Income taxation with multidimensional heterogeneity

Laurence Jacquet

THEMA, OFS

September 2016

1. Heterogeneous skills and preferences
   - Tagging literature
   - Models with participation decisions
   - Workfare literature
   - One dimensional aggregator
   - Negative marginal tax rates

2. Multidimensional heterogeneity: general case
   - Pooling of individuals
   - Derivation of the structural tax formula
   - Sign of optimal marginal Tax rates
   - Sufficient statistics tax formula

3. Reference

Mirrlees (ReStud 1971)'s model assumes unobserved heterogeneity to be one-dimensional, which is very restrictive. The optimal tax literature relaxes this assumption:

1. Heterogeneous preferences and skills
2. General case with multidimensional heterogeneity

We will study multidimensional heterogeneity in the Mirrlees' model i.e. a model with **one action**.

Outline

1. Heterogeneous skills and preferences
   - Tagging literature
   - Models with participation decisions
   - Workfare literature
   - One dimensional aggregator
   - Negative marginal tax rates

2. Multidimensional heterogeneity: general case

3. Reference

Heterogeneous skills and disutilities of work: Tagging literature

People differ in skill and their disutility of work in the **tagging literature**.(seminal paper: Akerlof AER 1978)
Assume people differ can be either able or disabled but their ability/disability status is imperfectly observable; signal/tag positively correlated with disability.

- Transfers to disabled must be based on self-identification or self-selection and will be restricted by IC.

IC (strongly) limits transfers to non-working so that able people do not mimic them.

- Gvt can acquire some individual info to relax the IC; tagging.

- The target of redistribution includes those with disabilities who are unable to work, e.g. Diamond and Sheshinsky (JPubE 1995), Parsons (JPubE 1996) who study the use of tagging to facilitate identifying the disabled.
- Boadway and Cuff (ITax 1999) who investigate how ex post monitoring of job search can assist in separating types.
- Boadway, Marceau and Sato (JPubE 1999) who analyze how agency problems between the government and social workers can undermine the tagging of disabled.
- etc.

Heterogeneous skills and disutilities of work (or of migration): When participation decision

- Models of optimal income taxation when labor supply is on the extensive margin assume **differences in the disutility of work as the factor determining participation**

e.g. Diamond (JPubE 1980), Saez (QJE 2002), Jacquet, Lehmann and Van der Linden (JET 2012).

- Heterogeneous skills and heterogeneous disutility of migration: Blumkin, Sadka and Shem-Tov (ITax 2015), Lehmann, Simula and Trannoy (QJE 2014).

Heterogeneous skills and disutilities of work: Workfare

Make transfer aimed to the needy contingent on engaging in some activity, e.g., community service jobs, job search activity, job-training.
E.g. in Australia (known as "mutual obligation"), Canada, U.K., U.S., Netherlands (known as "Work First").

Many rationales in the real world:

- Payment to society for transfer,
- Preserve and enhance skills, labor market experience,
- Prevent long-term dependency,
- Change attitudes to work,
- **Screening device to target needy** = rationale which is common in the Public Econ literature

Workfare in the optimal income tax literature

Workfare can be a useful **screening device** if the opportunity cost of workfare is lower for intended recipients than for potential mimickers

Gain of workfare: Workfare weakens the self-selection constraints applying to the non-needy types $\Rightarrow$ Savings from giving no transfers to the non-needy

Cost of workfare: You're forcing workfare recipients to produce something that has less value than if it was produced on the private market (i.e. workfare reduces the poor's private sector earnings)

Workfare in the optimal income tax literature

- Labor supply along the intensive margin
- Participation in workfare crowds out market work (people who enter workfare can continue to supply some labor to the private labor market)
- More costly to go on workfare for the high-skilled than for the low-skilled; opportunity cost = wage on the private market. Hence, larger opportunity cost for high-skilled people than for low-skilled people ⇒ Workfare screens low-skilled versus high-skilled people

Workfare in the optimal nonlinear income tax literature with intensive margin

Pareto-optimizing problem:

- There will be a value of workfare's productivity above that, workfare is **welfare improving.** Intuition? Workfare produces enough output to compensate for the foregone leisure. (Brett OEP 1998)
- Non-productive workfare is **welfare reducing** (Besley and Coate ReStud 1995, Cuff CJE 2000)

Sketch of the model behind these previous results

2 skill levels (wage rates): $n_2 > n_1$, pay attention to this notation for the skills!

N_1 and N_2: $\#$ of respective agents,

Gross labor earnings: y,

Labor hours: $\ell = y/n$,

Consumption (i.e. after tax income): $c = y - t(y)$,

Workfare hours: r (same r for all individuals in workfare)
r is perfectly substituable for ℓ in utility but produces less output: $\gamma < n_1$,

$v(c, y, r) \equiv c - h\left(\frac{y}{n} + r\right)$ with $h' > 0$, $h'' > 0$

Sketch of the model behind these previous results

Pareto optimizing problem:

$$\max_{c_1,\, c_2,\, y_1,\, y_2,\, r} \; c_1 - h\left(\frac{y_1}{n_1} + r\right) \equiv \mathcal{W}(.)$$

$$\text{s.to: } c_2 - h\left(\frac{y_2}{n_2}\right) \geq \overline{v}^2 \qquad [\rho]$$

$$N_1\left(y_1 - c_1 + \gamma r\right) + N_2\left(y_2 - c_2 + \gamma r\right) = R \quad \text{Budget constr. } [\lambda]$$

$$c_2 - h\left(\frac{y_2}{n_1}\right) \geq c_1 - h\left(\frac{y_1}{n_1} + r\right) \quad \text{ICC} \quad [\alpha]$$

Sketch of the model behind these previous results

$$\frac{d\mathcal{W}}{dr} = -h'\left(\frac{y_1}{n_1}+r\right) + \alpha h'\left(\frac{y_1}{n_2}+r\right) + \lambda\gamma$$

$$= \underbrace{-\left(1-\alpha\right)g'\left(\frac{y_1}{n_1}+r\right)}_{<0}\underbrace{-\alpha\left[g'\left(\frac{y_1}{n_1}+r\right)-g'\left(\frac{y_1}{n_2}+r\right)\right]}_{<0}+\lambda\gamma$$

$$\Rightarrow$$

- Unproductive workfare ($\gamma = 0$) is **welfare decreasing** (Besley and Coate ReStud 1995, Cuff CJE 2000).
- There is a value of γ such that for any productivity above that, workfare will be **welfare improving** (Brett OEP1998).

Other results from this model with labor supply along the intensive margin

- Max. consumption of low-ability (Paternalist):

Non-productive workfare is **welfare improving** (Besley and Coate AER 1992).

$$\max \quad \mathcal{W} \equiv c_1 \quad \text{under three previous constraints}$$

$$\Rightarrow \quad \frac{d\mathcal{W}}{dr} = \underbrace{\alpha g' \left(\frac{y_1}{n_2} + r \right)}_{>0} + \lambda \gamma$$

Key? This objective function does not value the disutility of labor; the welfare loss $-g' \left(\frac{y_1}{n_1} + r \right)$ we had in the Pareto optimizing problem does not appear here in $d\mathcal{W}/dr$.

Introduction of heterogeneous preferences modifies the outcomes

- **Heterogeneous preferences** and Pareto-optimizing problem:

when heterogeneous preferences for leisure (LH vs LL and H) and the Gvt redistributes towards the LL $\Rightarrow$ Workfare weakens the self-selection constraints on both mimickers H and LH (which may compensate for the fact that workfare is non-productive) $\Rightarrow$ **even non-productive workfare may be optimal** (Cuff CJE 2000).

Cuff (CJE 2000) notes that, along the intensive margin, with heterogeneous preferences, **workfare and marginal wage subsidies** push in the same direction and it is reasonable to expect the two policies to coexist.

Labor supply along the extensive margin and introduction of heterogeneous preferences

- **Heterogeneous preferences**
- Labor supply along the **extensive margin** (Brett and Jacquet CJE, 2015)
- Individuals differ along two dimensions: skill levels and onerousness of labor (market work)

$\Rightarrow$
When everyone has the same distaste for required work, unproductive workfare is suboptimal since it is less costly to give incentives to work through the nonlinear income tax.
There is a natural antipathy between workfare and EITC in tax model with extensive margin $\neq$ in model with intensive margin

Detailed intuition:

Assume **some** workfare at the optimum. We can keep the utility of people on workfare unaffected by $dr = -1$ and some $db < 0 \Rightarrow$ No effect on participation decisions. + No effect on workers' utility levels. However, **gain in budget** due to $db < 0$

$\Rightarrow$ Workfare suboptimal.

Unproductive workfare suboptimal when extensive margin

Intuition (cont'd), Screening interpretation:

- In the intensive margin model: (unproductive) Workfare helps to dissuade highly productive workers from claiming benefits (i.e. to screen in n) because it is the highly productive that have the highest opportunity cost of time.

$\neq$ Here, with extensive margin: (unproductive) Workfare cannot help to dissuade the highly productive workers from claiming workfare benefits since the opportunity cost of leisure does not depend on productivity.

- The welfare benefit already provides an instrument to screen over onerousness of labor m.

$\Rightarrow$ There remains no screening role for (unproductive) workfare.

Productive workfare can be optimal when extensive margin

- Introducing **productive** workfare is welfare increasing iff it produces enough output to compensate participants for their disutility of required work.

Unproductive workfare can be optimal when extra heterogeneity in disutility from workfare

- Adding a third dimension of heterogeneity: one in how people experience workfare

$\Rightarrow$ Introducing unproductive workfare can be welfare increasing iff the gain in tax revenue from behavioral responses offsets the increase in welfare benefit

Introducing unproductive workfare and increasing b to compensate average utility induce 2 behavioral responses:

- Some people increase their labor force participation $\Rightarrow$ **gains** in tax revenue (1)
- Some people decrease their labor force participation $\Rightarrow$ **losses** in tax revenue (2)

And, the increase in $b =$ **loss** in tax revenue (3)

Workfare is welfare improving when $(1) + (2) > |(3)|$

Multidimensional heterogeneity and the one dimensional aggregator of Brett and Weymark

- Many models with unobserved characteristics rely on the assumption that the action (which is the single dimension of observation for the gvt, i.e. gross income) only depends on a one-dimensional aggregation of the multidimensional heterogeneity.
- Agents differ in productivity and work opportunity cost (or disutility of work)
- Their **behavior**, however, **only depends on a unidimensional combination of the two underlying parameters**
- $\Rightarrow$ Easy for the government to identify the set of types that are associated with a given level of gross income.
- This agregator avoids the technicalities that typically go with multidimensional heterogeneity.

Multidimensional heterogeneity and the one dimensional aggregator of Brett and Weymark

- With a well designed incentive scheme, the government may infer the "aggregator characteristic" of an agent from his income y, but it is unable, say for a large income and a large agregator, to know whether it comes from a high productivity type or a low disutility of work type.
- This innovation was introduced in **Brett and Weymark (JPubE 2003)**
- Large use of this agregator: Boadway, Marchand, Pestieau and Racionero (JPET 2002), Choné and Laroque (AER 2010), Lockwood and Weinzierl (JPubE 2015), etc.

Multidimensional heterogeneity and a one dimensional aggregator

Let $\mathbf{t}$ denote the vector of unobserved characteristics and assume that intensive decisions depend only on a one-dimensional aggregator denoted $w = X(\mathbf{t})$, so that individuals of type $\mathbf{t}$ have preferences $u(c, y; X(\mathbf{t}))$ over consumption and income and solve $\max_y u(y - T(y), y; X(\mathbf{t}))$

$\Rightarrow$ All individuals with the same $w = X(\mathbf{t})$ are thus facing the same decision program; making the same intensive decisions.

Negative marginal tax rates with multidimensional heterogeneity

- Model with **intensive** margin only
- Individuals differ along their skills and preferences for effort/disutility of work
- $\Rightarrow$ Individuals who earn the same income level have the same aggregator but differ in terms of skill and disutility of work

Therefore at each income level, the social marginal welfare weight values people with distinct characteristics and these social marginal welfare weights may not decrease with income anymore. This is the so-called **composition effect**.

Negative marginal tax rates with multidimensional heterogeneity

- The composition effect may reduce marginal tax rates and may even induce them to become negative, see Cuff (CJE 2000), Boadway, Marchand, Pestieau, Racionero (JPET 2002) and Choné and Laroque (AER 2010).
- E.g., when some groups undervalued in the social objective (e.g. with high disutility of work) are overrepresented at low income levels may

$\Rightarrow$ individuals at the bottom of the income distribution may receive lower social welfare weights than individuals with larger incomes.
$\Rightarrow$ negative marginal tax rates at the bottom.

Outline

Optimal income taxation when people differ along many dimensions including behavioral responses to taxation

- Saez (ReStud 2001) conjectures that his optimal tax formula in terms of sufficient statistics, in the one-dimensional Mirrlees model, is also valid in the multidimensional case, but **did not prove it**.
- Relying on Mirrlees (1971) **structural** tax formula he shows the validity of the sufficient statistics tax formula when the unobserved heterogeneity is one-dimensional.
- Jacquet and Lehmann (WP 2016) derive the optimal tax formula, both the structural one and the one in terms of sufficient statistics when multidimensional heterogeneity. They allow for people earning the same income to have **distinct behavioral responses to taxation** hence do not rely on the usual (exogenous) one-dimensional aggregator.

Recall: Sufficient statistics versus structural tax formulae

Recall:
Sufficient statistics approach derives the optimal tax formula using a tax perturbation, i.e. by considering the effects of an inifinitesimal tax reform on the government's objective.

The sufficient statistics approach consists in focusing on empirical combinations of the primitives of the model, known as "sufficient statistics", that can be estimated using data, rather than considering the full economic structure (Chetty ARE 2009).

The sufficient statistics tax formula is enough to indicate the direction of desirable tax reforms but a structural formula is required to correctly implement the tax schedule.

The model

- Individuals differ along their skills w and a vector of characteristics $\theta \in \Theta$. Labor supply elasticity can be one of these characteristics.
- Call a *group* a subset of individuals with the same θ.
- CDF of θ is $\mu(\theta)$ over the potentially multidimensional set Θ (which is compat and measurable).
- The conditional skill density is $f(.|\theta)$ with support $\mathbb{R}_+$.

The first-order condition is:

$$1 - T'(Y(w,\theta)) = \frac{v'_y(Y(w,\theta); w, \theta)}{u'(C(w,\theta))} \quad (1)$$

where $\dfrac{v'_y(y; w, \theta)}{u'(c)}$ is the Marginal Rate of Substitution (MRS).

Identification of individuals who pool at the same level of income

Difficulty with this multidimensional screening problem: Being able to characterize the set of individuals who pool at the same income level.

A method: an endogenous **pooling function** (Jacquet and Lehman, WP 2016).

Within-group single crossing condition

Within-group single-crossing condition (AS 1):

$$v''_{yw}(y; w, \theta) < 0 \qquad \Leftrightarrow \qquad \text{MRS decreasing in } w$$

and limiting conditions:

$$\lim_{w \mapsto 0} v'_y(y; w, \theta) = +\infty \qquad \textit{and} \qquad \lim_{w \mapsto +\infty} v'_y(y; w, \theta) = 0$$

Example:

$$U(c, y; w, \theta) = u(c) - \left(\frac{y}{w}\right)^{1+\frac{1}{\theta}}$$

where $\theta \in \Theta \subset \mathbb{R}_+$ is the (Frish) labor supply elasticity

Taxation principle

- Individuals of type (w,θ) solves $\max_y \quad U(y-T(y),y;w,\theta)$
- Let $Y(w,\theta)$ be "'the" solution and let $C(w,\theta)=Y(w,\theta)-T(Y(w,\theta))$.

$$\forall y' \in \mathbb{R}_+ \qquad U(C(w,\theta),Y(w,\theta);w,\theta) \geq U(y'-T(y'),y';w,\theta)$$

- Take $y'=Y(\hat{w},\hat{\theta})$ leads to the incentive (self-selection) constraints:

$$u\left(C(w,\theta)\right)-v\left(Y(w,\theta);w,\theta\right) \geq u\left(C(\hat{w},\hat{\theta})\right)-v\left(Y(\hat{w},\hat{\theta});w,\theta\right)$$

- Choosing a nonlinear income tax schedule $y \mapsto T(y)$ and let individual choose their labor supply amounts to choose an allocation $(w,\theta) \mapsto (C(w,\theta),Y(w,\theta))$ that verify IC constraints (Revelation principle).
- The taxation principle (Hammond (ReStud 1979), Guesnerie (1995)) ensures the reciprocal holds.

The Government

The gvt maximizes a type-specific $\Phi(\cdot; w, \theta)$ social welfare function:

$$\int_{\theta\in\Theta} \left\{ \int_0^{+\infty} \Phi\left(U(w,\theta); w, \theta\right) \; f\left(w|\theta\right) \, dw \right\} d\mu(\theta)$$

subject to the budget constraint (multiplier $\lambda > 0$):

$$\int_{\theta\in\Theta} \left\{ \int_0^{+\infty} \left[Y\left(w,\theta\right) - C\left(w,\theta\right)\right] f\left(w|\theta\right) \, dw \right\} d\mu(\theta) \geq 0$$

and to incentive constraints (IC): $\forall (w, \hat{w}, \theta, \hat{\theta}) \in \mathbb{R}_+^2 \times \Theta^2$

$$u\left(C(w,\theta)\right) - v\left(Y(w,\theta); w, \theta\right) \geq u\left(C(\hat{w},\hat{\theta})\right) - v\left(Y(\hat{w},\hat{\theta}); w, \theta\right)$$

Incentive compatible allocations

Incentive Constraints (IC) implies within-group IC: $\forall (w, \hat{w}, \theta) \in \mathbb{R}_+^2 \times \Theta$:

$$u\left(C(w,\theta)\right) - v\left(Y(w,\theta); w, \theta\right) \geq u\left(C(\hat{w},\theta)\right) - v\left(Y(\hat{w},\theta); w, \theta\right)$$

Within each group θ:

- $w \mapsto Y(w,\theta)$ is nondecreasing.
- the envelope first-order incentive constraint holds:

$$\dot{U}(w,\theta) = -v'_w\left(Y(w,\theta); w, \theta\right) \qquad \text{(IC1)}$$

Incentive compatible allocations (cont'd)

"Smooth allocation" assumption (AS 2): For each θ, $Y(\cdot,\theta)$ is differentiable, with $\dot{Y}(w,\theta) > 0$, $Y(0,\theta) = 0$ and $\lim_{w \mapsto \infty} Y(w,\theta) = \infty$.

Pooling: Two individuals in different groups earn the same income.

- The "smooth allocation" assumption makes pooling unavoidable.
- Remark: We neglect the (unrealistic) possibility that two individuals in the same group θ with different skill levels earn the same income.
- "smooth allocation" assumption is automatically verified with isoelastic preferences.

Pooling function

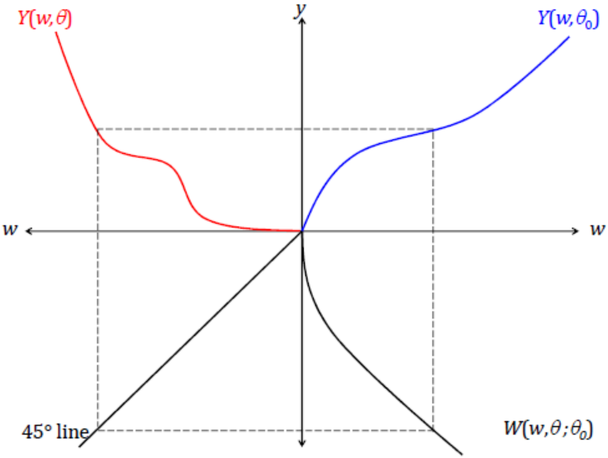
How can each within-group allocation $w \mapsto (Y(w,\theta), C(w,\theta))$ be mutually incentive compatible?

- From "smooth allocation" assumption, $w \stackrel{Y(\cdot,\theta)}{\longrightarrow} Y(w,\theta)$ is increasing from 0 to ∞.
- For each $y \in \mathbb{R}^+$, there exists a single w such that $Y(w,\theta) = y$.
- Take a reference group θ_0 and a skill level w, taking $y = Y(w,\theta_0)$, there thus exists a single skill denoted $W(w,\theta)$ such that:

$$Y(W(w,\theta),\theta) \stackrel{\text{AS } 2}{\equiv} Y(w,\theta_0) \text{ and } C(W(w,\theta),\theta) \stackrel{\text{IC}}{\equiv} C(w,\theta_0) \quad (2)$$

- The pooling function verifies: $w \stackrel{Y(\cdot,\theta_0)}{\longrightarrow} Y(w,\theta_0) \stackrel{Y^{-1}(\cdot,\theta)}{\longrightarrow} W(w,\theta)$ (and thus verifies "smooth allocation" assumption)

Pooling function



The pooling function $W(w,\theta)$ provides the allocation for any group θ from the allocation $w \mapsto (C(w,\theta_0), Y(w,\theta_0))$ specific to group θ_0:

Under assumptions AS 1 and AS 2, the bundle $(C(w,\theta_0), Y(w,\theta_0))$ designed for types (w,θ_0) is also designed for types $(W(w,\theta),\theta)$ where $W(w,\theta)$ solves:

$$\frac{v'_y\left(Y\left(w,\theta_0\right);w,\theta_0\right)}{u'\left(C\left(w,\theta_0\right)\right)} = \frac{v'_y\left(Y\left(w,\theta_0\right);W\left(w,\theta\right),\theta_0\right)}{u'\left(C\left(w,\theta_0\right)\right)} \tag{3}$$

From single-crossing assumption, a single skill level $W(w,\theta)$ solves the previous equation for each θ.

If two individuals belonging to two different groups earn the same income $Y(w, \theta_0)$

- They must face the same marginal tax rate $T'(Y(w, \theta_0))$
- They must thus have the same MRS.

Formal proof: Derive both sides of the definition (2) and use the individual's FOC.
As $U(c, y; w, \theta) = u(c) - v(y; w, \theta)$, the equality of MRS in Equation (3) simplifies to:

$$v'_y(Y(w, \theta_0); w, \theta_0) = v'_y(Y(w, \theta_0); W(w, \theta), \theta)$$

- This allows for endogenous pooling since it depends on $Y(\cdot, \theta_0)$.

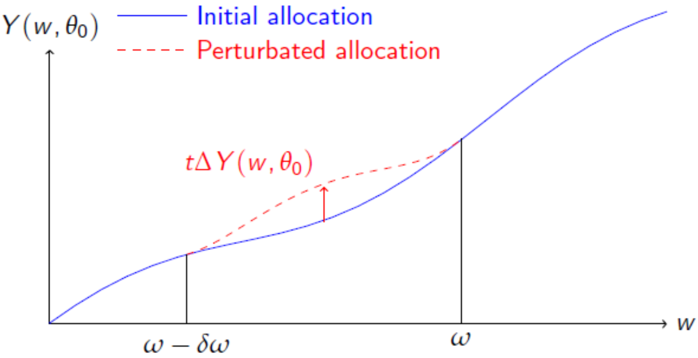
Solving of the optimization problem

This **pooling function** and an **allocation perturbation** method ⇒ able to derive the optimal structural tax formula.

- Calculus of variation
- A set of perturbations **of the allocation** in the reference group
- Thanks to pooling function ⇒ able to describe how the allocation is modified in other groups
- Computation of the **Gâteaux derivatives** of the Lagrangian **in the direction of these perturbations**. These Gâteaux derivatives must be equal to zero (since we are at the optium) ⇒ optimal structural tax formula.

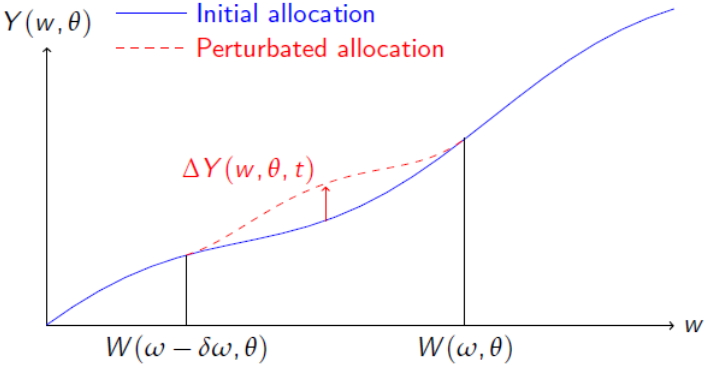
Allocation perturbation method

Consider a set of perturbations of the allocation in the reference group θ_0 such that $Y(\cdot, \theta_0)$ is only modified in the interval $[\omega - \delta\omega, \omega]$ by differentiable amounts $\Delta Y(w, \theta_0, t) = t\Delta Y(w, \theta_0)$, the perturbed allocations remaining increasing.


$Y(w, \theta_0)$
Initial allocation
Perturbed allocation
$t\Delta Y(w, \theta_0)$
w
$\omega - \delta\omega$
ω

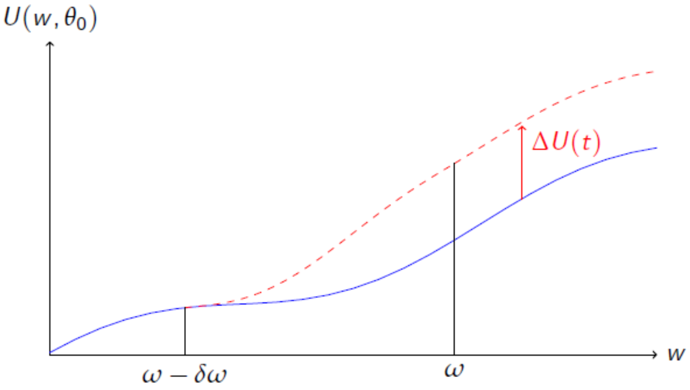
Allocation perturbation method

- Because of the pooling condition, in group $\theta \neq \theta_0$, the allocations are perturbed such that such that $Y(\cdot, \theta)$ is only modified in the interval $[W(\omega - \delta\omega, \theta), W(\omega, \theta)]$ by some $\Delta Y(w, \theta, x)$.


$Y(w, \theta)$
Initial allocation
Perturbated allocation
$\Delta Y(w, \theta, t)$
w
$W(\omega - \delta\omega, \theta)$
$W(\omega, \theta)$

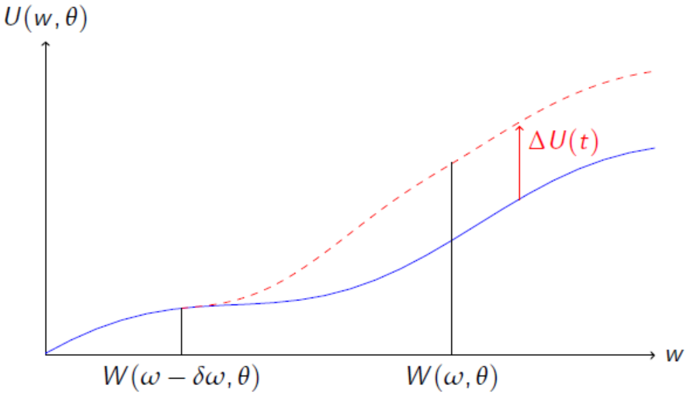
Allocation perturbation method

- By within-group IC, these perturbations induce no change in utility below $W(\omega - \delta\omega, \theta)$ but a uniform change in utility above $W(\omega, \theta)$.


$U(w, \theta_0)$
$\Delta U(t)$
w
$\omega - \delta\omega$
ω

Allocation perturbation method

- As pooling and income $Y(\cdot, \theta)$ are not modified above $W(\omega, \theta)$, these uniform increases in utilities above $W(\omega, \theta)$ must be equal to the same $\Delta U(t)$ for all groups θ:


$U(w,\theta)$
$\Delta U(t)$
w
$W(\omega - \delta\omega, \theta)$
$W(\omega, \theta)$

Allocation perturbation method (Jacquet and Lehmann 2016)

- Normalize these perturbations by ΔU instead of t.
- Compute the **Gateaux derivative** with respect to these perturbations on the Lagrangian.
- Take the limit of this derivative when $\delta\omega$ tends to 0.

Structural optimal tax formula

$$\frac{T'(Y(\omega,\theta_0))}{1-T'(Y(\omega,\theta_0))}\int_{\theta\in\Theta}\frac{v'_y\left\langle W(\omega,\theta),\theta\right\rangle}{-v''_{yw}\left\langle W(\omega,\theta),\theta\right\rangle}f\left(W(\omega,\theta)|\theta\right)\,d\mu(\theta)$$

$$=u'(C(\omega,\theta_0))\iint_{x\geq W(\omega,\theta),\theta\in\Theta}\left(\frac{1}{u'\left(C(x,\theta)\right)}-\frac{\Phi'_U\left\langle x,\theta\right\rangle}{\lambda}\right)f(x|\theta)dx\;d\mu(\theta)$$

$$0=\iint_{x\in\mathbb{R}_+,\theta\in\Theta}\left(\frac{1}{u'\left(C(x,\theta)\right)}-\frac{\Phi'_U\left\langle x,\theta\right\rangle}{\lambda}\right)f(x|\theta)dx\;d\mu(\theta)$$

This structural formula being expressed in terms of policy-invariant primitives, it is numerically implementable to real data.

Sign of optimal marginal Tax rates

- Under weighted utilitarian or Maximin social preferences, optimal marginal tax rates are positive.
- These objectives functions imply that all individuals at income y are socially valued identically, unlike in Boadway, Pestieau & Racionero (JPET 2006), Choné and Laroque (AER 2010), Lockwood and Weinzierl (JPubE 2015) $\Rightarrow$ no negative marginal tax rates.

Sufficient statistics tax formula

Rewrite the above structural formula in terms of sufficient statistics (i.e. social welfare weights, behavioral elasticity, income density).
To do so:

- Define **total** behavioral responses, taking into account the **circularity process (due to the nonlinearity of the tax schedule)**, i.e. that any income response to a tax reform induces a change in marginal tax rates that triggers a further income response.

- Direct responses, **which are the ones typically estimated**, ignore this circularity process. To obtain total behavioral responses, direct responses must be timed by a corrective term:

$$\begin{aligned}
\varepsilon(y;\theta) &= \frac{1-T'(y)}{1-T'(y)+y\ T''(y)\ \varepsilon^{\star}(y;\theta)}\ \varepsilon^{\star}(y;\theta) \\
\eta(y;\theta) &= \frac{1-T'(y)}{1-T'(y)+y\ T''(y)\ \varepsilon^{\star}(y;\theta)}\ \eta^{\star}(y;\theta)
\end{aligned}$$

$\varepsilon^{*}(y;\theta)$ is the compensated elasticity w.r.to the retention rate when $T''=0$.
$\eta^{*}(y;\theta)$ is the income effect when $T''=0$.

To obtain sufficient statistics

- The averaging procedure of sufficient statistics is far from intuitive!:

1. Every direct sufficient statistics has to be multiplied by a **group corrective term** ($\Rightarrow$ required structural approach)

$$\frac{1 - T'(y)}{1 - T'(y) + yT''(y)\,\varepsilon^*(y;\theta)}$$

to obtain the **total** sufficient statistics i.e. **those including the circularity process**.

Example:
The total compensated elasticity w.r.to the retention rate at income y in group θ is:

$$\varepsilon(y;\theta) = \underbrace{\frac{1 - T'(y)}{1 - T'(y) + yT''(y)\,\varepsilon^*(y;\theta)}}_{\text{corrective term}}\varepsilon^*(y;\theta)$$

2. Compute the weighted average of every total sufficient statistics across groups, the weights being the conditional income density for each group. Example: the mean total compensated elasticity at income y that will appear in the tax formula is:

$$\widehat{\varepsilon}(y) = \frac{\int\limits_{\theta\in\Theta} \frac{1-T'(y)}{1-T'(y)+yT''(y)\varepsilon^*(y;\theta)} \varepsilon^*(y;\theta)\, h(y|\theta)\, d\mu(\theta)}{\int\limits_{\theta\in\Theta} h(y|\theta)\, d\mu(\theta)}$$

$$= \frac{\int\limits_{\theta\in\Theta} \varepsilon(y;\theta)\, h(y|\theta)\, d\mu(\theta)}{\int\limits_{\theta\in\Theta} h(y|\theta)\, d\mu(\theta)}$$

The following could be an intuitive approach but this is **wrong**:

1. Compute the simple average of every direct sufficient statistic (for a given level of income).
2. Multiply the obtained average by the corrective term.

Be cautious when directly implementing the tax formula with sufficient statistics:

- **Sufficient statistics take distinct values at the optimum and when one estimates them in the actual economy!**
- **Sufficient statistics are endogenous: circularity process affect elasticities and income effects, multidimensional heterogeneity is a source of composition effect (individuals who earn a given income level are not the same in the actual and optimal economies)**

See Jacquet, Lehmann (WP 2016):

- They quantify this bias on US data.
- This error on T' can be up to 10 percentage points.

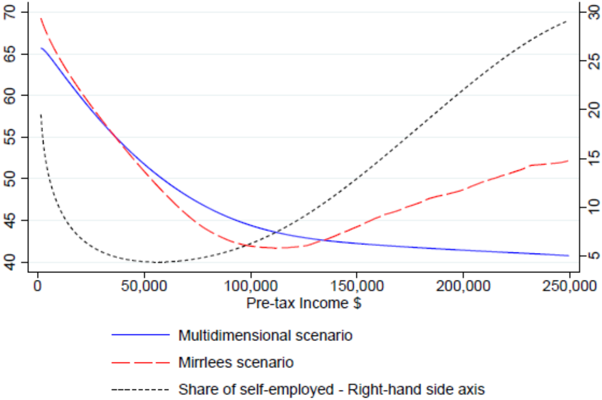
Optimal tax formula in terms of sufficient statistics

$$\frac{T'(y)}{1-T'(y)} = \frac{1}{\widehat{\varepsilon}(y)} \frac{\int_y^\infty \{1-\widehat{g}(z)-\widehat{\eta}(z)\,T'(Y(z))\}\cdot \widehat{h}(z)\,dz}{y\widehat{h}(y)}$$

where:

- $\widehat{h}(\cdot)$ is the **real** income density at the optimum.
- $\widehat{g}(y)$ is the mean of welfare weights across individuals who earn y.
- $\widehat{\varepsilon}(y)$ is the mean of **total** compensated elasticity.
- $\widehat{\eta}(y)$ is the mean of **total** income effects.
- Be **careful in the calculation of these means** (see above)!

Optimal marginal tax rates (in the U.S.) when self-employed and salary workers have distinct skills and di erent labor supply elasticities.



Tax perturbation versus allocation perturbation

Tax perturbation approach requires:
1. Restriction on the tax function that is perturbed: the tax function $T(\cdot)$ is twice differentiable.
2. Restrictions on the way the allocation is affected by the tax perturbation (which is internally inconsistent since the allocation is endogenous) which prevents jumps in the labor supply when a tax reform occurs and ensures that the individual FOC corresponds to the unique global maximum:

- For all $(w, \theta) \in \mathbb{R}_+ \times \Theta$, the second-order condition holds strictly: $Y_y\left(Y(w, \theta), 0, 0; w, \theta\right) < 0$.
- For all $(w, \theta) \in \mathbb{R}_+ \times \Theta$, the function $y \mapsto u(y - T(y)) - v(y; w, \theta)$ admits a unique global maximum.

Tax perturbation versus allocation perturbation (cont'd)

Allocation perturbation approach requires:

- Additive separable utility
- Restrictions (only) on the set of allocations to be perturbed; smooth allocations: $w \longmapsto Y(w, \theta)$ nondecreasing for each $\theta \in \Theta$.

Tax perturbation and allocation perturbation approaches are the two faces of the same coin.

Outline

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

Akerlof, G. A. (1978), The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Training, American Economic Review, 68, 8–19.

Besley, T. and Coate, S. (1992), Workfare versus welfare: Incentive arguments for work requirements in poverty-alleviation programs, American Economic Review, 82, 249–261.

Besley, T., Coate, S. (1995), The design of income maintenance programmes, Review ofEconomic Studies, 62, 187–221.

Blumkin, T., Sadka, E. and Shem-Tov, Y. (2015), International tax competition: zero tax rate at the top re-established, International Tax and Public Finance, 22(5), 760–776

Boadway, R. and Cuff, K. (1999), Monitoring job search as an instrument for targeting transfers, International Tax and Public Finance, 6, 317–337.

Boadway, R., Marceau, N. and Sato, M. (1999), Agency and the design of welfare systems, Journal of Public Economics, 73, 1–30.

Boadway, R., Marchand, M., Pestieau, P. and M. Del Mar Racionero (2002), Optimal redistribution with heterogeneous preferences for leisure, Journal of Public Economic Theory, 4(4), 475–498.

Brett, C. (1998), Who should be on workfare? The use of work requirements as part of an optimal tax mix, Oxford Economic Papers, 50, 607–622.

Brett, C. and L. Jacquet (2015), Workforce or workfare? The optimal use of work requirements when labour is supplied along the extensive margin, Canadian Journal of Economics, 48(5), 1855–1882.

Brett, C. and J. A. Weymark (2003), Financing education using optimal redistributive taxation, Journal of Public Economics, 87(11), 2549–2569.

**Chetty, R. (2009), Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods, Annual Review of Economics, 1(1), 451–488.

**Choné, P. and Laroque, G. (2010), Negative marginal tax rates and heterogeneity, American Economic Review, 100(5), 2532–47.

**Cuff, K. (2000) Optimality of workfare with heterogeneous preferences, Canadian Journal of Economics, 33, 149–174.

Diamond, P. (1980), Income taxation with fixed hours of work, Journal of Public Economics, 13(1), 101–110.

Diamond, P. and Sheshinski, E. (1995), Economic aspects of optimal disability benefits, Journal of Public Economics, 57, 1–23.

Guesnerie, R. (1995), A Contribution to the Pure Theory of Taxation, Cambridge: Cambridge University Press.

Hammond, P. (1979), Straightforward incentive compatibility in large economies, Review of Economic Studies, 46, 263–282.

Jacquet, L. and Lehmann, E. (2016), Optimal Taxation with Heterogeneous Skills and Elasticities: Structural and Sufficient Statistics Approaches, THEMA Working Paper n°2016/04, revised version of THEMA WP n°2014/01, March 2016. Latest version: June 2016.

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.

Laffont, J.-J., Maskin, E. and Rochet, J.-C. (1987), Optimal nonlinear pricing with two-dimensional characteristics, in Information, Incentives and Economic Mechanism, ed. by T. Grove, R. Radner, and S. Reiter, pp. 256–266. Univ. of Minnesota Press.

Lockwood, B. and Weinzierl, M. (2015), De gustibus non est taxandum: Heterogeneity in preferences and optimal redistribution, Journal of Public Economics, 124, 74–80.

Parsons, D. O. (1996), Imperfect 'Tagging' in Social Insurance Programs, Journal of Public Economics, 62, 183–207.

Piketty, T., and E. Saez (2013), Optimal Labor Income Taxation, in Handbook of Public Economics, vol. 5, ed. by M. F. Alan J. Auerbach, Raj Chetty, and E. Saez, chap. 7, pp. 391–474.

Saez, E. (2001), Using elasticities to derive optimal income tax rates, Review of Economic 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. (2003), Reply on Comparing Elasticity-based Optimal Income Tax Formulas by John T. Revecz, Public Finance, 53(3-4), 480–485.