac{y}{w}\right) + \frac{y}{w}h''\left(\frac{y}{w}\right)}{w^2 f(w)} \times \int_{w}^{\overline{w}} \left\{\frac{1}{v'(x(n))} - \frac{\Phi'(u(n))}{\lambda}\right\} f(n)\, dn \tag{9}$$

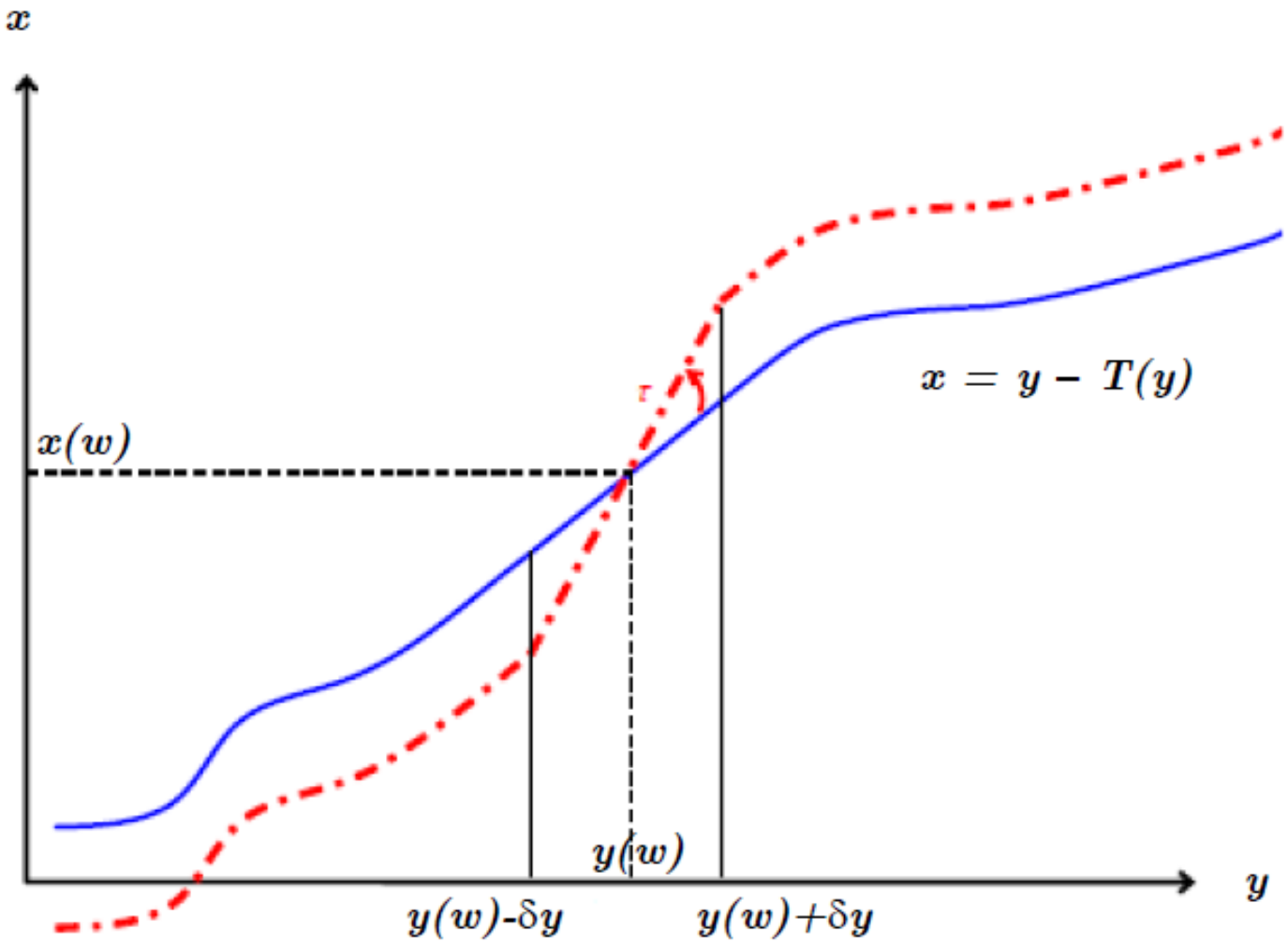
Interpretation of these FOC's in terms of behavioral elasticities

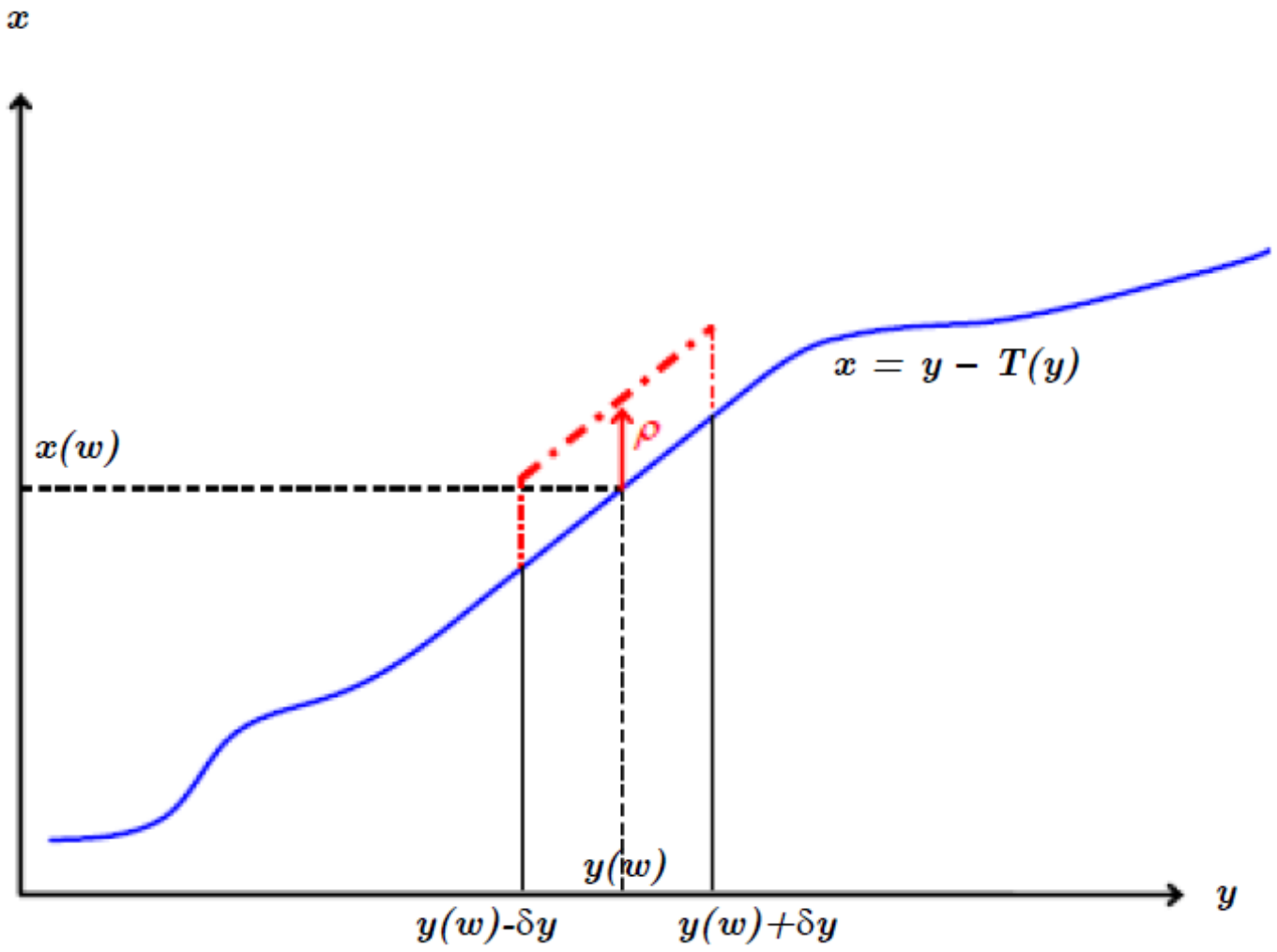
Tax perturbation approach, Saez (ReStud 2001)

Consider 2 types of tax perturbation, in the neighborhood of $y(w)$, the tax function becomes

$$y \longmapsto T(y) - \tau(y - y(w)) - \rho$$

- Uniform decrease in T' in the neighborhood of $y(w)$ (size of this change: τ), this perturbation captures the **substitution effect** along the optimal tax schedule for workers of skill w.
- Uniform decrease in the level of tax in the neighborhood of $y(w)$ (size of this change: ρ), this perturbation captures the **income effect along** the optimal tax schedule for workers of skill w.


x
x(w)
x = y − T(y)
y(w)
y
y(w)-δy
y(w)+δy


x
$x = y - T(y)$
$x(w)$
ρ
$y(w)$
y
$y(w)$-δy
$y(w)$+δy

Definitions of elasticities

We follow Jacquet, Lehmann and Van der Linden (2013 JET) in these definitions which **account for the nonlinearity of the income tax schedule.** Doing so, we avoid the use of *virtual densities* which prevail in Saez (ReStud 2001).

- Elasticity of y with respect to $\tau \equiv 1 - T'(y)$:

$$\varepsilon(w) \equiv \frac{1 - T'(y(w))}{y(w)} \frac{\partial y(w)}{\partial \tau}$$

- Response of before-tax income to a lump-sum increase in the level of taxation (level of taxation being denoted by ρ)

$$\eta(w) \equiv \frac{\partial y(w)}{\partial \rho}$$

- Elasticity of gross earning y with respect to w:

$$\alpha(w) \equiv \frac{w}{y(w)} \frac{\partial y(w)}{\partial w}$$

Compute the comparative static properties of the solution to the agents' optimization problem

$$\underset{y}{Max}\, v\left(y - T\left(y\right)\right) - h\left(\frac{y}{w}\right)$$

Using $\tau \equiv 1 - T'\left(y\right)$ (τ here is not the participation tax!), the FOC w.r.to y is

$$v'\left(y - T\left(y\right)\right)\tau - h'\left(\frac{y}{w}\right)\frac{1}{w} = 0 \tag{10}$$

Implicitly differentiating (10) w.r.to y and τ yields

$$\left[\tau^2 v''\left(y - T\left(y\right)\right) - h''\left(\frac{y}{w}\right)\frac{1}{w^2} - v'\left(y - T\left(y\right)\right)T''\left(y\right)\right]dy$$
$$+v'\left(y - T\left(y\right)\right)d\tau = 0$$

so that

$$\frac{\partial y}{\partial \tau} = \frac{-v'\left(y - T\left(y\right)\right)}{\tau^2 v''\left(y - T\left(y\right)\right) - h''\left(\frac{y}{w}\right)\frac{1}{w^2} - v'\left(y - T\left(y\right)\right)T''\left(y\right)}$$

Using (10) to re-express the numerator (as well as τ in the denominator) in the R.H.S. $\Rightarrow$

$$\varepsilon\left(w\right) \equiv \frac{1 - T'\left(y\left(w\right)\right)}{y\left(w\right)}\frac{\partial y\left(w\right)}{\partial \tau} = \frac{-h'\left(\frac{y}{w}\right)\frac{1}{w}}{y\left[\left(\frac{h'\left(\frac{y}{w}\right)\frac{1}{w}}{v'\left(y-T\left(y\right)\right)}\right)^2 v''\left(y - T\left(y\right)\right) - h''\left(\frac{y}{w}\right)\frac{1}{w^2} - v'\left(y - T\left(y\right)\right)T''\left(y\right)\right]}$$

Assume SOC of the individual maximization pgm holds strictly

The second-order condition associated to the individual maximization program is:

$$\tau^2 v''(y - T(y)) - h''\left(\frac{y}{w}\right)\frac{1}{w^2} - v'(y - T(y))\, T''(y) \leq 0$$

- The LHS is the denominator of our previous expression. If the SOC holds with a strict inequality, one can apply the **implicit function theorem** and the FOC implicitly defines $y(w)$ as a **(locally) differentiable function** of the skill level and the tax function $\Rightarrow$ We can **define our behavioral elasticities**.

Assume SOC of the individual maximization pgm holds strictly

- The second-order condition stipulates that the function $y \longmapsto y - T(y)$ is either concave or less convex than the indifference curve of workers of skill w at $(x(w), y(w))$. **Otherwise**, there is a **discontinuity of the tax function** (see Figure 3). The SOC is satisfied when the tax function is linear, convex or not "too concave".
- How the convexity of the tax function matters for the SOC depends on the term $T''(y)$ that captures the curvature of the tax function.

Implicitly differentiating (10) w.r.to y and ρ yields

$$\left[\tau^2 v''\left(y-T\left(y\right)\right)-h''\left(\frac{y}{w}\right)\frac{1}{w^2}-v'\left(y-T\left(y\right)\right)T''\left(y\right)\right]dy$$
$$-\rho v''\left(y-T\left(y\right)\right)d\rho=0$$

$$\eta\left(w\right)\equiv\frac{\partial y\left(w\right)}{\partial\rho}=$$
$$\frac{-\rho v''\left(y-T\left(y\right)\right)}{\left(\frac{h'\left(\frac{y}{w}\right)\frac{1}{w}}{v'\left(y-T\left(y\right)\right)}\right)^2 v''\left(y-T\left(y\right)\right)-h''\left(\frac{y}{w}\right)\frac{1}{w^2}-v'\left(y-T\left(y\right)\right)T''\left(y\right)}$$

Income effect of labor supply

$$\eta(w) \equiv \frac{\partial y(w)}{\partial \rho} = \frac{-\rho v''(y - T(y))}{\left(\frac{h'\left(\frac{y}{w}\right)\frac{1}{w}}{v'(y-T(y))}\right)^2 v''(y - T(y)) - h''\left(\frac{y}{w}\right)\frac{1}{w^2} - v'(y - T(y))\, T''(y)}$$

- $\eta(w)$ is response of before-tax income to a lump-sum increase in the level of taxation
- $\eta(w)$ captures the income effect of the labor supply.
- $\eta(w) < 0$ as long as leisure is a normal good.

Compensated elasticity of labor supply w.r.to retention rate

$$\varepsilon(w) \equiv \frac{1-T'(y(w))}{y(w)} \frac{\partial y(w)}{\partial \tau} = \frac{-h'\left(\frac{y}{w}\right)\frac{1}{w}}{y\left[\left(\frac{h'\left(\frac{y}{w}\right)\frac{1}{w}}{v'(y-T(y))}\right)^2 v''(y-T(y)) - h''\left(\frac{y}{w}\right)\frac{1}{w^2} - v'(y-T(y))\,T''(y)\right]}$$

- $\varepsilon(w)$ stands for the compensated elasticity of the labor supply with respect to $1-T'$.
- It is "compensated" because we leave unchanged the level of tax at $y(w)$.
- $\varepsilon(w) > 0$; a compensated decline in the marginal tax rates increases the marginal reward of effort in terms of additional consumption, which induces workers to substitute consumption for leisure.

Implicitly differentiating (10) w.r.to y and τ yields

$$\frac{\partial y}{\partial w} = \frac{-h'\left(\frac{y}{w}\right)\frac{1}{w^2} + h''\left(\frac{y}{w}\right)\frac{y}{w^3}}{\tau^2 v''\left(y - T\left(y\right)\right) - h''\left(\frac{y}{w}\right)\frac{1}{w^2} - v'\left(y - T\left(y\right)\right)T''\left(y\right)}$$

and

$$\alpha\left(w\right) \equiv \frac{w}{y\left(w\right)}\frac{\partial y\left(w\right)}{\partial w} = \frac{-\left[h'\left(\frac{y}{w}\right)\frac{1}{w} + h''\left(\frac{y}{w}\right)\frac{y}{w}\right]}{y\left[\left(\frac{h'\left(\frac{y}{w}\right)\frac{1}{w}}{v'\left(y-T\left(y\right)\right)}\right)^2 v''\left(y - T\left(y\right)\right) - h''\left(\frac{y}{w}\right)\frac{1}{w^2} - v'\left(y - T\left(y\right)\right)T''\left(y\right)\right]}$$

Elasticity of earning w.r.to skill

$$\alpha(w) \equiv \frac{w}{y(w)} \frac{\partial y(w)}{\partial w} = \frac{-\left[h'\left(\frac{y}{w}\right)\frac{1}{w} + h''\left(\frac{y}{w}\right)\frac{y}{w}\right]}{y\left[\left(\frac{h'\left(\frac{y}{w}\right)\frac{1}{w}}{v'(y-T(y))}\right)^2 v''(y-T(y)) - h''\left(\frac{y}{w}\right)\frac{1}{w^2} - v'(y-T(y))\,T''(y)\right]}$$

- $\alpha(w)$ stands for the elasticity of earning with respect to w.
- $\alpha(w) > 0 \Rightarrow$ We retrieve that along an incentive-compatible allocation, $y(w)$ have to be a nondecreasing with w.
- Bunching occurs only when $\alpha(w) = 0$, i.e. when the curvature term $T''(y(w))$ in the denominator tends to infinity, i.e. kink in the tax function (see Figure 3).

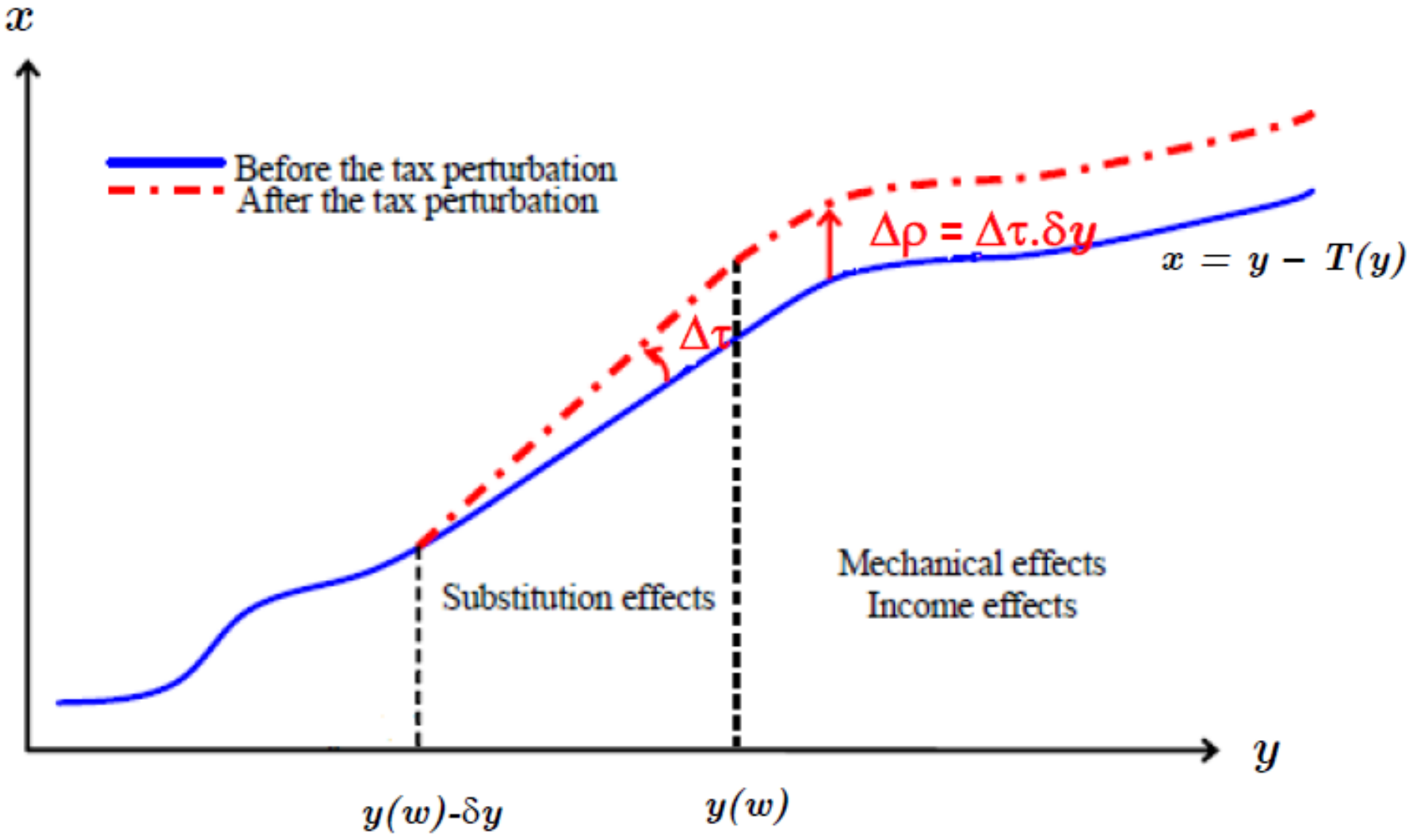
Endogeneity of behavioral elasticities

Behavioral elasticities/responses are endogenous:

- They depend on $(x(w), y(w))$ which is endogenous.
- $T''(y)$ in their denominators which account for the nonlinearity of the tax schedule ($\neq$ from Saez (ReStud 2001)).They depend on the curvature of the tax function, as captured by the term term $T''(y)$ in their denominators; they include a(n endogenous) "**circular process**": an exogenous change in either w, τ or ρ induces a change in earnings $\Delta_1 T' = T''(y(w)) \Rightarrow$ change in T' by $\Delta_1 T' = T''(y(w)) \times \Delta_1 y(w) \Rightarrow$ further change in y (due to substitution effects).
- The literature often considers only the "direct effects" by **assuming that marginal tax rates are exogenous** in the **computation of elasticities and income responses**, thereby taking $T''(y(w)) = 0$.

Tax perturbation method to obtain optimal tax formula

Consider a uniform decrease of $T'(y(w))$ over $[y(w) - \delta y, y(w)]$.
$\Rightarrow$ Tax function unchanged for earnings below $y(w) - \delta y$
$\Rightarrow$ Tax function uniformly decreased by an amount $\Delta\rho = \Delta\tau \times \delta y$ for earnings above $y(w)$.


x
Before the tax perturbation
After the tax perturbation
$\Delta\rho = \Delta\tau.\delta y$
$x = y - T(y)$
$\Delta\tau$
Substitution effects
Mechanical effects
Income effects
y
$y(w)$-δy
$y(w)$

Tax perturbation method to obtain optimal tax formula

To apply the tax perturbation method, we need **further regularity assumptions:**

The optimal allocation is such that Y is **differentiable everywhere** in skill, there is **no bunching** and the tax function T is **twice differentiable everywhere** in earnings.

Mechanical effects

Absent any behavioral response (and income effect), for each taxpayer of skill n above w:

- the government receives $\Delta\rho$ Euros of tax less from each of these workers.
- welfare of these individuals increases by $v'(x(n))\Delta\rho$, which is valued $\Phi'(u(n))v'(x(x))\Delta\rho/\lambda$ Euros by the government.

$\Rightarrow$ Total mechanical effect for all tax payers with skills above w equals

$$M(w) = -\Delta\rho \int_{w}^{\overline{w}} \left\{ 1 - \frac{\Phi'(u(n))v'(x(x))}{\lambda} \right\} f(n)\, dn$$

Income effects

Workers of skill n above w change their labor supply decisions because of the income effect. Their earnings change by $\Delta y(n) = \eta(n)\Delta\rho$ which induces a change in tax revenue of $T'(y(n))\eta(n)\rho$. This behavioral response has only a second-order effect on the social objective.
$\Rightarrow$ Total income effect for all tax payers with skills above w equals

$$I(w) = \Delta\rho \int_{w}^{\overline{w}} T'(y(n))\eta(n) f(n)\, dn$$

Substitution effect

Workers whose y before the reform lie in the interval $[y(w) - \delta y, y(w)]$ of the earnings distribution have a productivity that belongs to an interval $[w - \delta w, w]$ of the skill distribution.
The elasticity $\alpha(w)$ of earnings with respect to w links the widths of these two intervals through

$$\delta w = \frac{w}{\alpha(w)\, y(w)} \delta y$$

Therefore, the number of these individuals is

$$f(w)\, \delta w = \frac{w f(w)}{\alpha(w)\, y(w)} \delta y$$

Substitution effect (cont'd)

Each of them face a decline $\Delta\tau$ of their T' and then substitute consumption for leisure so their earnings increase by

$$\Delta y(w) = \frac{\varepsilon(w) y(w)}{1 - T'(y(w))} \Delta\tau$$

Hence each of them generates $T'(y(w)) \Delta y(w)$ additional tax to the government. Since their change of labor supply induces only a second-order effect on the utility of these individuals (and social welfare) and since $\Delta\rho = \Delta\tau \cdot \delta y$, the substitution effect is valued, by the government, as

$$S(w) = \frac{T'(y(w))}{1 - T'(y(w))} \frac{\varepsilon(w)}{\alpha(w)} w f(w) \Delta\rho$$

Starting from the optimal tax schedule, a tax perturbation should have no first-order effect. Therefore $M(w) + I(w) + S(w) = 0$:

$$\frac{T'(w)}{1 - T'(w)} = \tag{11}$$

$$\underbrace{\frac{\alpha(w)}{\varepsilon(w)}}_{A(w)} \cdot \underbrace{\frac{\int_{w}^{\overline{w}} \left\{1 - \frac{\Phi'(u(n))v'(x(x))}{\lambda} - \eta(n)\, T'(y(n))\right\} f(n)\, dn}{1 - F(w)}}_{B(w)} \cdot \underbrace{\frac{1 - F(w)}{wf(w)}}_{C(w)}$$

The derivation of this optimal tax formula in terms of **sufficient statitics** was heuristic $\Rightarrow$ **We still have to verify that (11) is consistent with (9)**.

Detailed proof in e.g. Jacquet, Lehmann and Van der Linden (JET 2013), Jacquet and Lehmann (WP Thema 2016) or in Boadway, Brett and Jacquet (WP Thema 2015).

Efficiency term A(w)

$$\frac{T'(w)}{1-T'(w)} =$$

$$\underbrace{\frac{\alpha(w)}{\varepsilon(w)}}_{A(w)} \cdot \underbrace{\frac{\int_{w}^{\overline{w}} \left\{1 - \frac{\Phi'(u(n))v'(x(x))}{\lambda} - \eta(n)\, T'(y(n))\right\} f(n)\, dn}{1-F(w)}}_{B(w)} \cdot \underbrace{\frac{1-F(w)}{wf(w)}}_{C(w)}$$

$\Rightarrow$ A higher efficiency term $A(w)$ implies a lower T' (ceteris paribus).

Efficiency term A(w)

Numerator: $\alpha(w)$ of $A(w) = \frac{\alpha(w)}{\varepsilon(w)}$:

A higher skill elasticity $\alpha(w)$ reduces the **income density** $h(y(w))$ hence the **distortions due to substitution effects** $S(w)$.
$\Rightarrow$ Ceteris paribus, **higher** $\alpha(w)$ **implies a higher** T' (Jacquet, Lehmann and Van der Linden JET 2013).

To see this, note that

$$y(w)\, h(y(w)) = \frac{wf(w)}{\alpha(w)}$$

where $h(y)$ denote the income density and $H(y)$ the corresponding CDF. Differentiating both sides of $H(y(w)) = F(w)$ and using $\alpha(w) \equiv \frac{w}{y(w)} \frac{\partial y(w)}{\partial w}$, we obtain the above equality.

Efficiency term A(w)

Denominator: $\varepsilon(w)$ of $A(w) = \frac{\alpha(w)}{\varepsilon(w)}$:

A higher skill elasticity $\varepsilon(w)$ increases **distortions due to substitution effects** $S(w)$
$\Rightarrow$ Ceteris paribus, **higher** $\varepsilon(w)$ **implies a lower** T' (inverse Ramsey rule)

Efficiency term A(w)

Remark:

$\alpha(w)/\varepsilon(w)$ is a ratio $\Rightarrow$ $T''(.)$ that appear in the denominator of $\alpha(w)$ as well as in the denominator of $\varepsilon(w)$ cancel out.

$\Rightarrow$ The efficiency term is the same whether we define behavioral elasticities $\alpha(w)$ and $\varepsilon(w)$ along the optimal tax schedule (as in Jacquet, Lehmann and Van der Linden JET 2013) or along a linear tax schedule (as in Saez 2001).

Density term C(w)

$$C(w) = \frac{1 - F(w)}{wf(w)}$$

Increasing the marginal tax rate of type-w individuals:

- increases **distortions due to substitution effects**; disincentive proportional to $wf(w)$
- but allows to **increase the tax paid** by the $(1 - F(w))$ individuals **above** w

(Diamond AER 1998)

Equity term B(w)

$$B(w) = \frac{\int_{w}^{\overline{w}} \left\{1 - \frac{\Phi'(u(n))v'(x(x))}{\lambda} - \eta(n)\, T'(y(n))\right\} f(n)\, dn}{1 - F(w)}$$

$B(w)$ is the average of **mechanical** and **income effects** above skill w.

- The term $1 - \frac{\Phi'(u(n))v'(x(x))}{\lambda} - \eta(n)\, T'(y(n))$ captures the total cost for the government **to decrease by one unit the level of tax paid by workers of skill** n, including their change in labor supply due to the income effect.

Equity term B(w)

- The government values giving one more euro to each of the $f(n)$ individuals of skill n as a gain of $\frac{\Phi'(u(n))v'(x(x))}{\lambda}$ of government spending. $\Phi(.)$ is increasing and concave, so $\Phi'(u(n))$ is positive and decreases in $u(n)$. We have shown that $u(n)$ is increasing in skill. Therefore $\Phi'(u(n))$ is positive and **decreasing** in skill.
- We have shown that $x(x)$ is nondecreasing in skill. As $v(.)$ is increasing and weakly concave, $v'(x(n))$ is **nonincreasing** in skill.

Equity term B(w)

$\Rightarrow$ As a consequence, the mechanical term $1 - \frac{\Phi'(u(n))v'(x(x))}{\lambda}$ is **increasing in skill**.

Therefore, in the absence of income effects (i.e. if $\eta(n) = 0$), like in Diamond (AER 1998), $B(w)$ hence T' increases with w (thereby with y), ceteris paribus.

Equity term B(w)

$$B(w) = \frac{\int_{w}^{\overline{w}} \left\{1 - \frac{\Phi'(u(n))v'(x(x))}{\lambda} - \eta(n)\, T'(y(n))\right\} f(n)\, dn}{1 - F(w)}$$

- Typically, leisure is a normal good : $\eta(n) < 0$. Rising marginal tax rate at one earning level induces, through **income effects,** more labor supply for all tax payers above.
- Larger $\eta(n)$ in absolute value tend(s) to increase $B(w)$ hence T', ceteris paribus.

Optimal tax formula in terms of earnings y

Using tax perturbation $(M(y) + I(y) + S(y) = 0)$ or structural approach (i.e. from the FOC), it is also very easy to derive tax schedule in terms of y:

$$\frac{T'(y)}{1 - T'(y)} = \underbrace{\frac{1}{\varepsilon(y)}}_{A(y)} \cdot \underbrace{\frac{\int_{y}^{y(\overline{w})} \left\{1 - \frac{\Phi'(u(y))v'(.)}{\lambda} - \eta(y)\, T'(y)\right\} h(y)\, dy}{1 - H(y)}}_{B(y)} \cdot \underbrace{\frac{1 - H(y)}{yh(y)}}_{C(y)}$$

with some abuse of notations.

Marginal social welfare weight

Define marginal social welfare weight associated with y-workers by:

$$g(y) \equiv \frac{\Phi'(u(y))\, v'}{\lambda}$$

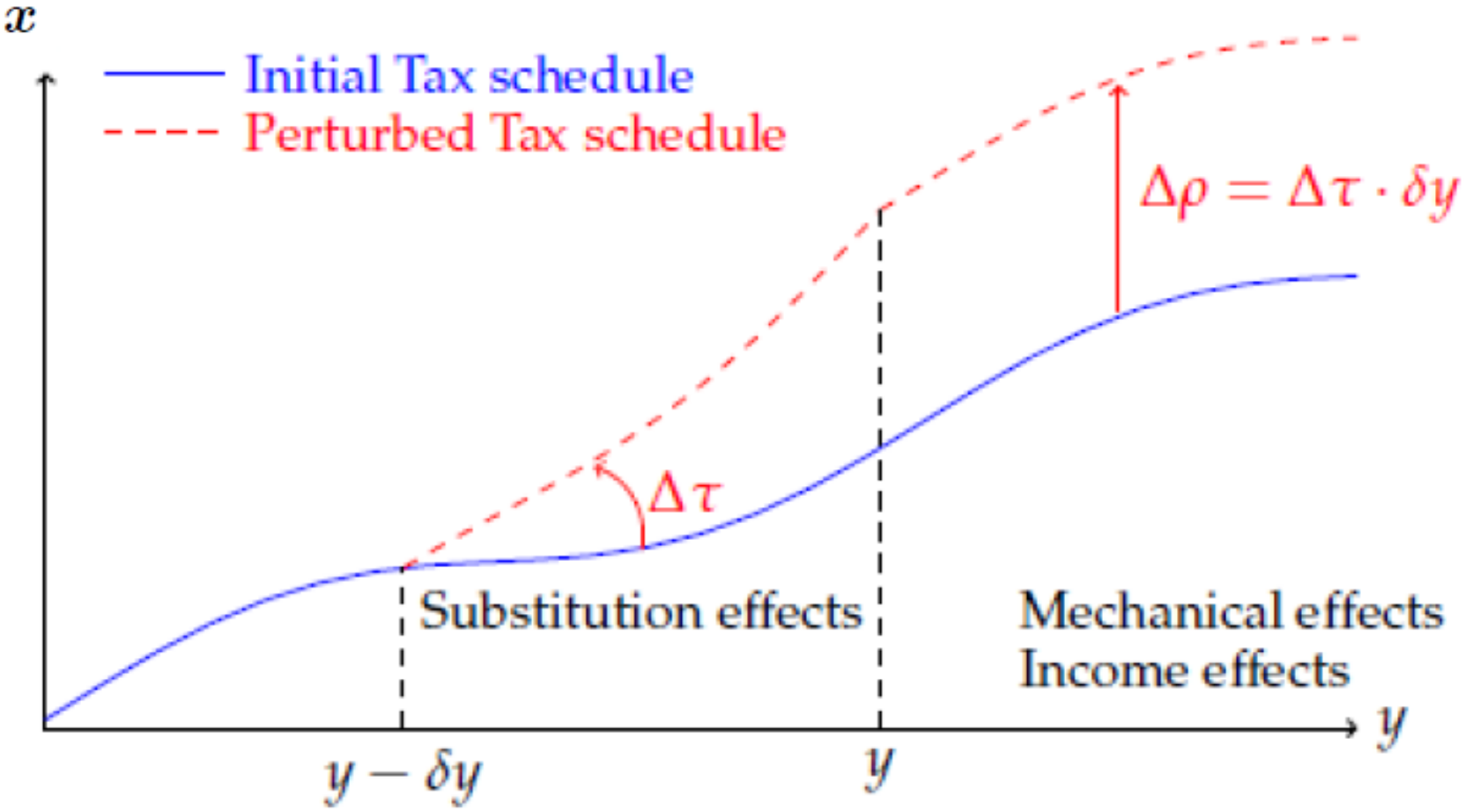
The government values giving one extra dollar to a y-worker as a gain of $g(y)$ dollar(s) of public funds.

This is the social marginal value of consumption for worker with income y at the optimum expressed in terms of the value of public funds (Saez 2001).

Tax perturbation method to obtain optimal tax formula

Consider a uniform decrease $\Delta\tau$ of T' over $[y - \delta y, y]$.

$\Rightarrow$ Tax function unchanged for earnings below $y - \delta y$
$\Rightarrow$ Tax function uniformly decreased by an amount $\Delta\rho = \Delta\tau \cdot \delta y$ for earnings above y.


x
Initial Tax schedule
Perturbed Tax schedule
$\Delta\rho = \Delta\tau \cdot \delta y$
$\Delta\tau$
Substitution effects
Mechanical effects
Income effects
y
$y - \delta y$
y

Mechanical effects

Absent any behavioral response and an income effect, for taxpayers whose income levels were above y before reform:

- the government receives $\Delta\rho$ Euros of tax less from each of these workers.
- welfare of these individuals increases which is valued $g(y)$ Euros by the government.

$\Rightarrow$ Total mechanical effect for all tax payers with incomes above y equals

$$M(y) = -\Delta\rho \int_{y}^{y(\overline{w})} \{1 - g(y)\} h(y) \, dy$$

Income effects

Workers with earnings above y are induced to reduce their labor supply because of the **income effect**.
Their earnings change by $\Delta y(n) = \eta(n)\Delta\rho$ which induces a change in tax revenue of $T'(y(n))\eta(n)\Delta\rho$.
$\Rightarrow$ Total income effect for all tax payers whose incomes above y equals

$$I(y) = \Delta\rho \int_{y}^{y(\overline{w})} T'(y)\eta(y)h(y)\,dy$$

where

$$\eta(y) \equiv \frac{\partial y}{\partial \rho}$$

Substitution effect

The lower T' implies that individuals whose income before the tax reforms lies within $[y - \delta y, y]$ increase their income by Δy due to a substitution effect:
Each of them face a decline τ of their T' and then substitute consumption for leisure. Their earnings increase by

$$\Delta y = \frac{\varepsilon(y) y}{1 - T'(y)} \Delta \tau$$

where

$$\varepsilon(y) \equiv \frac{1 - T'(y)}{y} \frac{\partial y}{\partial \tau}$$

Substitution effect (cont'd)

Hence each of them generates $T'(y)\Delta y$ additional tax to the government.

Since their change of labor supply induces only a second-order effect on the utility of these individuals (and social welfare) and since $\Delta\rho = \Delta\tau \cdot \delta y$, the substitution effect is valued, by the government, as

$$S(y) = \frac{T'(y)}{1 - T'(y)} \varepsilon(y)\, yh(y)\, \Delta\rho$$

Starting from the optimal tax schedule, a tax perturbation should have no first-order effect. Therefore $M(y) + I(y) + S(y) = 0$:

$$\frac{T'(y)}{1 - T'(y)} = \underbrace{\frac{1}{\varepsilon(y)}}_{A(y)} \cdot \underbrace{\frac{\int_{y}^{y(\overline{w})} \{1 - g(y) - \eta(y)\, T'(y)\}\, h(y)\, dy}{1 - H(y)}}_{B(y)} \cdot \underbrace{\frac{1 - H(y)}{yh(y)}}_{C(y)}$$

This optimal tax formula is written in terms of **sufficient statistics**.
Remark: Again, we have included the circularity into the definitions of behavioral elasticities and income responses.

Sign of marginal tax rates for interior skill levels

- $T'(.)$ below 100% and cannot be negative

Mirrlees (ReStud, 1971)

With negative T', people are working too much (due to substitution effects) $\Rightarrow$ Increase T' improves efficiency and also equity by distributing the proceeds lump sum.

Marginal tax rates at the top

At the very top:

If the skill distribution is **bounded**, then the transversality condition $q(\overline{w}) = 0$ and (7) implies

$$T'(\overline{w}) = 0$$

i.e. zero marginal tax rate at the very top, Sadka (ReStud, 1976), Seade (JPubE, 1977).

Marginal tax rates at the top (cont'd)

However, Diamond (AER 1998) argues that when the distribution of skill is **unbounded**, $T' > 0$ at the very top. More specifically, Diamond (1998) argues that empirically, the distribution of high skills is approximated by a Pareto for which $\frac{1-F(w)}{wf(w)}$ is constant.

Saez (ReStud 2001) made a similar point about income distributions and the pattern of $\frac{1-H(y)}{yh(y)}$.

Diamond and Saez (JEP 2011): Top is uncertain: actual distribution is finite draw from an underlying Pareto distribution.

C(w) computed in Diamond (AER 1998)

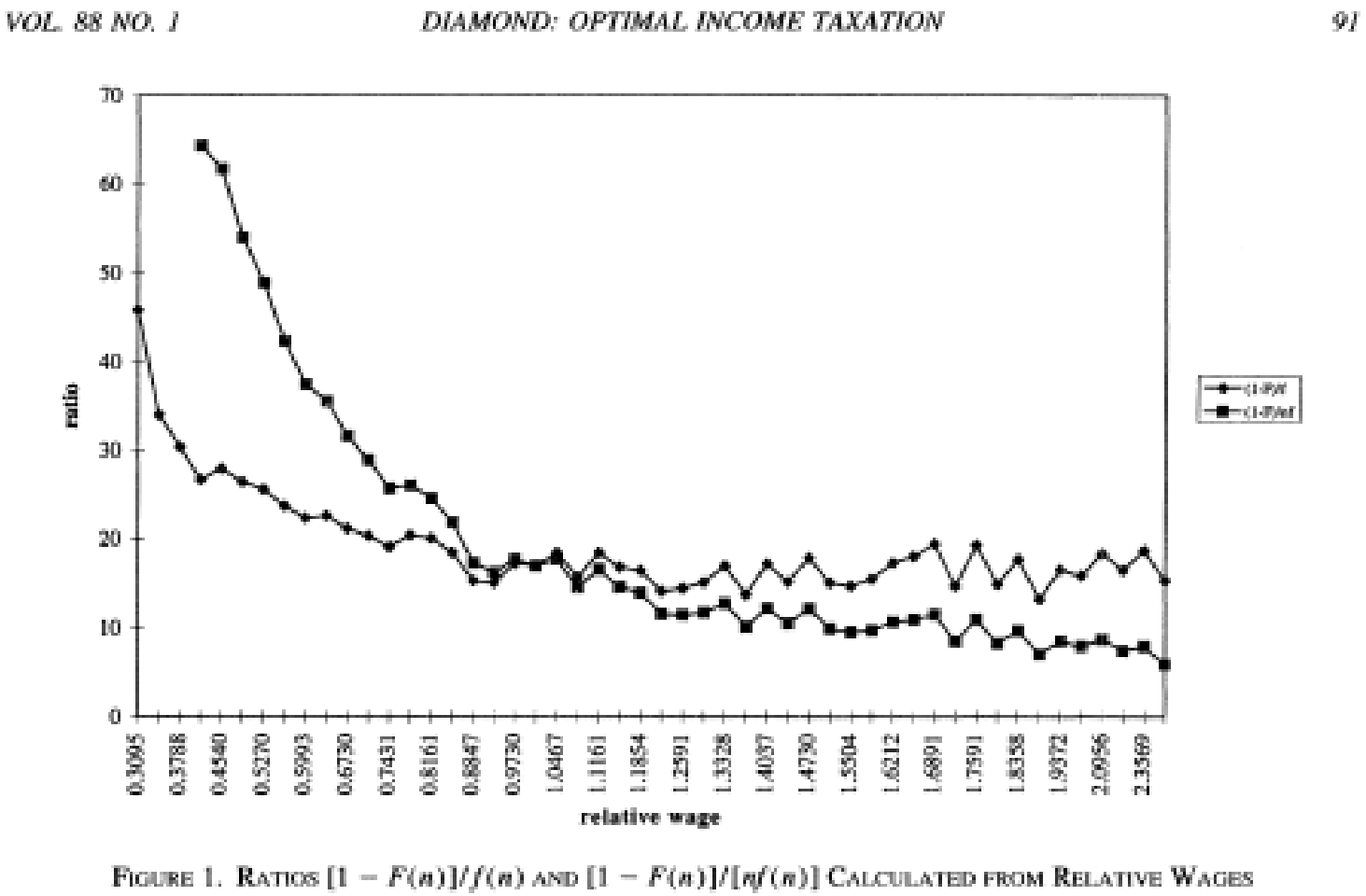


FIGURE 1. RATIOS $[1 - F(n)]/f(n)$ AND $[1 - F(n)]/[nf(n)]$ CALCULATED FROM RELATIVE WAGES

$C(w)$ from Diamond (1998, AER)

C(y) computed in Saez (ReStud 2001)

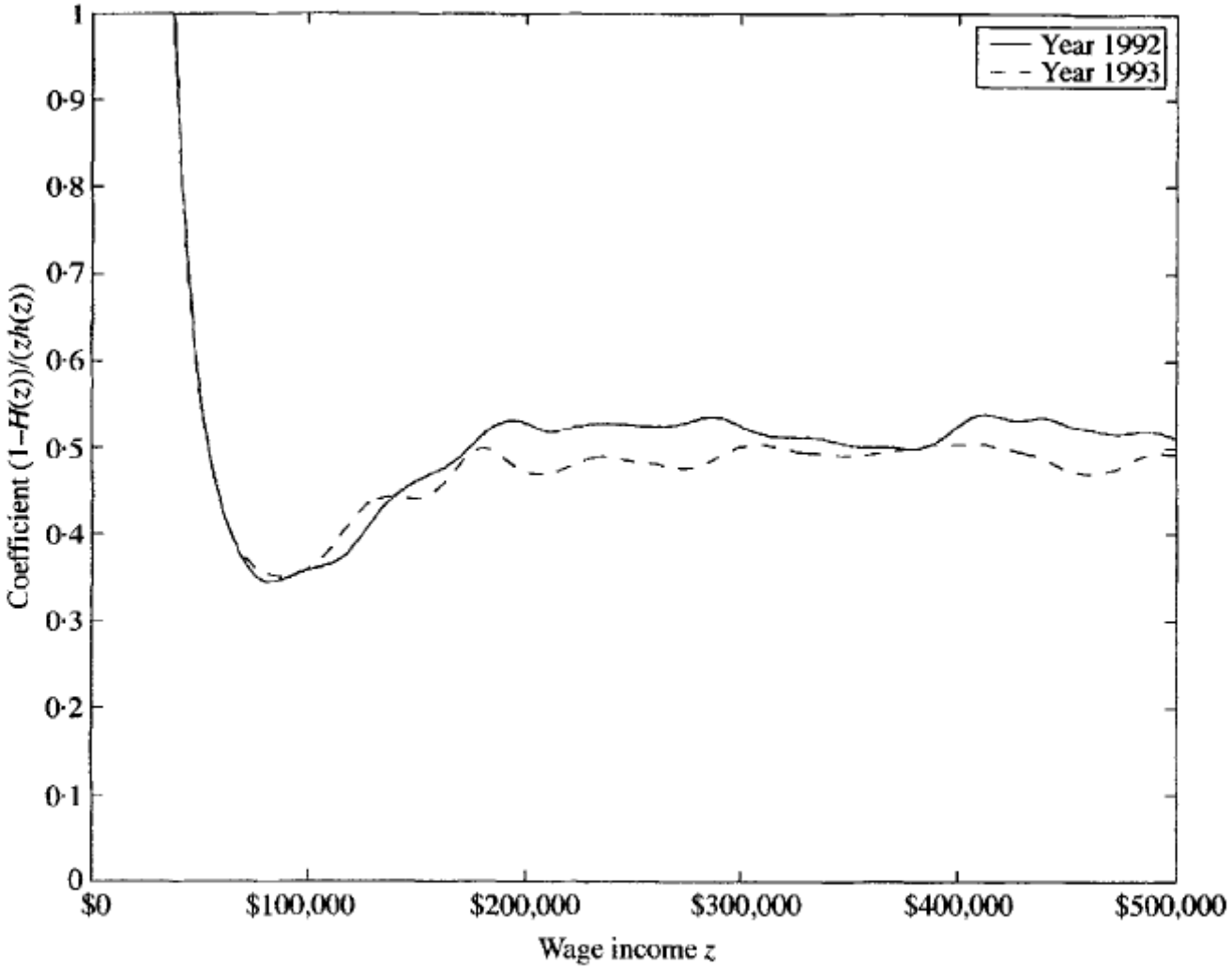


FIGURE 4

Hazard ratio $(1 - H(z))/(zh(z))$, years 1992 and 1993

Optimal tax formula for top incomes

- Assume no income effects.
- Assume that the top tax rate above a fixed income level y^* is constant and equal $\overline{T}'$
- Denote $\overline{y}$ the average income in top bracket
- Denote $\overline{\varepsilon}$ the average elasticity of income w.r.to net-of-tax rate $1 - \overline{T}'$ for top bracket individuals
- Denote $\overline{g}$ the average marginal social welfare weight for individuals in top bracket
- Denote a the (top) tail-parameter of the Pareto income distribution

Optimal tax formula for top incomes (cont'd)

Saez (2001), Piketty and Saez (201, pp.423-424), Optimal top tax rate:

$$\overline{T}' = \frac{1-\overline{g}}{1-\overline{g}+a\overline{\varepsilon}}$$

Interpretation:

- $\overline{T}'$ decreases $\overline{g}$. In the limit case where Gvt does not put any value on the marginal consumption of top earners, the formula simplifies to $\overline{T}' = 1/\left(1+a\overline{\varepsilon}\right)$ which is the revenue maximizing top tax rate (Diamond and Saez JEP 2011).
- $\overline{T}'$ decreases with $\overline{\varepsilon}$ (inverse Ramsey rule).
- $\overline{T}'$ decreases with $a \geq 1$ which measures the thinness of the top tail of the income distribution.

Marginal tax rate at the top: Migration effects

Migration issues may be particularly important at the top end (brain drain). Some theory papers: Mirrlees (JPubE 1982), Lehmann, Simula,Trannoy (QJE 2014).
Migration depends on average tax rate.
Define $P(y - T(y)|\, y)$ fraction of y earners in the country.
Define migration elasticity as:

$$\eta^m = \frac{y - T(y)}{P} \frac{\partial P}{\partial (y - T(y))}$$

Marginal tax rate at the top: Migration effects (cont'd)

Optimal top tax rate can now be written as:

$$\overline{T}' = \frac{1-\overline{g}}{1-\overline{g}+a\overline{\varepsilon}+\eta^m}$$

- Ceteris paribus, a larger $\eta^m \Rightarrow$ lower T'.
- η^m depends on size of jurisdiction (large for cities, zero worldwide),
- Redistribution easier in large jurisdictions,
- Tax coordination across countries increases ability to redistribute. Big issue currently in EU.

Marginal tax rates at the bottom

- In the absence of bunching at the bottom (i.e. If everybody works) and if the government has no maximin objective:

$$T'(.) = 0 \text{ at the (very) bottom,}$$

Seade (ReStud 1977)

Proof. The transversality condition $q(\underline{w}) = 0$, (1) and (7) yield this result.

Marginal tax rates at the bottom: Maximin

- If maximin:

$$T'(.) \quad > \quad 0 \text{ at the (very) bottom,}$$

Boadway and Jacquet (JET 2008)

Proof. Social weights are concentrated on the lowest skill level only. So, there is a positive mass of social weights there, even without bunching at the bottom. In contrast, social weights $= 0$ for all $w > \underline{w}$.

$\Rightarrow$ the term $B(w) \geq 1$ (depending on whether income effects are nil, or negative when leisure is a normal good), **including for** $w = \underline{w}$.

Marginal tax rates at the bottom: Bunching

- If there is bunching of non workers at the bottom (for instance because of a non-negativity constraint $y(w) \geq 0$ is binding at the bottom):

$T'(.) > 0$ at the bottom (i.e. at the end of the bunching interval)
Ebert (JPubE, 1992)

Proof.
There is a positive mass of workers with the lowest earning level, and the highest skilled workers among them have a skill $w > \underline{w}$ hence face a positive marginal tax rate.

Optimal marginal tax rates at the bottom

Mirrlees model predicts that optimal transfer at bottom takes the form of a **Negative Income Tax**:

- Lump-sum grant $-T(0)$ for those with no earnings
- High T' at the bottom to phase-out the lump-sum grant quickly

Intuition: high T' at bottom are efficient because:
(a) they target transfers to the most needy.
(b) earnings at the bottom are low to start with so intensive response does not generate large output losses.

Other theoretical contributions

And IC_2? See Lollivier and Rochet (JET, 1983) for the case $U(x, \ell) = v(x) - \ell$. They do not need regularity assumption.

Two types of bunching:

- due to the violation of the SOIC conditions: Lollivier and Rochet (JET, 1983)
- due to a binding non-negativity income constraint: See Boadway, Cuff and Marchand (JPET, 2000)

Calibration of the model using real data

e.g. Saez (ReStud 2001), US data.

Two specifications of utility functions:
Type I: $\log\left(x - \frac{\ell^{1+1/k}}{1+k}\right)$; corresponds to Diamond (AER 1998) without income effects ($\eta(w) = 0$), $\frac{\alpha(w)}{\varepsilon(w)} = 1 + k$.
Type II: $\log x - \log\left(1 - \frac{\ell^{1+1/k}}{1+k}\right)$

Remark: Under both utility functions, the compensated elasticity of labor supply (along a linear tax schedule) simply equals $1/k$ and is exogenous.

Calibration of the model using real data

Two values for the compensated elasticity of labor supply, ζ^c in Saez' notations: 0.25 and 0.5.

Numerical simulations: method

$H(y)$ and also $g(y)$'s are endogenous to $T(.)$.

- Take data of gross labor earnings
- **Specify utility** function (e.g. constant elasticity):

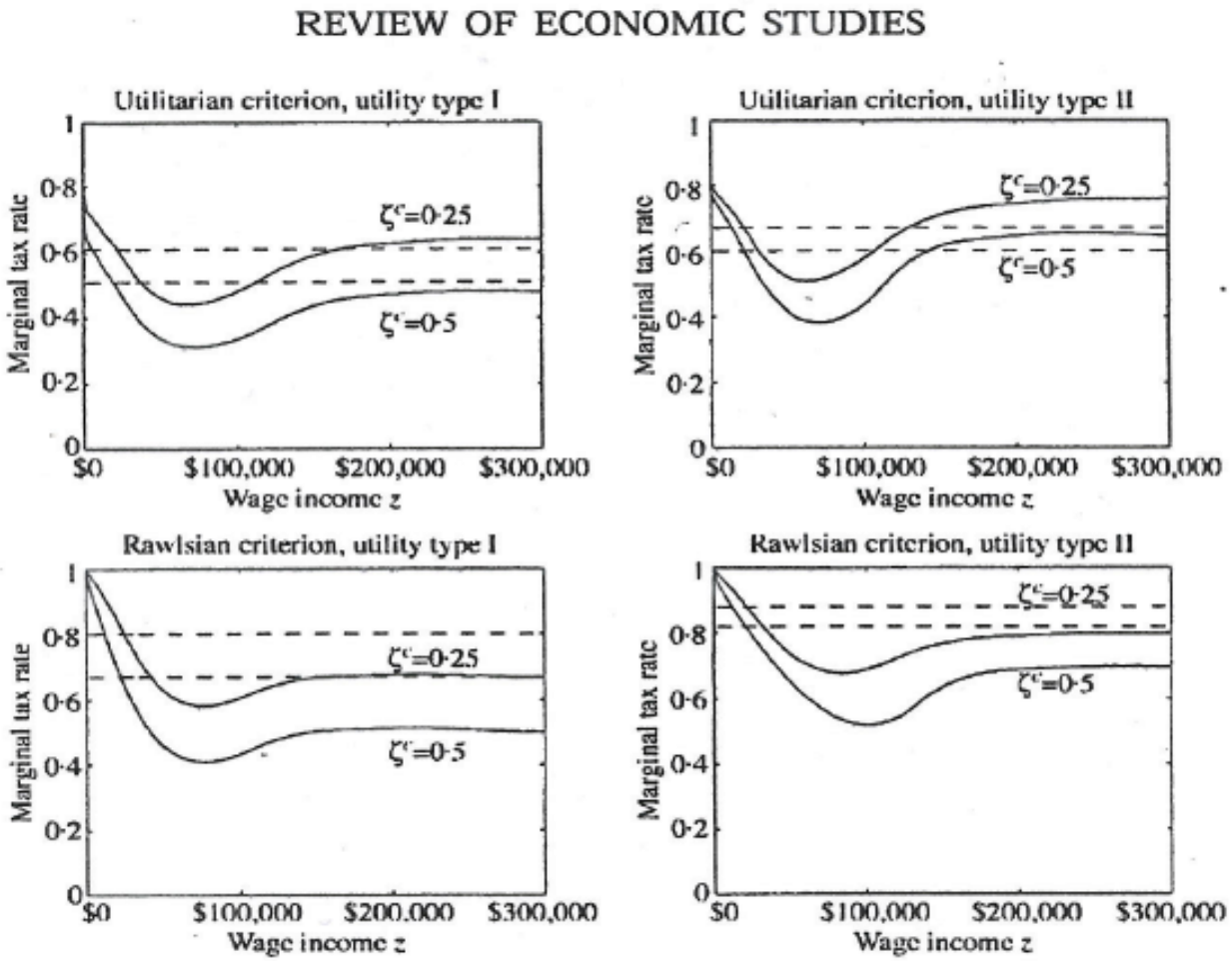
e.g. $u(x, \ell) = x - \frac{1}{1+\frac{1}{e}} \left(\frac{y}{n}\right)^{1+\frac{1}{e}}$

- **From the individual F.O.C., rewrite the gross earnings as a function of the skill level, other (possible) parameters and** T'.

e.g. Individual FOC $\Rightarrow y = n^{1+e} (1 - T') e$

- **Calibrate the exogenous skill distribution** $F(w)$ so that, using **actual** $T'(.)$, you recover **empirical** $H(y)$.
- Use optimal tax formula **expressed in terms of skills** to obtain the optimal tax rate schedule T'.

Calibration of the model using real data



Saez (2001).

Numerical simulations

Optimal $T'(y)$ is U-shaped!

Outline

Extensive margin

The extensive margin implies a discrete choice between:

- Not working/Not participating to the labor market.
- Working at least a strictly positive, and typically fairly large, number of hours.

Fixed utility costs of working can be psychological costs, cost of getting organized, commuting cost.

Diamond (JPubE 1980), Laroque (Ecta 2005) and Saez (QJE 2002) study the optimal income tax model with extensive margin only. Saez (QJE 2002) and Jacquet, Lehmann and Van der Linden (JET 2013) study the optimal income tax model with **both** intensive and extensive margins.

Optimal taxation when labor is along the extensive margin

$g(y) \equiv [\Phi'(u(.))\, v'(.)] / \lambda$ denote the value the government places on increasing income of individuals with income y.

Define the participation tax as $\tau(y) \equiv [T(y) - T(0)] / y$ as earlier or rather $\tau(y) \equiv [T(y) + b] / y$ with $b \geq 0$: welfare benefit, demogrant for the non-employed.

Diamond (JPubE 1980), Saez (QJE 2002) participation choice/extensive margin model

- Workers differ both in skill and disutility of work
- Workers can participate or not
- **Focus for now only on participation in job for own skill**
- Gvt can observe skill (= earnings) of those working, not of non-participants
- Perfectly elastic supply of jobs for each skill level
- Income and effort per job fixed (and increasing in skill)

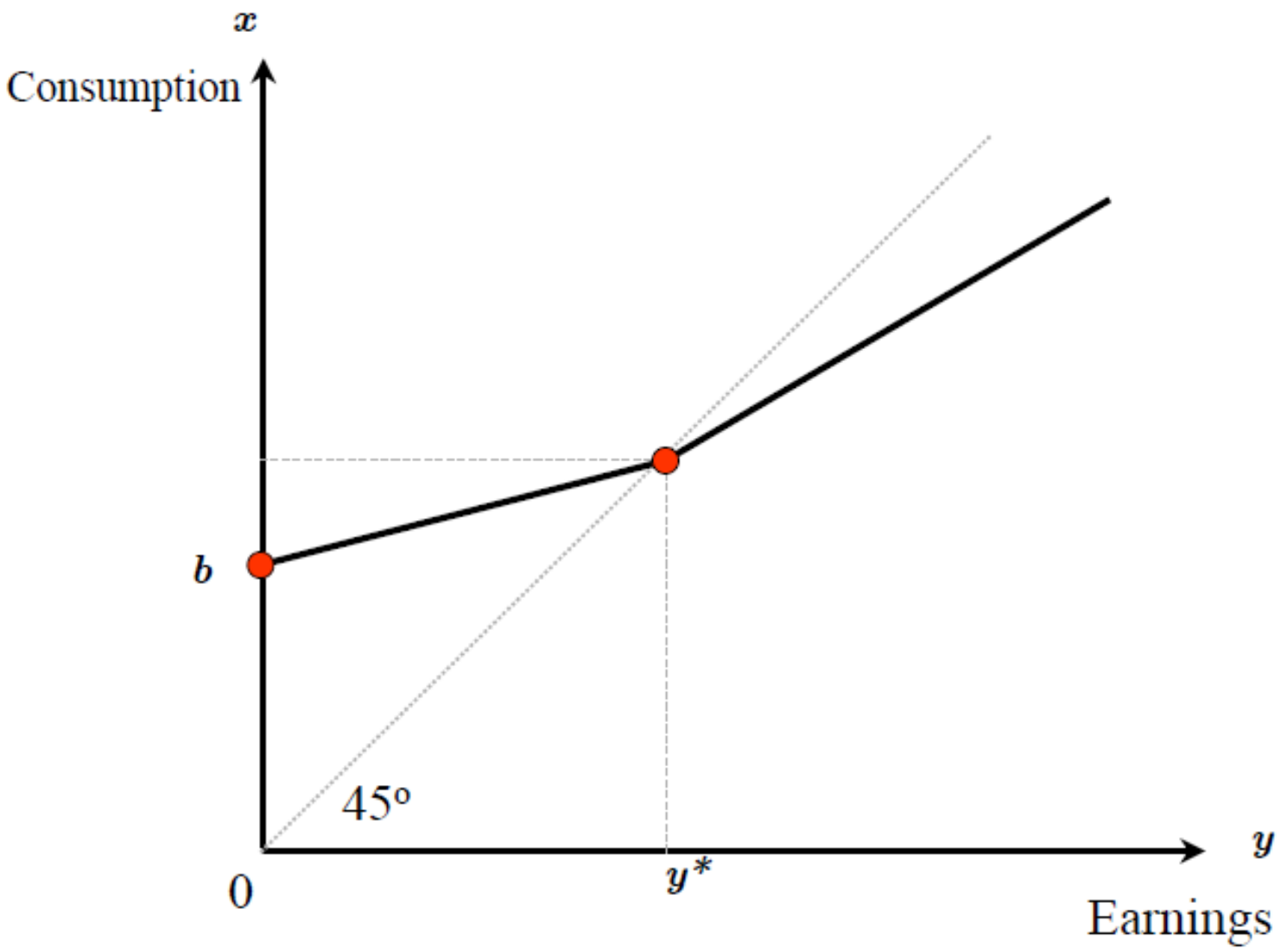
Saez (QJE 2002): participation/extensive margin

- The elasticity κ measures the percentage number of employed workers (at a given y) who decide to leave the labor force when the difference between their net earnings (i.e. consumption) in employment and non-employment decreases by 1%.

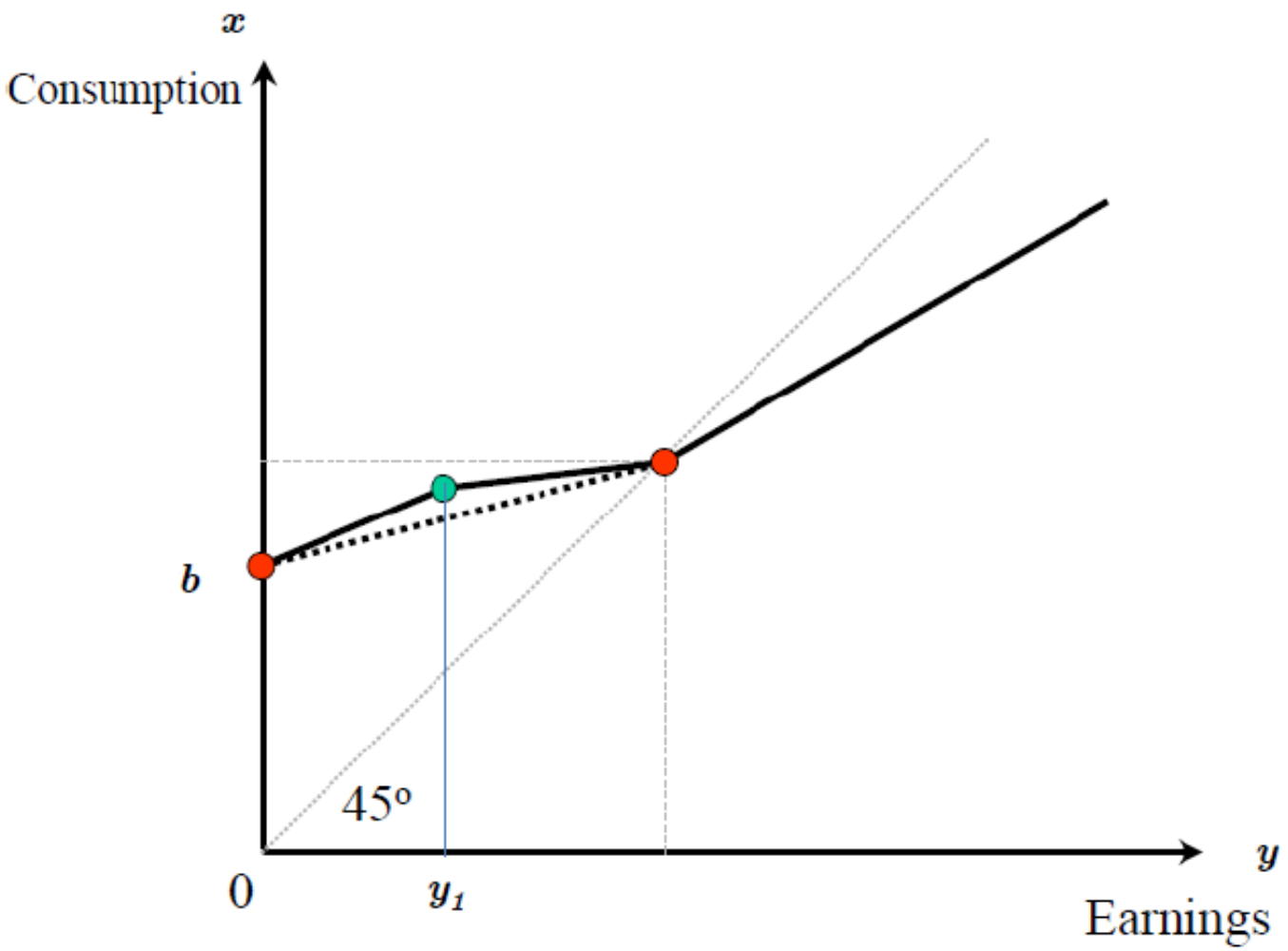
- Participation elasticity

$$\kappa_i \equiv \frac{\partial h_i}{\partial x_i} \frac{x_i}{h_i}$$

Start from a Negative Income Tax (NIT) i.e. apositive phasing-out rate



Introducing a "very small EITC" is desirable for tax revenue

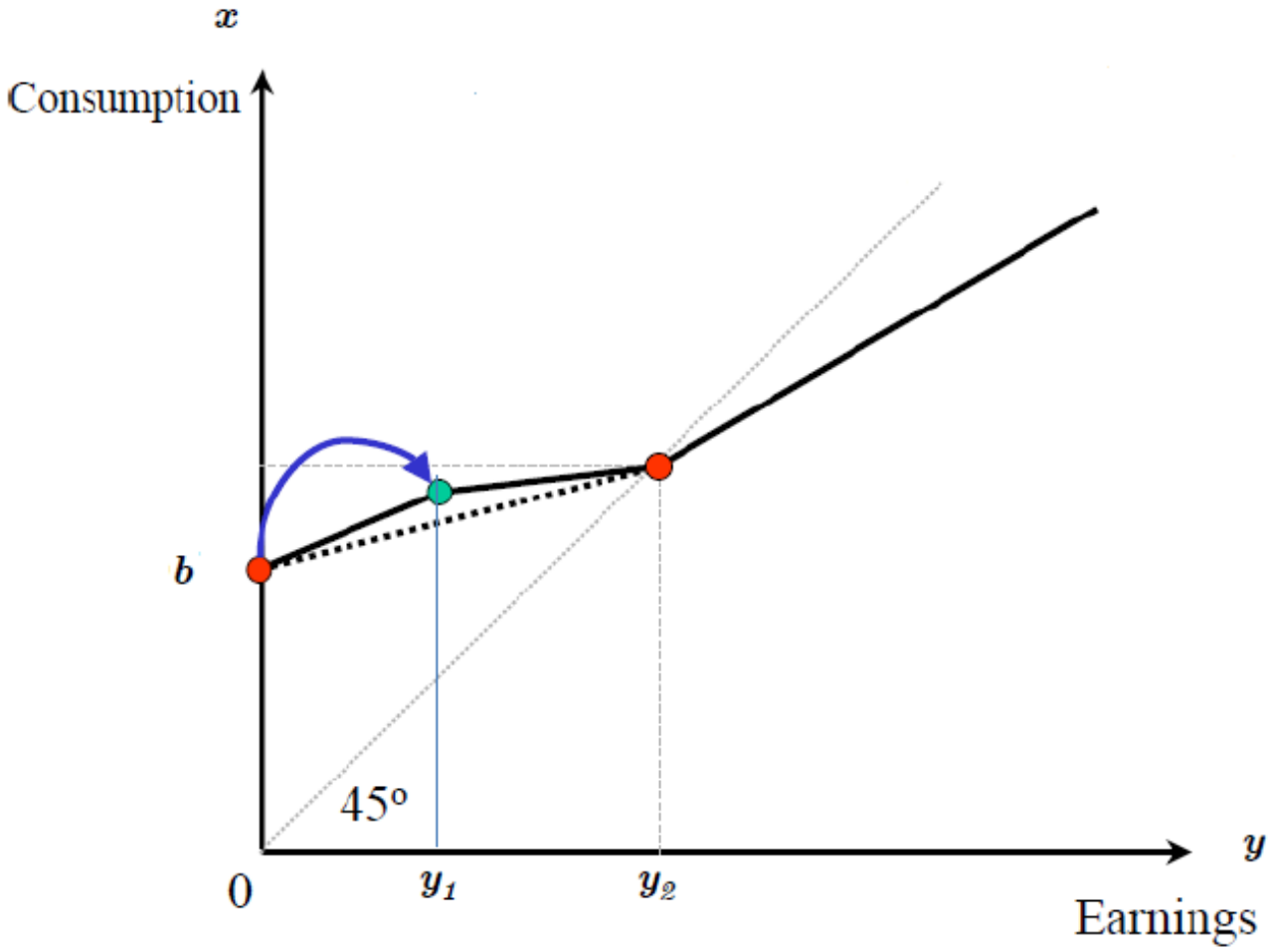


Introducing a "very small EITC" is desirable for tax revenue

Increasing transfer in occupation $i = 1$ is desirable for tax revenue (redistribution) if $g_1 > 1$:
Mechanical effects:

$$M_1 = (g_1 - 1)\, h_1\, dx_1 > 0 \quad \text{if } g_1 > 1$$

Participation responses save government revenue



Participation responses save government revenue

Behavioral responses i.e. participation responses (P) saves government revenue:

$$P_1 = \tau_1 h_1 dy_1 = \kappa_1 \frac{\tau_1}{1-\tau_1} h_1 dx_1 > 0$$

where we have used $dx_i = y_i - \tau_i y_i$ and $\kappa_i \equiv \frac{\partial h_i}{\partial x_i} \frac{x_i}{h_i}$.

Optimal participation tax

$$M_1 + P_1 = 0$$

$$(g_1 - 1)\, h_1\, dx_1 + \kappa_1 \frac{\tau_1}{1 - \tau_1} h_1\, dx_1 = 0$$

$$\frac{\tau_1}{1 - \tau_1} = \frac{1 - g_1}{\kappa_1}$$

Optimal taxation when labor is along the extensive margin

- Optimal participation tax rate as:

$$\frac{\tau(y)}{1-\tau(y)} = \frac{1-g(y)}{\kappa(y)}$$

- see Diamond (JPubE 1980) or Saez (QJE 2002, Prop.1)
- This formula is a simple inverse elasticity tax rule for the participation tax rate on work: The participation tax rate decreases with the elasticity η and with $g(y)$, the social value of marginal consumption for individual earnings y.

Optimal taxation when labor is along the extensive margin

The intuition for this result can be understood as follows.
Starting from a tax and benefit system with a **positive** participation tax rate for low-skilled workers, and suppose the government contemplates strengthening work incentives for low-skilled individuals by **reducing the taxes that they would pay when working**. This has the following effects:

- The cut in taxes means that **tax revenues fall**, and this is a cost.

- Depending on the $g(y)$ associated to the low-skilled workers, the associated i**ncrease in income of these workers can be viewed positively by the government.** For instance, when $g(y) > 1$ for the low-paid workers, the government would prefer these individuals to have income more than the average individual and this benefit effect has to outweigh the costs of reduced revenue.
- The behavioural response from cutting the participation tax rate is to **induce some low-skilled individuals to start working**, and this increases government revenue (because the individuals who move into work pay positive net taxes).

Optimal taxation when labor is along the extensive margin

In-work subsidies $\tau_i < 0$ (such as EITC) become optimal when labor supply responses are concentrated along extensive margin and social marginal welfare weight on low skilled workers > 1 (Saez 2002).

Optimal taxation when labor is along the extensive margin

$$\frac{\tau(y)}{1-\tau(y)} = \frac{1-g(y)}{\eta}$$

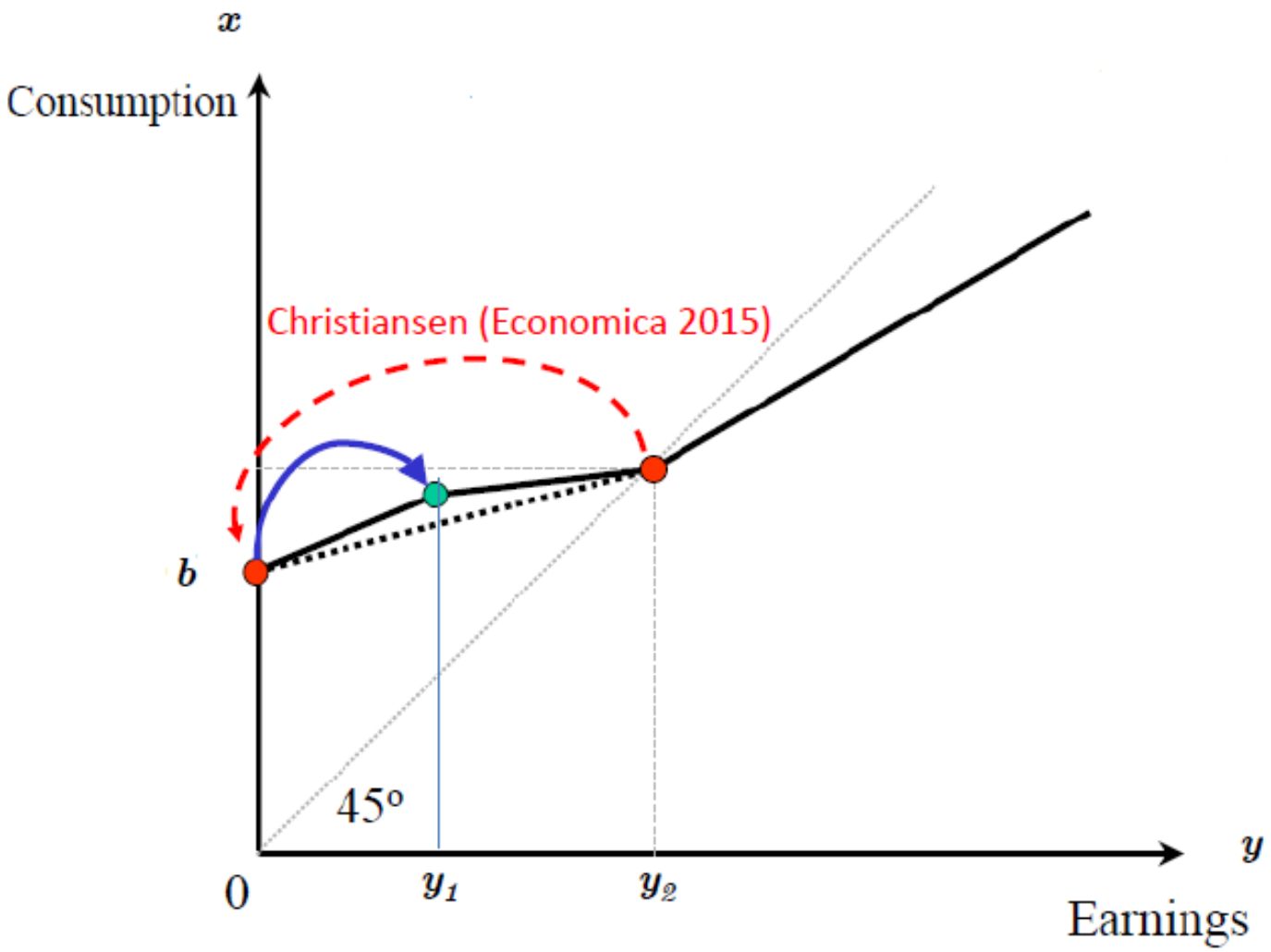
- If the government values redistribution so that $1 - g(y) < 0$, then the participation tax rate should be **negative**—in other words, low income workers should receive an earnings subsidy; **EITC**.
- In sharp contrast to the intensive model, the extensive model implies that earnings subsidies or work-contingent credits (such as the earned income tax credit or the working tax credit) should be part of an optimal tax system.

More intuition from Christiansen (Economica 2015)

- Extensive margin model
- Utilitarian criterion
- Keep b low as otherwise too many skilled people would be induced to opt for b and stop participating $\Rightarrow$ EITC prevail.
- Whether EITC prevails depends on labor supply responses of agents in the income brackets beyond that of the working poor.
- "The main concern is to discourage the better off from not working rather than subsidizing the poor to encourage their labour market participation" (Christiansen 2015, p.595).

Christiansen (Economica 2015)

Key: Increasing x_1 rather than b avoids that type 2-workers stop working.



Utilitarian criterion and extensive margin

We have seen that $\tau(y) < 0$ can prevail at the bottom.
The negative participation tax makes it possible to redistribute to the working poors without distorting the labor supply of higher productivity individuals.

Remark: The negative participation tax induces **some poor individuals to work**: This is a **distortion**!
In a first-best environment, the planner would give resources to the poor without trying to raise their participation

Maximin and extensive margin only

Maximin: Always positive participation taxes, $\tau(y) > 0$, Laroque (Ecta 2005)

Intuition: The government only cares about the non-working people $\Rightarrow$ Their well-being cannot be reduced compared to the one of working-poor, therefore $\tau(y) < 0$ never used

Outline

Income taxation creates 2 types of distortion on the individual labor supply

- Responses along the **extensive** margin, i.e. to work ($Y > 0$) or not ($Y = 0$)
relevant: **participation tax rates:** $\frac{T(Y)+b}{Y}$
where $b \geq 0$: welfare benefit for the non-employed.

- Responses along the **intensive** margin, i.e. to choose "continuously" your labor earnings Y (your labor hours or effort)
relevant: **marginal tax rates:** $T'(Y)$

Empirical evidence emphasizes the need to include both margins in the optimal tax literature

- Both labor supply margins empirically matter.

See the large literature in microeconometrics, e.g., Heckman (AER 1993), Blundell and MaCurdy (1999), Blundell, Bozio and Laroque (AER PP 2011)

- Saez (2010 AEJ: Econ.Pol.):
  - Intensive response implies bunching at budget kinks of the individual budget constraint (under piecewise linear taxation).
  - Saez finds limited bunching at kink points except for self-employed $\Rightarrow$ extensive margin relevant for employed workers.
- Empirical literature shows that participation labor supply responses (due to fixed costs of working) are large at the bottom, much larger and clearer than intensive responses.

Optimal income taxation with both intensive and extensive margins

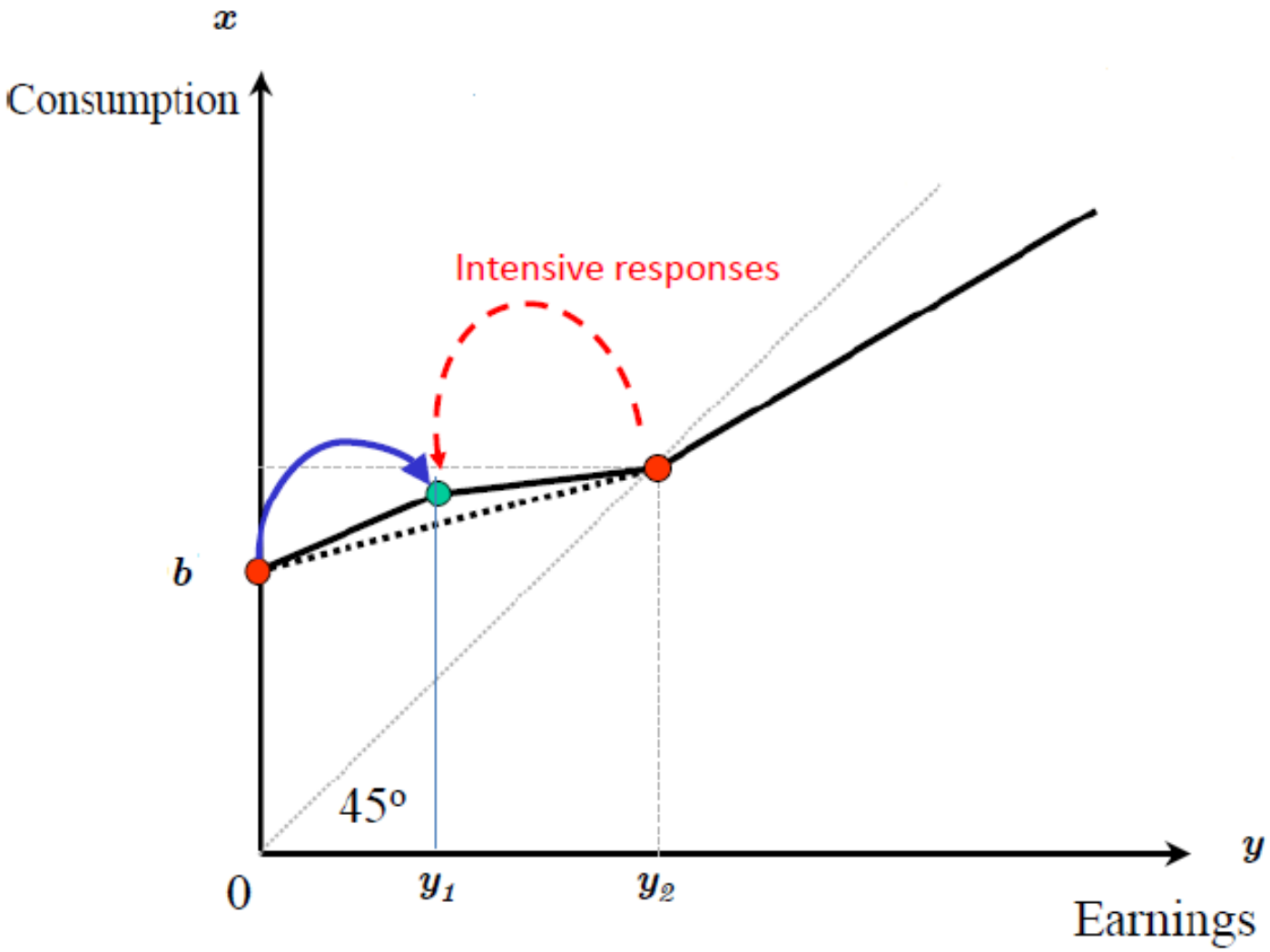
- Saez (QJE 2002), participation and occupational choice. Participation can be in job for own skill or next lower skill.
- Jacquet, Lehmann and Van der Linden (JET 2013) analytically derive the signs of the **marginal** tax rates in a model with intensive and extensive margins simultaneously.

With both margins, win-win reform if intensive responses are small

With both intensive and extensive margins, if intensive response is small, the previous result regarding negative participation tax rates is still valid (Saez QJE 2002).

Again, start from a tax and benefit system with a **positive** participation tax rate for low-skilled workers, and suppose the government reduces **the taxes that low-paid workers would pay when working**. See next figure.

With both margins, win-win reform if intensive responses are small



Saez (2002) numerically show that with both margins, negative participation tax rates can prevail at the bottom of the distribution.

What we already know

- Model with intensive margin only and continuum of earnings (Mirrlees 1971): **Positive** marginal tax rates $T'(Y) \geq 0 \quad \forall Y$

With negative T', people are working too much (due to substitution effects) $\Rightarrow$ Increase T' improves efficiency and also equity by distributing the proceeds lump sum

- Model with extensive margin only and continuum of earnings (Diamond 1980) or with both margins and a finite number of earnings (using simulations, Saez 2002):

Negative participation tax rates at the bottom $\frac{T(Y)+b}{Y} < 0$ may prevail, i.e. EITC

Redistribute more money to low-paid workers when they have a relatively large social welfare weight

Both intensive and extensive margins

- Fixed wage rates $\equiv$ skills: w

  on $[w_0, w_1]$ with $0 \leq w_0 < w_1 \leq +\infty$.
- Fixed disutilities of working (e.g., commuting, job search, reduced amount of time for home production) net of the stigma of being non-employed: χ

  on $(-\infty, \chi^{\max}]$ with $\chi^{\max} \leq +\infty$
- Joint density of types (χ, w): $k(\chi, w)$ continuous and positive.

Agents differ along two dimensions and preferences over consumption and effort (or earnings) are heterogeneous

- Preferences over earnings, $Y > 0$, and consumption, $C = Y - T(Y)$:

$$\mathcal{U}(C, Y, w) - \mathbb{I}_{Y>0} \cdot \chi \qquad \text{with } \mathcal{U}'_C > 0 > \mathcal{U}'_Y$$

This utility allows **preferences over consumption and earnings to vary with skill** w

- Single-Crossing condition : $w \mapsto \underbrace{-\frac{\mathcal{U}'_Y}{\mathcal{U}'_c}(C, Y, w)}_{\text{MRS}_{CY}}$ is decreasing

Individuals' choices

- Intensive choice: Employed workers choose different (positive) earnings Y because of skill heterogeneity:

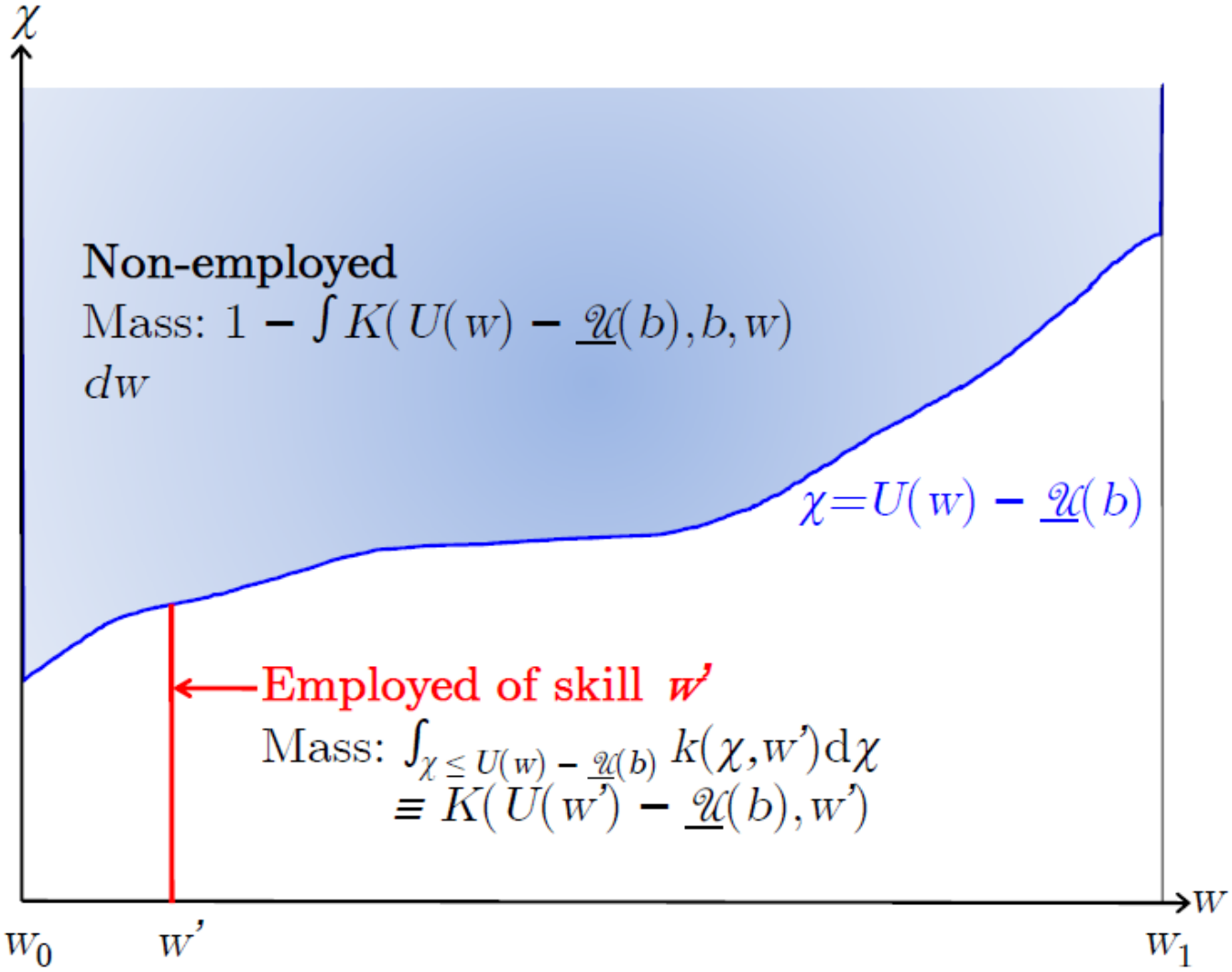
$$U(w) \equiv \max_{Y} \quad \mathcal{U}(Y - T(Y), Y, w) \tag{12}$$

- Extensive choice: Individuals of the same w take different participation decisions because they have different disutilities of participation χ:

  An individual (w, χ) chooses to work iff:

$$\begin{aligned} U(w) - \chi &\geq \mathcal{U}(b, 0, w) \equiv \underline{\mathcal{U}}(b) \\ \text{where } b &: \quad \text{(endogenous) welfare benefit} \end{aligned}$$

At each skill level, we then have employed and non-employed people



Mass of workers of skill w given by:

$$K\left(U\left(w\right)-\underline{\mathcal{U}}\left(b\right)\right)\equiv\int\limits_{U(w)-\underline{\mathcal{U}}(b)}k\left(\chi,w\right)d\chi \tag{13}$$

The Social Preferences generalize Bergson-Samuelson SWF

Sum a transformation $G(U, w, \chi)$ of individuals' utility U over all individuals:

$$\begin{aligned}\Omega &= \int_{w_0}^{w_1} \left\{ \int_{\chi \leq U(w) - \underline{\mathcal{U}}(b)} G(U(w) - \chi, w, \chi) \cdot k(\chi, w)\, d\chi \right. \\ &\quad \left. + \int_{\chi \geq U(w) - \underline{\mathcal{U}}(b)} G(\underline{\mathcal{U}}(b), w, \chi) \cdot k(\chi, w)\, d\chi \right\} dw\end{aligned}$$

$G'_U > 0$ and either $G''_{UU} < 0$ or $G''_{Uw} < 0$ or both.

G depends on utility levels as with Bergson-Samuelson SWF but also on the (w, χ)-type.
Relevant when the gvt wants to **compensate only for some characteristics (e.g. skill levels)**

Government's optimization problem

Standard solving in terms of **incentive compatible allocations**:

Max. Welfarist Criterion s.to:

- Government budget constraint
- Incentive-Compatibility Constraints

The optimal taxation problem

Revelation principle $\Rightarrow$ The planner's problem is to find the best allocation implementable by a direct truthful/revealing mechanism.

In a direct mechanism, an individual who claims to be of type (w, χ) obtains the allocation

$$\{C(w,\chi), Y(w,\chi)\}.$$

Incentive compatibility constraints among working individuals of productivity w imply that (with some abuse of notations):

$$\begin{aligned} U(C(w,\chi), Y(w,\chi), w) - \chi &\geq U(C(w,\chi'), Y(w,\chi'), w) - \chi \\ U(C(w,\chi'), Y(w,\chi'), w) - \chi' &\geq U(C(w,\chi), Y(w,\chi), w) - \chi' \end{aligned}$$

The optimal taxation problem

Thus

$$\mathcal{U}\left(C\left(w,\chi'\right),Y\left(w,\chi'\right),w\right)=\mathcal{U}\left(C\left(w,\chi\right),Y\left(w,\chi\right),w\right)$$

All participating workers of productivity w must have the same utility ignoring the fixed cost.

This level utility should be provided at minimum cost, which implies:

- $C\left(w,\chi\right)=C\left(w,\chi'\right)=C\left(w\right)$
- $Y\left(w,\chi\right)=Y\left(w,\chi'\right)=Y\left(w\right)$

The optimal taxation problem

Incentive compatibility constraint among non-working individuals imply that:

$$\begin{aligned} \mathcal{U}\left(C\left(w,\chi\right),0,w\right) &\geq \mathcal{U}\left(C\left(w',\chi'\right),0,w\right) \\ \mathcal{U}\left(C\left(w',\chi'\right),0,w'\right) &\geq u\left(C\left(w,\chi\right),0,w'\right) \end{aligned}$$

Thus, all non-working individuals must have the same level of consumption:

$$C\left(w,\chi\right) = C\left(w',\chi'\right) = b$$

Thus, without loss of generality, we can consider that the goal of the optimal taxation problem is to find the optimal allocation of resources:

$$\left\{C\left(w\right), Y\left(w\right), b\right\}$$

The optimal taxation

- The government' s problem consists in finding the optimal tax schedule and welfare benefit to maximize the social objective, subject to the budget constraint and the labor supply decisions along the intensive (12) and extensive (13) margins.
- According to the **taxation principle**, the set of allocations induced by an income tax function and a welfare benefit corresponds to the set of allocations that verify (13) and incentive compatibility constraints:

$$\forall (w, x) \in [w_0, w_1]^2$$
$$U(w) = \mathcal{U}(C(w), Y(w), w) \geq \mathcal{U}(C(x), Y(x), w)$$

- Since the strict single-crossing condition holds, the above ICC are thus equivalent to imposing the monotonicity constraint that earnings Y are non-decreasing in skill as well as the envelope condition on (12):

$$U'(w) = \mathcal{U}'_w(C(w), Y(w), w) > 0$$

i.e. first-order ICC.

Deriving the optimal tax formulae with Hamiltonian

Hamiltonian (note: with dual problem, see Boadway and Jacquet JET 2008):

$$\underbrace{[Y - C(U, Y, w) + b] \cdot K(U - \underline{\mathcal{U}}(b), w)}_{\text{Tax revenue}} + q \cdot \underbrace{\mathcal{U}'_w(C(U, Y, w), Y, w)}_{\text{ICC}}$$

$$\underbrace{+\frac{1}{\lambda}\left\{\int_{-\infty}^{U-\underline{\mathcal{U}}(b)} G(U - \chi, w, \chi) \cdot k(\chi, w) \cdot d\chi + \int_{U-\underline{\mathcal{U}}(b)}^{\chi^{\max}} G(\mathcal{U}(b), w, \chi) \cdot k(\chi, w) \cdot d\chi\right\}}_{\text{Social Welfare function}}$$

The optimal tax formulae

For the F.O.C.'s and the details of the derivation of the structural optimal tax formulae, see Jacquet, Lehmann and Van der Linden (JET 2013).

$$\mathcal{H}'_Y = \left(1 + \frac{\mathcal{U}'_Y}{\mathcal{U}'_C}\right) \cdot K\big(U - \underline{\mathcal{U}}(b), w\big) + q \cdot \mathcal{X}'_Y$$

$$\begin{aligned}
\mathcal{H}'_U &= \big\{Y - \mathcal{C}(U, Y, w) + b\big\} \cdot k\big(U - \underline{\mathcal{U}}(b), w\big) \\
&\quad + \frac{(g(w) - 1) \cdot K(U - \underline{\mathcal{U}}(b), w)}{\mathcal{U}'_C(C(w), Y(w), w)} + q \cdot \mathcal{X}'_U \\
&= \left\{Y - \mathcal{C}(U, Y, w) + b - \frac{1 - g(w)}{\kappa(w)}\right\} \cdot k\big(U - \underline{\mathcal{U}}(b), w\big) + q \cdot \mathcal{X}'_U
\end{aligned}$$

$$\begin{aligned}
\mathcal{H}'_b &= K\big(U - \underline{\mathcal{U}}(b), w\big) - \big(Y - \mathcal{C}(U, Y, w) + b\big) \cdot \underline{\mathcal{U}}'_b(b) \cdot k\big(U - \underline{\mathcal{U}}(b), w\big) \\
&\quad + \int_{U - \underline{\mathcal{U}}(b)}^{\chi^{\max}} G'_U\big(\underline{\mathcal{U}}(b), w, \chi\big) \cdot \underline{\mathcal{U}}'_b(b) \cdot k(\chi, w) \cdot d\chi
\end{aligned}$$

An heuristic interpretation of the FOC thanks to an infinitesimal tax perturbation method

- Define the **behavioral elasticities along the nonlinear** tax schedule hence they incorporate the "circular process".
- A **tax perturbation method** along the **nonlinear** tax schedule ($\neq$ along a linearized tax schedule).
- $\Rightarrow$ Further assumptions: The optimal allocation is such that Y is **differentiable everywhere** in skill, there is **no bunching** and the tax function T is **twice differentiable everywhere** in earnings.
- Method: **Uniform decrease of marginal tax rates over a small range of earnings**:
- The sum of these effects $= 0$ since we are at the optimum. This gives the optimal tax formula as a function of the social weights and behavioral responses.

Behavioral elasticities and income effects

Similar derivation and definitions to the ones used above, in the model with intensive margin only. Details in Jacquet, Lehmann and Van der Linden (2013).
Remark: Elasticities **account for the nonlinearity of the income tax schedule.**

- Elasticity of y with respect to $\tau \equiv 1 - T'(y)$:

$$\varepsilon(w) \equiv \frac{1 - T'(Y)}{Y} \frac{\partial Y}{\partial \tau}$$

$$= \frac{\frac{\mathcal{U}'_Y}{Y(w)}}{\left[\mathcal{U}''_{YY} - 2\left(\frac{\mathcal{U}'_y}{\mathcal{U}'_c}\right)\mathcal{U}''_{CY} + \left(\frac{\mathcal{U}'_y}{\mathcal{U}'_c}\right)^2 \mathcal{U}''_{CC} - T''(y)\mathcal{U}'_C\right] \cdot \mathcal{U}'_C} > 0$$

Behavioral elasticities and income effects

- Response of before-tax income to a lump-sum increase in the level of taxation (level of taxation being denoted by ρ)

$$\eta(w) \equiv \frac{\partial y(w)}{\partial \rho}$$

$$\frac{\left(\frac{\mathcal{U}'_y}{\mathcal{U}'_c}\right)\mathcal{U}''_{CC} - \mathcal{U}''_{CY}}{\left[\mathcal{U}''_{YY} - 2\left(\frac{\mathcal{U}'_y}{\mathcal{U}'_c}\right)\mathcal{U}''_{CY} + \left(\frac{\mathcal{U}'_y}{\mathcal{U}'_c}\right)^2\mathcal{U}''_{CC} - T''(y)\mathcal{U}'_C\right]\cdot\mathcal{U}'_C}$$

which is < 0 if leisure is a normal good.

Behavioral elasticities and income effects

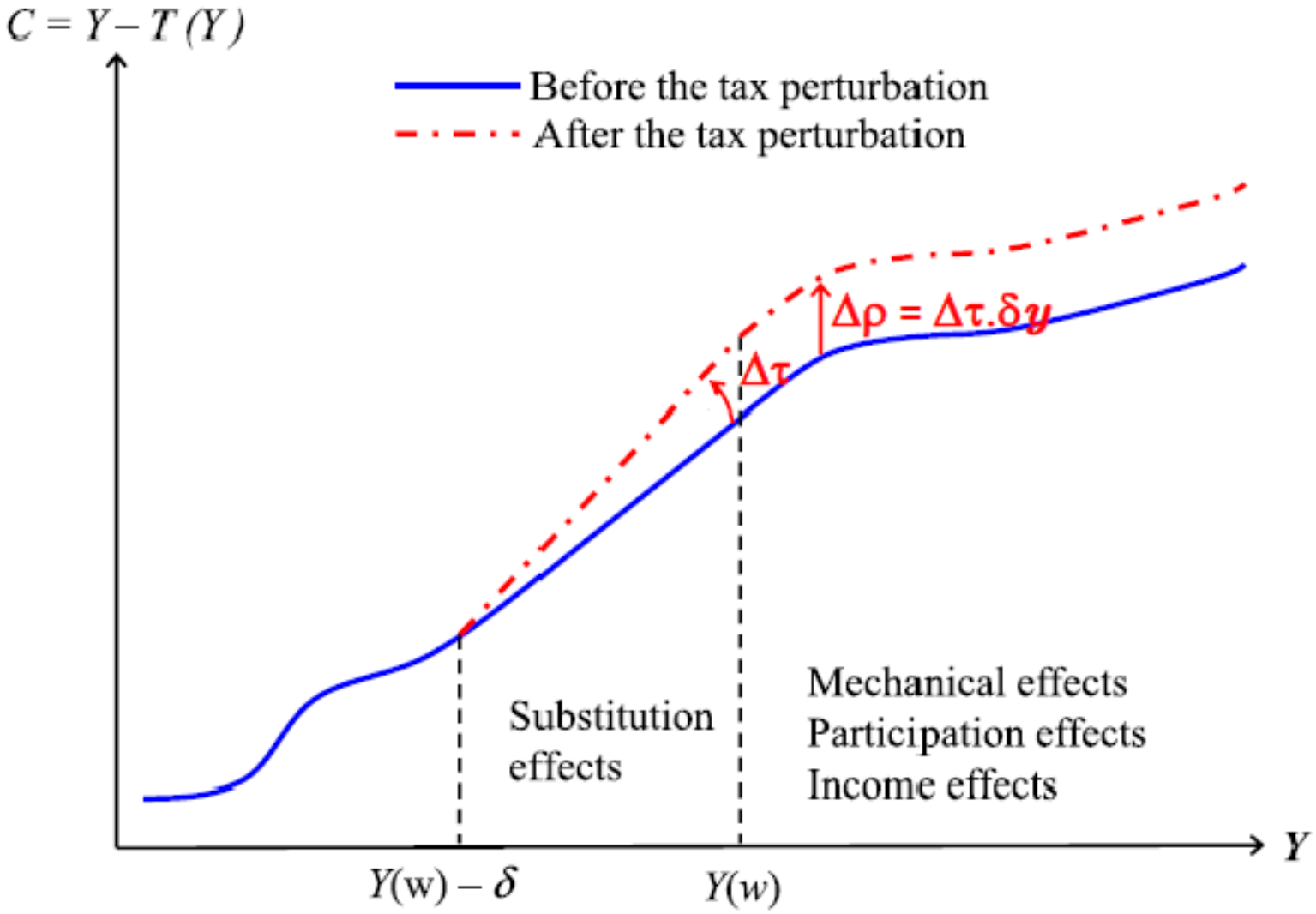
- Elasticity of gross earning y with respect to w:

$$\alpha(w) \equiv \frac{w}{Y(w)} \frac{\partial Y(w)}{\partial w}$$

$$= \frac{-\frac{w}{Y(w)}\left[\mathcal{U}''_{Yw} \cdot \mathcal{U}'_C - \mathcal{U}''_{Cw} \cdot \mathcal{U}'_Y\right]}{\left[\mathcal{U}''_{YY} - 2\left(\frac{\mathcal{U}'_y}{\mathcal{U}'_c}\right)\mathcal{U}''_{CY} + \left(\frac{\mathcal{U}'_y}{\mathcal{U}'_c}\right)^2 \mathcal{U}''_{CC} - T''(y)\mathcal{U}'_C\right] \cdot \mathcal{U}'_C} > 0$$

Tax perturbation method to obtain optimal tax formula

Consider a uniform decrease of $T'(Y(w))$ over $[Y(w) - \delta Y, Y(w)]$.
$\Rightarrow$ Tax function unchanged for earnings below $Y(w) - \delta y$
$\Rightarrow$ Tax function uniformly decreased by an amount $\Delta\rho = \Delta\tau \times \delta Y$ for earnings above $Y(w)$.


$C = Y - T(Y)$
Before the tax perturbation
After the tax perturbation
$\Delta\rho = \Delta\tau.\delta y$
$\Delta\tau$
Substitution effects
Mechanical effects
Participation effects
Income effects
Y
$Y(w) - \delta$
$Y(w)$

Tax perturbation method

To apply this method, we need **further regularity assumptions:**

The optimal allocation is such that Y is **differentiable everywhere** in skill, there is **no bunching** and the tax function T is **twice differentiable everywhere** in earnings.

Tax perturbation method

Let $h(w) \equiv K\left(U\left(w\right) - \underline{\mathcal{U}}\left(b\right), w\right)$: skill density **among the employed** population.

Let $H(w) \equiv \int_{w_0}^{w_1} h(x)dx$ the corresponding CDF.

$H(w_1)$: total mass of employed individuals, in general < 1.

φ : income density and $\Phi\left(.\right)$ corresponding CDF. Differentiating both sides of the equality $\Phi\left(Y\left(w\right)\right) = H\left(w\right)$ and using the definition of $\alpha\left(w\right)$, the two densities are linked by:

$$Y\left(w\right)\varphi\left(Y\left(w\right)\right) = \frac{wh\left(w\right)}{\alpha\left(w\right)}$$

Mechanical effects

Absent any behavioral response (, participation and income effects), for each taxpayer of skill n above w:

- the government receives $\Delta\rho$ Euros of tax less from each of the $h(n)$ workers of skill n.
- welfare of these individuals increases (thanks to the increase in consumption), which is valued $g(n)$ by the government.

$\Rightarrow$ Total mechanical effect for all tax payers with skills above w equals

$$M(w) = -\Delta\rho \int_{w}^{w_1} \{1 - g(n)\} h(n)\, dn$$

Income effects

Workers of skill n above w change their intensive labor supply decisions because of the income effect. Their earnings change by $\Delta y(n) = \eta(n)\Delta\rho$ which induces a change in tax revenue of $T'(y(n))\eta(n)\rho$. This behavioral response has only a second-order effect on the social objective.
$\Rightarrow$ Total income effect for all tax payers with skills above w equals

$$I(w) = \Delta\rho \int_{w}^{\overline{w}} T'(y(n))\eta(n) f(n)\, dn$$

Substitution effect

Workers whose y before the reform lie in the interval $[Y(w) - \delta Y, Y(w)]$ have their marginal tax rates that uniformly decrease by $\Delta\tau$. This induces a substitution effect that lead to a rise in earnings, for the $\varphi(Y(w))$ affected workers:

$$\Delta Y(w) = \frac{\varepsilon(w) Y(w)}{1 - T'(Y(w))} \Delta\tau$$

Hence each of them generates $T'(Y(w)) \Delta Y(w)$ additional tax to the government. From $\Delta\rho = \Delta\tau \cdot \delta Y$, the total substitution effect equals:

$$S(w) = \frac{T'(Y(w))}{1 - T'(Y(w))} \frac{\varepsilon(w)}{\alpha(w)} wh(w) \Delta\rho$$

Participation effect

The reduction in tax levels induces $\kappa(n)h(n)\Delta\rho$ individuals of skill n to enter employment. Each additional worker of skill n pays $T\left(Y\left(n\right)\right)$ taxes, and the government saves the welfare benefit b, so the participation effect at skill n is $\kappa(n)\left(T\left(Y\left(n\right)\right)+b\right)\Delta\rho$. The participation effects for all skills n above w are:

$$P\left(w\right)=\Delta\rho\int_{w}^{\overline{w}}\kappa(n)\left(T\left(Y\left(n\right)\right)+b\right)h\left(n\right)dn\Delta\rho$$

Starting from the optimal tax schedule, a tax perturbation should have no first-order effect. Therefore $M(w) + I(w) + S(w) + P(w) = 0$:

The optimal marginal tax formula at each skill level in terms of behavioral responses and social weights

$$\frac{T'(Y(w))}{1-T'(Y(w))} = \underbrace{\frac{\alpha(w)}{\varepsilon(w)}}_{\text{Efficiency}} \times \underbrace{\frac{H(w_1)-H(w)}{w \cdot h(w)}}_{\text{Skill distribution among employed}} \times$$

$\alpha(w)$: Elasticity of earnings Y w.r.to the skill level w

$\varepsilon(w)$ (>0): Compensated elasticity of earnings w.r.to $1-T'(Y)$

$$\frac{1}{H(w_1)-H(w)} \int_w^{w_1} \underbrace{\{1-g(n)}_{\text{Mechan.}} - \underbrace{\kappa(n)(T(Y(n))+b)}_{\substack{\text{Partic. effects} \\ \kappa(n)\equiv\frac{1}{h(w)}\frac{\partial h(w)}{\partial \rho}>0}} - \underbrace{\eta(n)\,T'(Y(n))\}}_{\substack{\text{Income effects} \\ \eta(n)\equiv\frac{\partial Y}{\partial \rho}<0 \\ \downarrow |\eta(n)| \Rightarrow \text{lower income effects}}} h(n)\,dn$$

$$g(w) \equiv \mathbb{E}\left[\frac{G'_v(v,..)\cdot \mathcal{U}'_C(C(w),..)}{\lambda} \,|\, w, \chi \leq U(w) - \mathcal{U}(b,0,w)\right]$$

Average marginal social weight associated with employed agents of skill w, expressed in terms of public funds:

Marginal tax rates at the bottom and at the top

From FOC's:

- If skill distribution bounded ($w_1 < \infty$) and no bunching at the top: $T'(Y(w_1)) = 0$
- If $w_0 > 0$ and no bunching at the bottom: $T'(Y(w_0)) = 0$

i.e. same results as in the model with intensive margin only, Sadka (ReStud 1976), Seade (JPubE 1977).

Marginal tax rates for interior skill levels

A sufficient condition for **positive marginal tax rates** ($\forall w \in (w_0, w_1)$) thanks to a new method to sign distortions along the intensive margin.

Method to sign distortions (valid in any model of monopoly screening with random participation):

- Characterize the optimum when the government observes the skills of workers but neither the skills of the non-employed nor the χ of anyone (first-and-a-half-best):

$$T\left(Y\left(w\right)\right)+b=\frac{1-g\left(w\right)}{\kappa\left(w\right)},$$

where $g\left(w\right)$: average mg social weight of w-workers,
$\kappa\left(w\right)$: participation response.

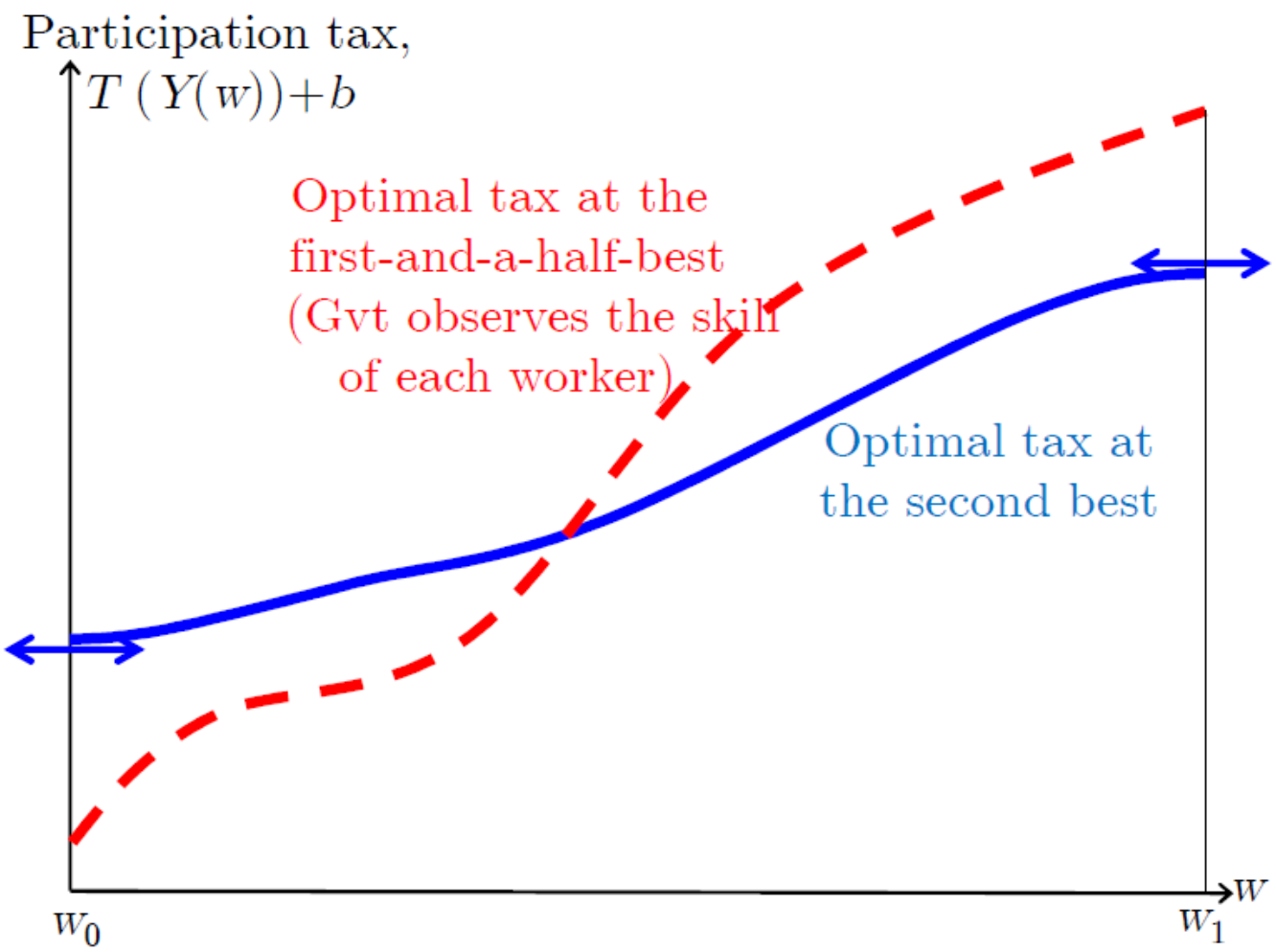
- Find a property on the first-and-a-half best optimum that allows to sign distortions along the intensive margin:

Property: $w \mapsto \frac{1-g(w)}{\kappa(w)}$ **admits a positive derivative everywhere** guarantees $T'(Y(w)) > 0$ in the first-and-a-half-best.

- It can be shown that this property is still valid when the government does not observe the skills and the χ (second-best).

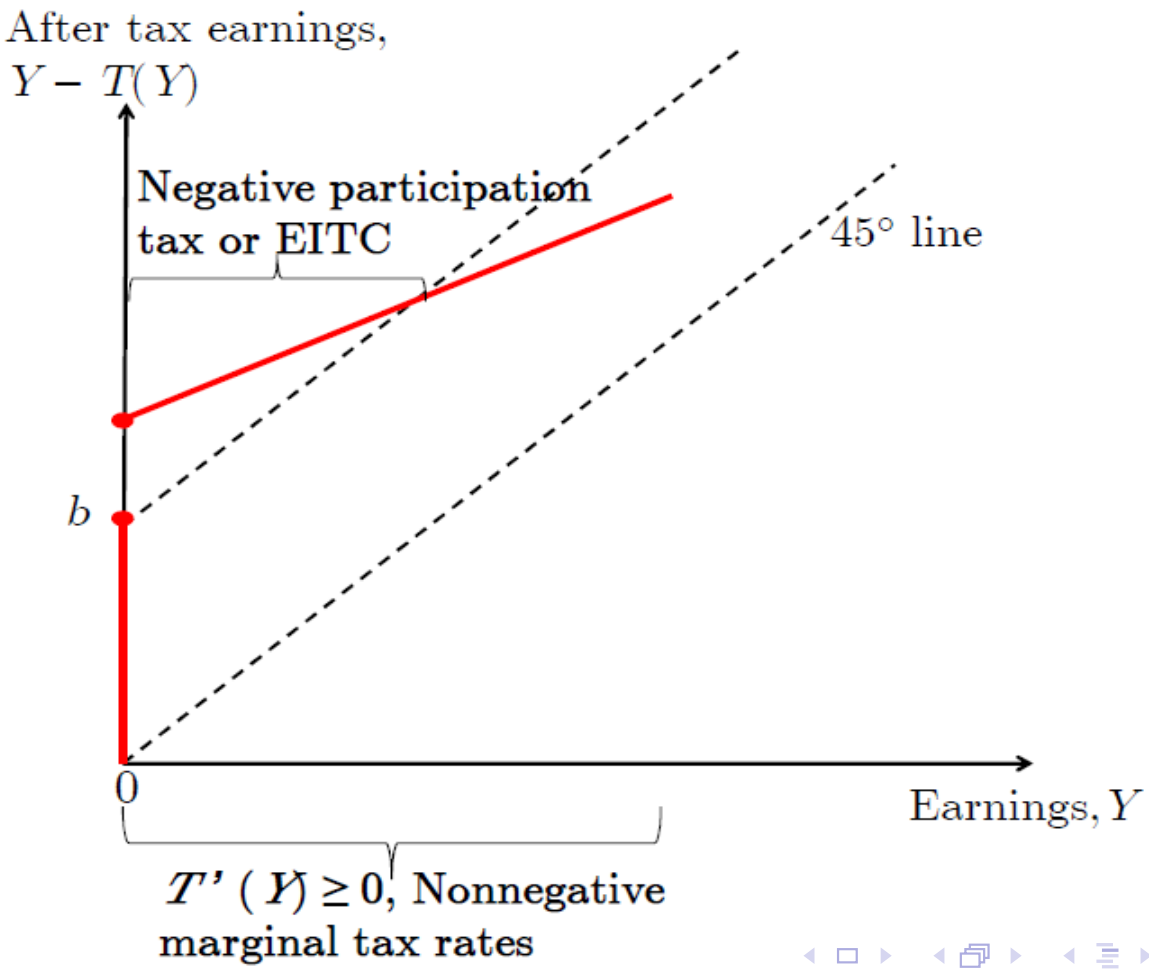
A sufficient condition for nonnegative mg tax rates:

Along the second-best allocation, $w \mapsto \frac{1-g(w)}{\kappa(w)}$ **admits everywhere a positive derivative is sufficient** for $T'(Y(w)) > 0$



Positive marginal tax rates does not preclude negative participation tax rates (discontinuity at the very bottom)

Using US data, Jacquet et al. (2013) show that negative participation tax rates (EITC) at the bottom and nonnegative marginal tax rates everywhere are optimal, under Benthamite preferences, i.e.



Examples on primitives where marginal tax rates are positive and participation tax rates are nonnegative (NIT)

Assume:

- Additive separable utility function

$$u(C) - v(Y, w) - \mathbb{I}_{Y>0} \cdot \chi$$

- If $w \mapsto k(\chi, w) / K(\chi, w)$ admits everywhere a negative partial derivative in χ and a non-positive partial derivative in w

- **Either** Maximin

  or Benthamite social preferences and $g(w_0) \leq 1$

  $\Rightarrow T'(Y(w)) > 0$ and NIT.

Check the empirical relevance of the condition using US data

- Social preferences: Bentham and Maximin
- Individuals' preferences:

$$\mathcal{U}(C, Y, w) = \frac{\left(C - \left(\frac{Y}{w}\right)^{1+\frac{1}{\varepsilon}} + 1\right)^{1-\sigma}}{1-\sigma}$$

  No income effect along the intensive margin

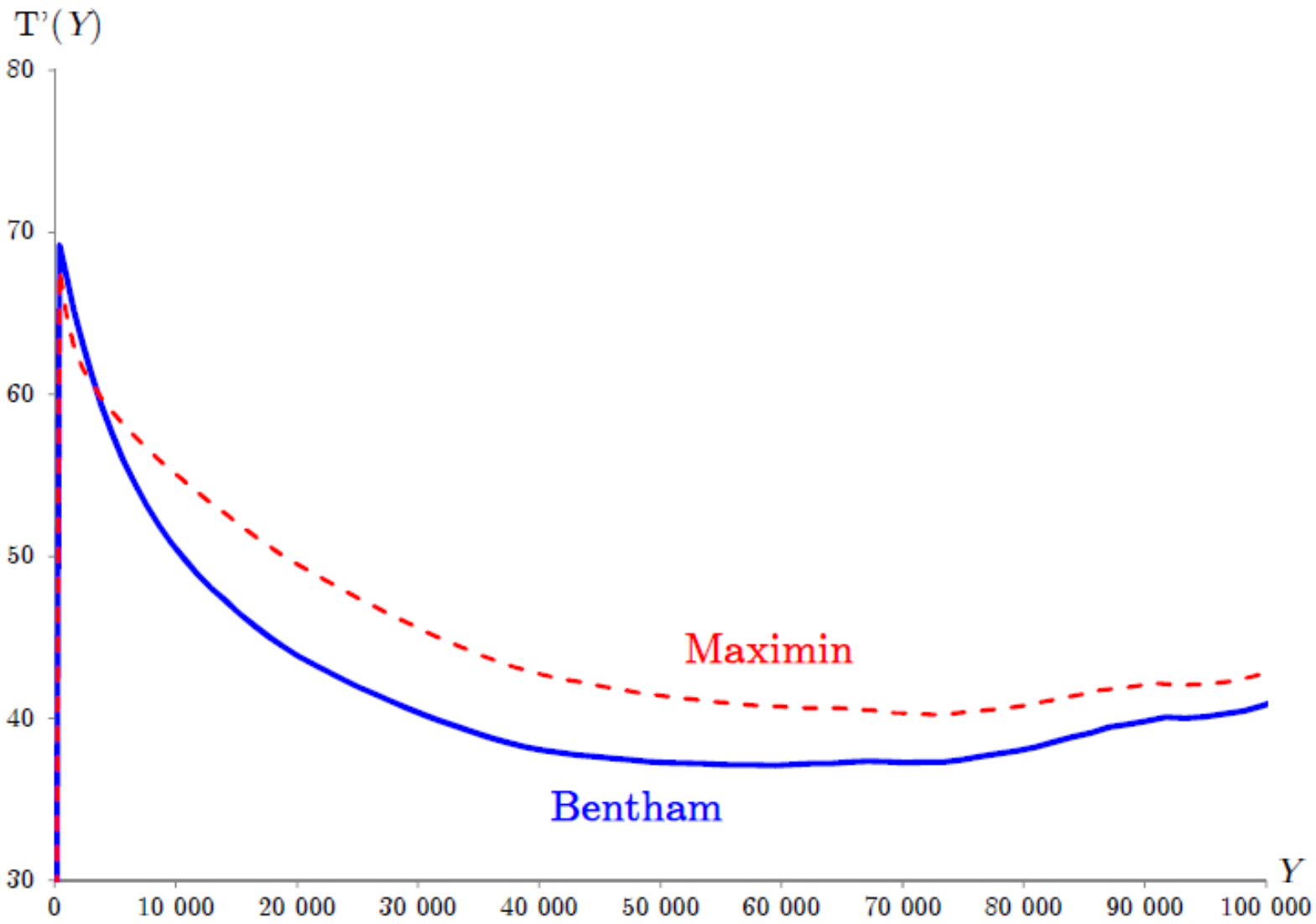
- $\varepsilon = 0.25$ in the benchmark (In the US, $\varepsilon \in [0.12, 0.4]$, Saez et al. JEL 2012)
- $\sigma = 0.8$ in the benchmark
- Weakly earnings in 2007 CPS for singles (without kids)

- The skill distribution among employed workers is calibrated such that the actual $T(.)$ yields **empirical earnings**
- The skill density is smooth, using a **quadratic kernel** (bandwidth \$3822)
- The **top** (3.3%) of the distribution is approximated by a **Pareto distribution** with Pareto index $a = 2$ following Diamond (AER 1998) and Saez (ReStud 2001)
- Remark: $w_0 = \$0.1$ (that corresponds to an annual $Y(w_0) < \$1$)
- CDF of χ, conditional on skill level is logistic:

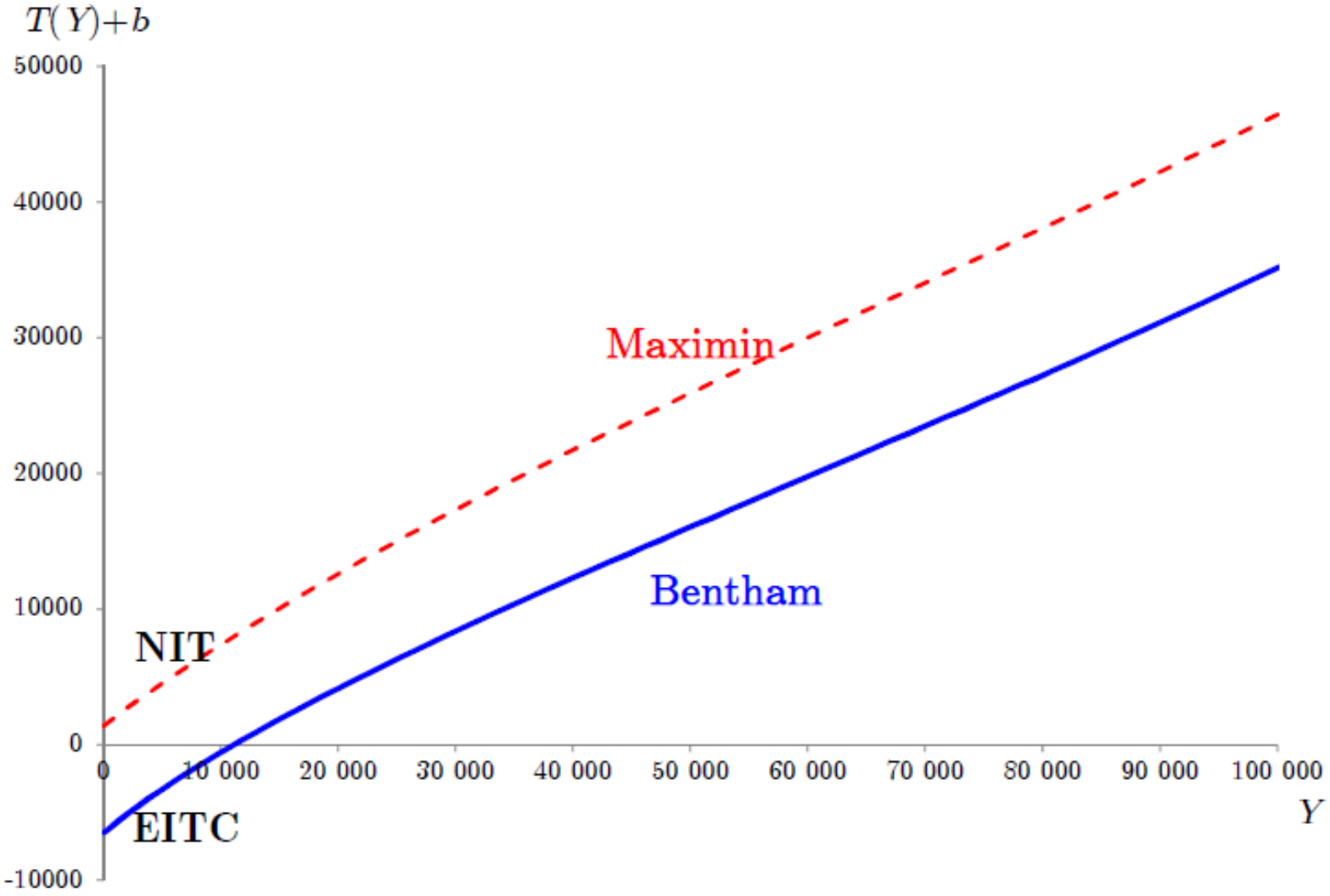
$$K(\chi, w) = \frac{\exp(-\alpha(w) + \beta(w)\ \chi)}{1 + \exp(-\alpha(w) + \beta(w)\ \chi)}$$

Parameters $\alpha(w)$ and $\beta(w)$ are calibrated to obtain **skill-specific** employment rates and **skill-specific** elasticities of employment rates w.r.to $C(w) - b$ that are empirically relevant.

Optimal U-shape profiles for the marginal tax rates

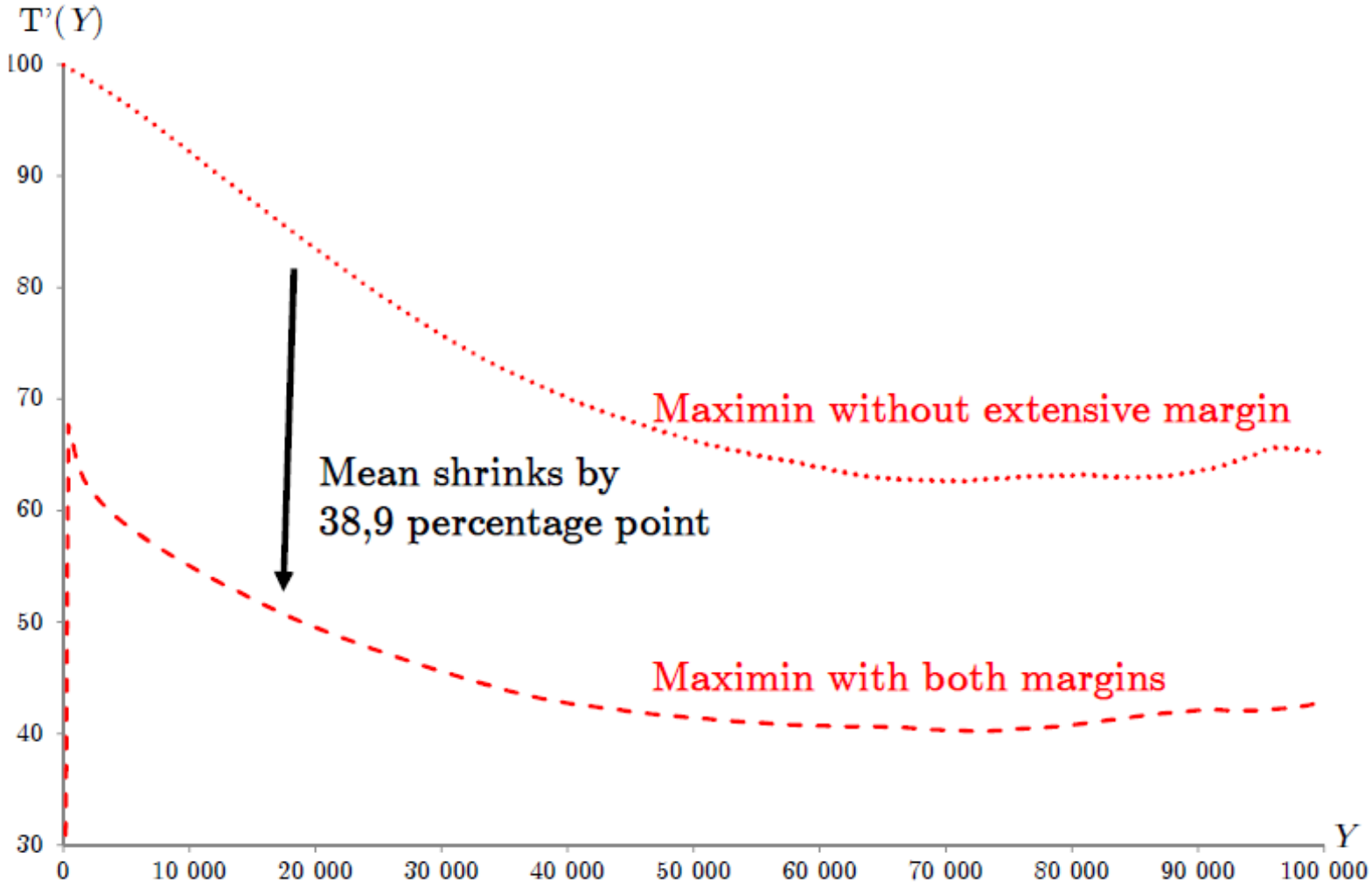


NIT under Maximin and EITC under Bentham



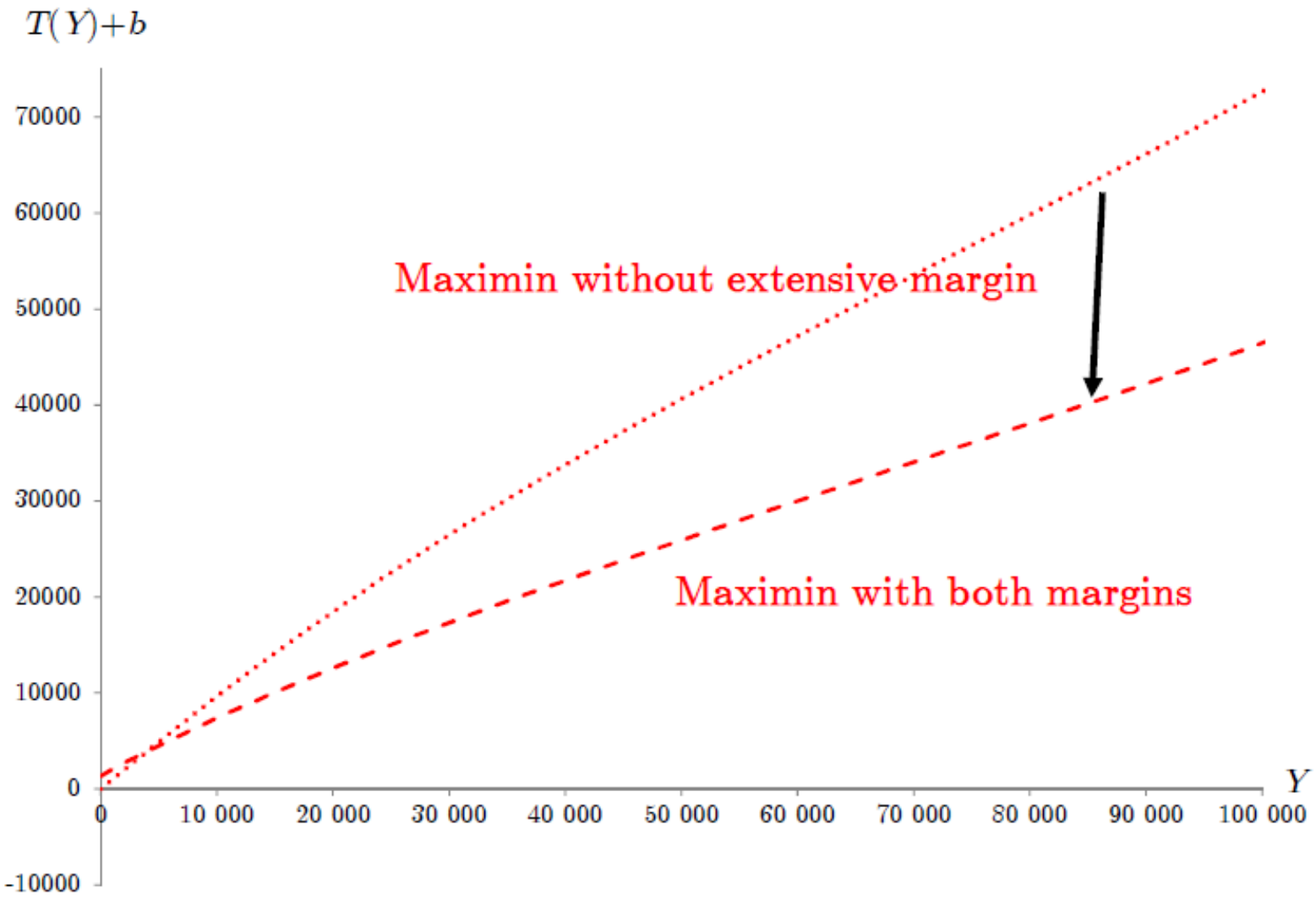
Introducing the extensive margin drastically reduces the optimal marginal tax rates

e.g., under Maximin:

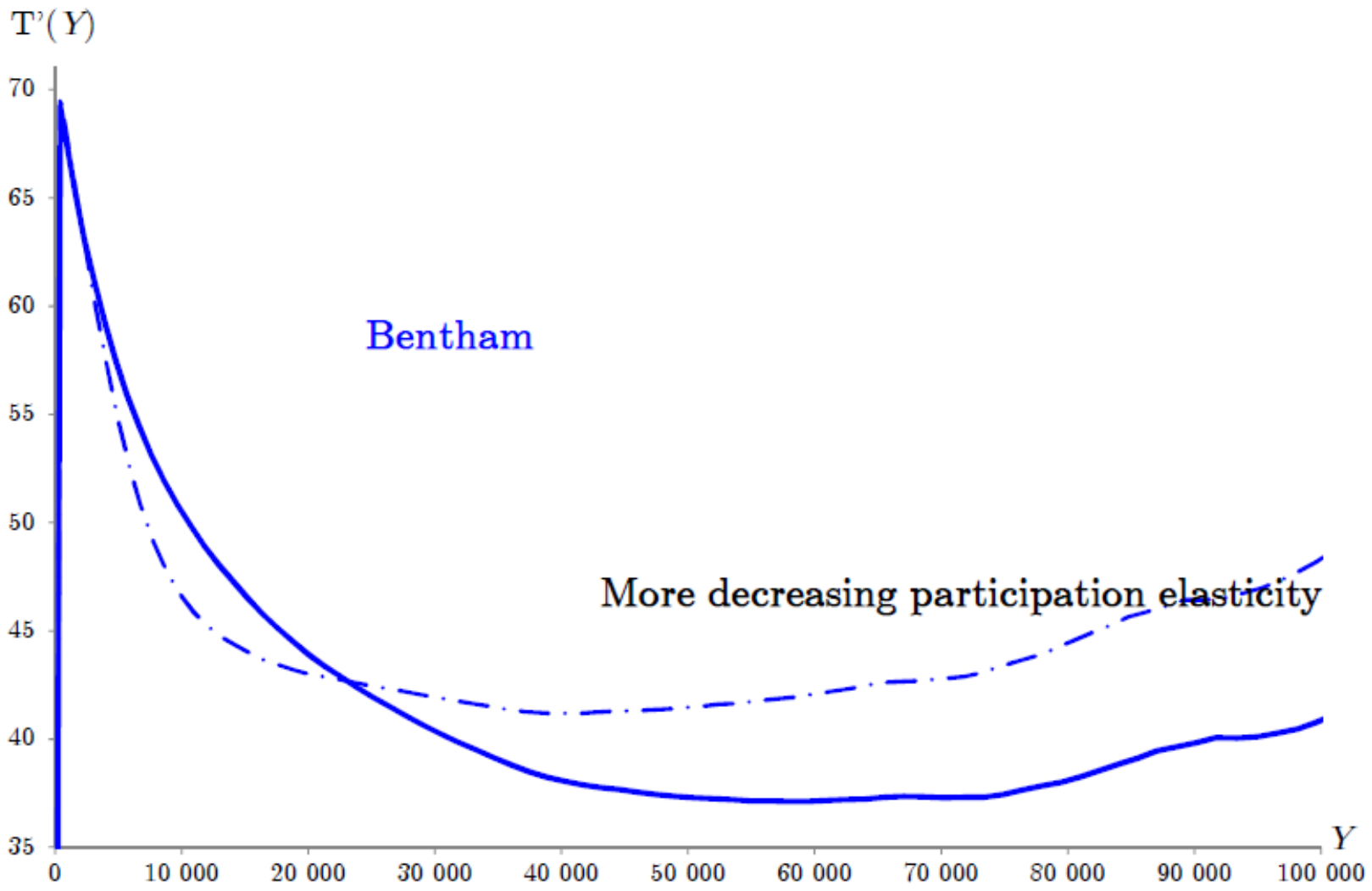


Introducing the extensive margin drastically modifies the optimal participation taxes

e.g., under Maximin:



A more decreasing participation elasticity shifts the marginal tax rates upwards



Sensitivity exercises

All sensitivity exercises show:

- Marginal tax rates always positive (and nil at the two extremes)
- The sufficient condition always holds
- Marginal tax rate are always U-shaped

Optimal income taxation with intensive and extensive margins

- Nonnegative marginal tax rates may coexist with negative participation tax rates at the bottom
- Under Bentham: EITC vs under Maximin: NIT

EITC in the US

The negative participation tax justifies the implementation in the U.S. of the Earned Income Tax Credit (EITC)

- The EITC is a subsidy that individuals can only get conditional on participating
- It is the largest cash anti-poverty program in the U.S. ($50billion allocated to 25 million families earning less than $40000)
- The EITC is large for families with children: It can raise their income by up to 40%

Optimal income tax schedule with intensive and extensive margin for the **Netherlands**: Zoutman, Jacobs and Jongen (2014)

Outline

All references are recommended but references preceded by asteriks are compulsory.

Atkinson, A., Piketty, T. and Saez, E. (2011), Top incomes in the long run of history, Journal of Economic Literature, 49:1, 3–71.

Blundell, R. Bozio, A. and Laroque, G. (2011), Labour supply responses and the extensive margin, American Economic Review, Papers and Proceedings, 101(3), 482–486.

Blundell, R. and T. MaCurdy, 1999, Labour Supply: A Review of Alternative Approaches, Handbook of Labor Economics.

Boadway, R. and L. Jacquet (2008), Optimal Marginal and Average Income Taxation under Maximin, Journal of Economic Theory, 143, 425–441.

Boadway, R., Cuff K. and Marchand, M. (2000), Optimal income taxation with quasi-linear preferences revisited, Journal of Public Economic Theory, 2, 435–460.

Choné, Ph., Laroque, G. (2011), Optimal taxation in the extensive model, Journal of Economic Theory, 146(2), 425–453.

Christiansen, V. (2015),.Optimal participation taxes, Economica, 82(328), 595–612.
Diamond, P. (1980), Income taxation with fixed hours of work, Journal of Public Economics, 13(1), 101–110.
**Diamond, P. (1998), Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates, American Economic Review, 88(1), 83–95.
Diamond, P. and Saez, E. (2011), From Basic Research to Policy Recommendations: The Case for a Progressive Tax", Journal of Economic Perspectives, 25(4), 2011, 165–190.
Ebert, U., 1992, A reexamination of the optimal nonlinear income tax, Journal of Public Economics, 49, 47–73.
Guesnerie, R. (1995), A Contribution to the Pure Theory of Taxation, Cambridge: Cambridge University Press.

Guesnerie, R. and Laffont, J-J. (1984), A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm, Journal of Public Economics, 25, 329–369.
Hammond, P. (1979), Straightforward incentive compatibility in large economies, Review of Economic Studies, 46, 263–282.
Hammond, P. (1987), Altruism. In J. Eatwell, M. Milgate, and Peter Newman, eds., The New Palgrave: A Dictionary of Economics, London: Macmillan, 165–167.
Heckman, J. (1993), What Has Been Learned about Labor Supply in the Past Twenty Years?, American Economic Review, 83(2), 116–21.
**Jacquet, L., Lehmann, E. and Van der Linden, B..(2013), The optimal marginal tax rates with both extensive and intensive responses, Journal of Economic Theory, 148(5): 1770–1805.

Kleven, H.J., Kreiner, C.T. and Saez, E. (2009), The optimal income taxation of couples, Econometrica, 77(2), 537–560.
Laroque, G. (2005), Income maintenance and labor force participation, Econometrica, 2005, 73(2), 341–376.
Lehmann, E., Simula, L., Trannoy, A. (2014),Tax Me if You Can! Optimal Non-linear Income Tax between Competing Governments," Quarterly Journal of Economics, 129(4), 2014, 1995–2030.
Lindert, P. (2004), Growing public: Social spending and economic growth since the eighteenth century. (Two volumes). Cambridge, MA: Cambridge University Press.
Lollivier S. and Rochet, J-C. (1983), Bunching and second-order conditions: a note on optimal tax theory, Journal of Economic Theory, 31, 392–400.
Mirrlees, J. (1971), An exploration in the theory of optimum income taxation, Review of Economic Studies, 38(1), 175–208.

Mirrlees, J. (1982), Migration and Optimal Income Taxes", Journal of Public Economics, 18, 19–341.
Piketty, T. and Saez, E. (2013), Optimal Labor Income Taxation, Handbook of Public Economics, 5, 391–474.
Rochet, J-C. (1985),The taxation principle and multi-time Hamilton-Jacobi equations, Journal of Mathematical Economics, 14(2), 113–128.
Sadka, E., 1976, On Income Distribution, Incentive Effects and Optimal Income Taxation, Review of Economic Studies, 43, 261–267.

**Saez, E. (2001), Using elasticities to derive optimal income tax rates, Review of Economics Studies, 68, 205–229.
**Saez, E. (2002), Optimal income transfer programs: Intensive versus extensive labor supply responses, Quarterly Journal of Economics, 117(3), 1039–1073.
Saez, E. Slemrod, J. and Giertz, S. (2012), The elasticity of taxable income with respect to marginal tax rates: A critical review, Journal of Economic Literature, 50(1), 3–50.
Saez, E. (2010), Do Taxpayers Bunch at Kink Points?, American Economic Journal: Economic Policy, 2, 180–212.
Salanié, B. (1997), The Economics of Contract, MIT Press.

Seade, J., 1977, On the Shape of Optimal Tax Schedules, Journal of Public Economics, 7, 203–235.
Seade, J., 1982, On the Sign of the Optimum Marginal Income Tax, Review of Economic Studies, 49, 637–643.
**Stiglitz, J. (1982), Self-selection and Pareto efficient taxation, Journal of Public Economics, 17, 49–68.
Stiglitz, J. (1982), Self-selection and Pareto efficient taxation, NBER Working Paper, 632.
Zoutman, F. Jacobs, B. and Jongen E. (2014), Optimal Redistributive Taxes and Redistributive Preferences in the Netherlands, Mimeo.