Optimal taxation of polluting goods...and also clean ones?

February 20, 2012

Abstract

The paper studies the changes in the optimal tax system when an externality is “discovered”, in a Mirlees setting with heterogeneous agents. We consider a world in which people differ in their (separable) preferences and ability. We study how the tax system should be modified when taking into account an externality, compared to the initial optimal system (with no externalities). We introduce three types of goods, a dirty good (polluting transport and energy), a clean substitute good (non polluting transport and energy), and a composite good that includes other types of goods. We assume that people have to pay an access cost in order to consume the clean substitute. We find that if ability and the access cost to the clean (perfect) substitute are negatively correlated, then no good should be taxed if there is no externality. With externality, we find that the optimal way to redistribute amongst agents is, in addition to set a non linear income tax, to tax the dirty good less than the pigovian tax and to levy a positive tax on the clean good as well.

1 Introduction

In recent years, there have been debates, for instance in France, on the desirability of a carbon tax. The main criticism of its detractors was that such a tax reform is regressive, because the energy part is larger in the expenses of the poorer. In addition, some argued that poorer people have less substitution possibilities, in France, because they live far from city centers and thus do not have access to public transportation or city gaz.
If the reform is to be done with a constant budget constraint, the way the tax proceeds are redistributed is of foremost importance for equity reasons. To offset the potential regressive bias of environmental taxes, the French government suggested to give lump sum transfers (“chèques verts”) to people, based on criteria such as income and localization\footnote{Pour nos concitoyens les plus défavorisés et des habitants des territoires ruraux, nous étudions en outre des mesures financières d’accompagnement. Il est hors de question d’appliquer uniformément ce dispositif à des Français qui ont le choix et à d’autres qui ne l’ont pas » François Fillon (08|2009).}. However, this solution was not very popular and its implementation seemed complicated, as it required accurate informations on agents. The aim of the paper is to study how the tax system should be modified when an externality is “discovered”, in order to maximize social welfare, and thus taking into account equity considerations.

We introduce three types of goods, a dirty good (polluting transport and energy), a clean good (non polluting transport and energy), and a composite good that includes other types of goods. We seek to find the optimal indirect tax system, in a second best Mirlees setting when a nonlinear income tax and linear taxes on commodities can be used, and when neither productivity nor hours worked are observable. When a carbon tax is introduced, we find that redistribution could be achieved by modifying the relative level of other indirect taxes. In particular, if ability and the access cost to the clean (perfect) substitute are negatively correlated, then the tax on the clean good should increase (relative to the composite good) when the externality is “discovered”.

If agents are “indistinguishable”, that is to say if their consumption choices depend only on their disposable income, the Atkinson Stiglitz theorem (Atkinson and Stiglitz (1976)) holds even when there is an externality : it
is unnecessary to use indirect taxes, apart from the tax on the commodity that generates the externality (Gauthier and Larouque (2008), Kaplow (2004), Kaplow (2006)). As stated in Cremer et al. (1998) "Any Given commodity tax and income-tax system, differential commodity taxation can be eliminated in a manner that results in a Pareto improvement". The optimal carbon tax is then equal to the Pigouvian tax, i.e. the marginal damage from consumption of polluting goods divided by the marginal cost of public funds. In particular, this is the case if agents differ only in their productivity and if utility function (common to all) is separable between leisure and other consumption goods.

The Atkinson-Stiglitz theorem is no longer valid, even in the case of a separable utility function, when agents differ in another characteristic than their productivity (Boadway and Pestieau (2002), Kaplow (2008) Saez (2002), Cremer, Pestieau and Rochet (2001)). In this case, it is difficult to obtain general results (see Rochet and Chone (1998)). Indirect taxes on goods can be useful because they allow to relax the incentive constraints, and thus allow better redistribution.

When combining separable heterogeneous preferences with externalities, one finds that the tax on the polluting good is the sum of the Pigouvian tax and the usual formula for commodity taxation (see Cremer et al. (1998), Cremer, Galvani and Ladoux (2001) ). Specifically, as explained by Kopczuk (2003), "one can build the following policy prescription: correct the externality directly, using the Pigouvian tax (imposed on the dirty commodity), then find and apply the optimal taxes while ignoring the externality by using prices corrected by the Pigouvian tax and taking into account that some income is collected by it."

The general formulation corresponds to the classical calculation rule, to which is added a tax equal to marginal damage only on the good that causes the externality, the principle of “targeting” of Dixit (1985) is generalized (Kopczuk (2003) ): we must correct the externality by targeting the source of the externality directly.

This result does not imply, however, that taking into account the externality does not change the existing tax system, calculation rules remain the same, but the externality can change the relative level of all other taxes. In their applied paper, Cremer et al. (2002) look at the optimal tax system when externality is taken into account. They consider heterogeneous agents who consume two types of goods (polluting goods and a composite good that puts together all the other types of goods). As a result, the only room for redistribution, in their paper, is to set up a tax on the dirty good which is less than the Pigouvian tax, and to change the direct tax system in order to make it more redistributive. We account here for the fact that the inequality between people has to do with the differences in the ease of access to clean substitute for the dirty goods. We introduce a third type of good, which can be consumed easily only by some agents. We focus on changes in the system of direct and indirect taxes when an externality is taken into account (in comparison with cases where there
is no attempt to correct it), and when people differ along two dimensions. Individuals do not all have the same productivity nor the same preferences: some have the choice between meeting their demand for energy and transport with clean technologies (city gas, transportation) or dirty (car, home heating oil) while others only have access to dirty energy and transport.

2 The model

The utility achieved by an individual consuming $x$, $e_1$ (dirty), $e_2$ (clean), and working to generate the rough income $z$ is:

$$u(x, e_1 + e_2) - \theta \psi(e_2) - \sigma \varphi(z)$$

where $u$ is homogenous with degree one, $\psi$ a convex function measuring a potential transaction or access cost to the clean energy and $\varphi$ the desutility of effort (labor). $\theta \in \Theta \subseteq \mathbb{R}^+$ and $\sigma \in \Sigma \subseteq \mathbb{R}^+$, are two individual parameters measuring the intensity of the transaction cost and desutility of effort. The marginal cost of the consumption good is $c$ and the common marginal cost of both dirty and clean energy is $\alpha$. Let $t, \tau_1, \tau_2$ be the linear taxes on goods. And let $T(.)$ be the non linear income tax, chosen optimally given the indirect tax system $t, \tau_1, \tau_2$. Denote $N(z) = z - T(z)$ the net income.

Facing this tax structure, the consumer solves (for $N(.) t, \tau_1, \tau_2$ given):

$$\max_{x, e_1, z} \{ u(x, e_1 + e_2) - \theta \psi(e_2) - \sigma \varphi(z), (c + t)x + (\alpha + \tau_1)e_1 + (\alpha + \tau_2)e_2 = N(z) \}$$

Let $U(\theta, \sigma)$ be the achieved utility.

**Lemma 1.** For all $N(.) t, \tau_1, \tau_2$, let $z^*, x^*, e_1^*, e_2^*$ functions of $(\theta, \sigma)$ solution of the program,

1. $U$ is a decreasing convex function of $(\theta, \sigma)$,
2. $\partial_\theta U = -\psi(e_2^*), \partial_\sigma U = -\varphi(z^*), u(x^*, e^*) = U - \theta \partial_\theta U - \sigma \partial_\sigma U, (c + t)x^* + (\alpha + \tau_1)e_1^* + (\alpha + \tau_2)e_2^* = N(z)$

**Proof.** $U$ is convex as maximum of linear functions. Envelope theorem insures the two first equalities.

Define the expense function :

**Definition 1.** For $c$ and $\alpha$ given let the function $D$:

$$D(v) = \min \{ cx + \alpha e, u(x, e) = v \}$$

Since $u$ is homogeneous with degree one, $D$ is linear increasing.

**Lemma 2.** $cx^* + \alpha(e_1^* + e_2^*) \geq D(U - \theta \partial_\theta U - \sigma \partial_\sigma U)$

**Proof.** straightforward since we have $u(x^*, e^*) = U - \theta \partial_\theta U - \sigma \partial_\sigma U$. 

4
3 Taxation with no externality

Assume that the government uses a concave Social Welfare function \( W \). The Social surplus achieved when taxes are \( t, \tau_1, \tau_2 \) writes:

\[
\int_{\Sigma} \int_{\Theta} W(U(\theta, \sigma)) f(\theta, \sigma) d\theta d\sigma + \lambda \int_{\Sigma} \int_{\Theta} [z^*(\theta, \sigma) - cx^*(\theta, \sigma) - \alpha e^*(\theta, \sigma)] f(\theta, \sigma) d\theta d\sigma
\]

We are going now to show that there exists another tax structure involving no indirect taxation that gives a larger social surplus. Given \( U \) we build a function \( V \) of \( \sigma \) such that the weighted sum of agents welfare is at least as large as with the tax system \((N(), \tau_1, \tau_2)\) if all agents with type \( \sigma \) reach utility \( V(\sigma) \), and such that the government revenues can be increased. Define:

\[
V(\sigma) = E(U/\sigma) - \int_{\Theta} U(\theta, s) \frac{\partial_s f(\theta/s)}{f(\theta/s)} d\theta ds + v
\]

Where \( v \) is some constant. We have:

\[
V'(\sigma) = E(\partial_\sigma U/\sigma)
\]

It is possible to choose \( v \) such that:

\[
\int_{\Theta} U(\theta, s) f(\theta, \sigma) d\theta = \int_{\Sigma} E(U/\sigma) f(\sigma) d\sigma = \int_{\Sigma} V(\sigma) f(\sigma) d\sigma
\]

The value of \( v \) must be:

\[
v = \int_{\Sigma} \int_{\Theta} \left[ \int_{\Theta} U(\theta, s) \frac{\partial_s f(\theta/s)}{f(\theta/s)} d\theta \right] f(\sigma) ds d\sigma
\]

Lemma 3. Let \( F(\theta/\sigma) = \int_0^\theta f(t/\sigma)dt \).

\[
\partial_\sigma (F(\theta/\sigma) \leq 0 \Rightarrow 0 \geq V'(\sigma) \geq \partial_\sigma [E(U/\sigma)]
\]

Proof. It is straightforward that \( V'(\sigma) = E(\partial_\sigma U/\sigma) \leq 0 \).

\[
\partial_\sigma E(U/\sigma) = E(\partial_\sigma U/\sigma) + \int_{\Theta} U(\theta, \sigma) \partial_\sigma f(\theta/\sigma) d\theta
\]

\[
= \int_{\Theta} U(\theta, \sigma) \partial_\sigma f(\theta/\sigma) d\theta + \int_{\Theta} U(\theta, \sigma) \partial_\sigma F(\theta/\sigma) d\theta
\]

As \( f(\theta/\sigma) \) is a density w.r. \( \theta \), \( \int_{\Theta} \partial_\sigma f(\theta/\sigma) d\theta = 0 \). An integration by part gives:

\[
\int_{\Theta} U(\theta, \sigma) \partial_\sigma f d\theta = \int_{\Theta} \partial_\sigma U(\theta, \sigma) \partial_\sigma F(\theta/\sigma) d\theta
\]

As \( \partial_\sigma U(\theta, \sigma) = -\phi(e_2^*) < 0 \), \( \int_{\Theta} U(\theta, \sigma) \partial_\sigma f(\theta/\sigma) d\theta \) has the sign of \( \partial_\sigma F(\theta/\sigma) \). \( \square \)
Now compute the Social welfare when agents with type \((\theta, \sigma)\) reach \(U(\theta, \sigma)\):

\[
W_U = \int_\Sigma \int_\Theta W(U(\theta, \sigma)) f(\theta, \sigma) d\theta d\sigma
\]

As \(W\) is concave:

\[
W_U \leq \int_\Sigma W(E(U/\sigma)) f(\sigma) d\sigma
\]

By definition, \(E(U/\sigma)\) and \(V(\sigma)\) have the same expectation. But if \(V\) is steepless than \(E(U/\sigma)\), the values of \(V\) are more concentrated than those of \(E(U/\sigma)\). Indeed, formally, assume that \(V'(\sigma) \geq \partial_\sigma [E(U/\theta)]\) (Which is implied by \(\partial_\sigma (F(\theta/\sigma) \leq 0, see lemma 3\)). Let \(F(\sigma)\) the cumulative distribution function (CDF) of \(\sigma\) Note \(\hat{U}(\sigma) = E(U/\sigma)\). Call \(F_{\hat{U}}\) the CDF of \(\hat{U}\) and \(F_V\) the CDF of \(V\). Random variable \(\hat{U}\) is a mean-preserving spread of \(V\) (or, equivalently, \(V\) is second-order stochastically dominant over \(\hat{U}\)) if, for all \(\sigma\):

\[
\int_{v_{\text{min}}}^{V} (F_{\hat{U}}(t) - F_V(t)) dt \geq 0 \tag{1}
\]

But \(F_V(t) = P(V(\sigma) \leq v) = P(\sigma \geq V^{-1}(t)) = 1 - F(V^{-1}(t))\). Similarly, \(F_{\hat{U}}(t) = 1 - F(\hat{U}^{-1}(t))\). So that Eq.1 is equivalent to, for all \(v\):

\[
\int_{v_{\text{min}}}^{V} F(\hat{U}^{-1}(t)) - F(\hat{U}^{-1}(v)) dt \leq 0 \tag{2}
\]

Consider:

\[
F(\hat{U}^{-1}(v)) - F(\hat{U}^{-1}(v))
\]

As \(0 \geq V' \geq \hat{U}'\) and \(E(U) = E(V)\) the curves \(V\) and \(\hat{U}\) cross once and once only. Then for low values of \(v\): \(\hat{U}^{-1}(v) \leq V^{-1}(v)\) and for large ones \(\hat{U}^{-1}(v) \geq V^{-1}(v)\), and so \(F(\hat{U}^{-1}(v)) - F(\hat{U}^{-1}(v))\) is first negative and then positive. But as \(E(\hat{U}) = E(V)\), \(\int_{v_{\text{min}}}^{v_{\text{max}}} (F(\hat{U}^{-1}(v)) - F(V^{-1}(v))) dv = 0\). So we have:

\[
\forall v, \int_{v_{\text{min}}}^{v} [F(\hat{U}^{-1}(t)) - F(\hat{U}^{-1}(t))] dt \leq 0
\]

which implies that \(E(U/\sigma)\) is a mean-preserving spread of \(V\) and then that for all concave function \(W\),

\[
\int_{\Sigma} W(E(U/\sigma)) f(\sigma) d\sigma \leq \int_{\Sigma} W(V(\sigma)) f(\sigma) d\sigma
\]

So that the social welfare is higher if all agents with type \(\sigma\) have utility \(V(\sigma)\), than if agents with types \((\theta, \sigma)\) have utility \(U(\theta, \sigma)\).
Let us show that \( V \) can be obtained as the solution of the first problem with some tax levels such that \( \tau_1 = \tau_2 = t = 0 \). In this case, the problem amounts to:

\[
\max \{ u(x, e) - \sigma \varphi(z), cx + \alpha e = N(z) \}
\]

We seek a function \( \tilde{N} \) of \( z \) such that:

\[
V(\sigma) = \max \{ u(x, e) - \sigma \varphi(z), cx + \alpha e = \tilde{N}(z) \}
\]

Indeed let \( \tilde{z}(\sigma) = \varphi^{-1}(-V'(\sigma)) \) and define:

\[
D(v) = \min \{ cx + \alpha e, u(x, e) = v \}
\]

Assume that \( V'' > 0 \). Then \( V(\sigma) - \sigma V'(\sigma) \) is decreasing, as \( \tilde{z}(\sigma) \) is also decreasing, we can define \( \tilde{N} \) such that \( \tilde{N}(\tilde{z}(\sigma)) = D(V(\sigma) - \sigma V'(\sigma)) \). It is easy then to check that:

\[
V(\sigma) = \max \{ u(x, e) - \sigma \varphi(z), cx + \alpha e = \tilde{N}(z) \}
\]

Call \( \tilde{c}(\sigma) \) and \( \tilde{\sigma}(\sigma) \) the achieved consumptions. We have:

\[
\{ \tilde{z}(\sigma), \tilde{c}(\sigma) \} = \arg \min \{ cx + \alpha e, u(x, e) = V(\sigma) - \sigma V'(\sigma) \}
\]

= \arg \max \{ u(x, e), cx + de = \tilde{N}(\tilde{z}(\sigma)) \}

We move now to government revenues with this new tax system. Collected taxes by the government, \( G_V \), are equal to:

\[
G_V = \int_\Sigma [\tilde{z}(\sigma) - \tilde{c}(\sigma) - \alpha \tilde{e}(\sigma)] f(\sigma) d\sigma
\]

That is:

\[
G_V = \int_\Sigma [\varphi^{-1}(-V'(\sigma)) - D(V(\sigma) - \sigma V'(\sigma))] f(\sigma) d\sigma
\]

Consider now the Government revenues with the former tax system \((\tau_1, \tau_2, N())\):

\[
G_U = \int_\Sigma \int_\Theta [\tilde{z}^*(\theta, \sigma) - cx^*(\theta, \sigma) - \alpha e^*(\theta, \sigma)] f(\theta, \sigma) d\theta d\sigma
\]

Using the lemmas in section 2 gives:

\[
G_U \leq \int_\Sigma \int_\Theta [\varphi^{-1}(-\partial_\sigma U) - D(U - \theta \partial_\theta U - \sigma \partial_\sigma U)] f(\theta, \sigma) d\theta d\sigma
\]

By linearity of \( D \)
\[ G_U \leq \int_{\Theta} \int_{\sigma} \varphi^{-1}(-\partial_{\sigma} U) f(\theta, \sigma) d\theta d\sigma - \int_{\Theta} D \left[ \int_{\sigma} (U - \theta \partial_{\theta} U - \sigma \partial_{\sigma} U) f(\theta, \sigma) d\theta \right] d\sigma \]

Since \( \partial_{\theta} U \leq 0 \)

\[ G_U \leq \int_{\Theta} \int_{\sigma} \varphi^{-1}(-\partial_{\sigma} U) f(\theta, \sigma) d\theta d\sigma - \int_{\Theta} D \left[ \int_{\sigma} (U - \sigma \partial_{\sigma} U) f(\theta, \sigma) d\theta \right] d\sigma \]

And then :

\[ G_U \leq \int_{\Theta} \int_{\sigma} \varphi^{-1}(-\partial_{\sigma} U) f(\theta, \sigma) d\theta d\sigma - \int_{\Theta} D \left[ (E[U/\sigma] - \sigma E[\partial_{\sigma} U/\sigma]) f(\sigma) d\sigma \right] \]

But as \( \varphi^{-1} \) is concave :

\[ G_U \leq \int_{\Theta} \varphi^{-1}(E(-\partial_{\sigma} U/\sigma)) f(\sigma) d\sigma - \int_{\Theta} D \left[ (E[U/\sigma] - \sigma E[\partial_{\sigma} U/\sigma]) f(\sigma) d\sigma \right] \]

Then replacing \( E[\partial_{\sigma} U/\sigma] \) by \( V'(\sigma) \) in Eq.3 and using the linearity of \( D \) gives :

\[ G_U \leq \int_{\Theta} \varphi^{-1}(-V'(\sigma)) f(\sigma) d\sigma - \int_{\Theta} D \left[ (V(\sigma) - \sigma V'(\sigma)) f(\sigma) d\sigma \right] \]

We have then found a function of \( \sigma \) only that is such that the total surplus is better than the one for \( U \).

We have to check that \( V \) is convex.

**Proof.**

\[ V''(\sigma) = \partial_{\sigma} \left[ \int_{\Theta} \partial_{\sigma} U f(\theta/\sigma) d\theta \right] \]

That is :

\[ V''(\sigma) = \int_{\Theta} \partial_{\sigma}^2 U f(\theta/\sigma) d\theta + \int_{\Theta} \partial_{\sigma} U \partial_{\sigma} f(\theta/\sigma) d\theta \]

An integration by part gives:

\[ V'' = \int_{\Theta} (\partial_{\sigma}^2 U f(\theta/\sigma) d\theta - \partial_{\sigma} \partial_{\sigma} f(\theta/\sigma) d\theta) \]

The first term is positive. As \( f(\theta/\sigma) \) is a density w.r. \( \theta, \int_{\Theta} \partial_{\sigma} f(\theta/\sigma) d\theta = 0 \)
Let \( F(\theta/\sigma) = \int_0^\theta f(\theta/\sigma) d\theta \), we have \( \partial_\sigma F(\theta/\sigma) = \int_0^\theta \partial_\sigma f(\theta/\sigma) d\theta \). An integration by part gives:

\[
\int_\Theta \partial_\sigma U \partial_\sigma f(\theta/\sigma) d\theta = - \int \partial_{\sigma^2} U \partial_\sigma F d\theta
\]

If \( \partial_{\sigma^2} U \geq 0 \), a sufficient condition for \( V'' \geq 0 \) is \( \partial_\sigma F \leq 0 \). We check under what condition \( \partial_{\sigma^2} U \) is greater than 0. We denote:

\[
w(\theta, N) = \max_{x, e_1, e_2} \{ u(x, e_1 + e_2) - \theta \psi(e_2), (c + t)x + (\alpha + \tau_1)e_1 + (\alpha + \tau_2)e_2 = N \}
\]

Take the optimal income tax schedule \( z \to \hat{N}(z) \) for some given indirect tax system \((t, \tau_1, \tau_2)\), with \( \hat{N}' \geq 0 \). We have that:

\[
U(\theta, \sigma) = \max_z \left\{ w(\theta, \hat{N}(z) - \sigma \varphi(z) \right\}
\]

\[
z(\theta, \sigma) = \arg \max_z \left\{ w(\theta, \hat{N}(z) - \sigma \varphi(z) \right\}
\]

\[
\partial_{\sigma^2} U = -\varphi' \partial_\theta z = \partial_{\sigma^2} U = \partial_{\sigma^2} w \hat{N}' \partial_\sigma z
\]

We have that \( \partial_\sigma z < 0 \) (as \( \partial_{\sigma^2} U > 0 \)). In order to determine whether \( \partial_{\sigma^2} U > 0 \), we study the sign of \( \partial_{\sigma^2} w \). Recall that:

\[
w(\theta, N) = \max_{x, e_1} \{ u(x, e_1 + e_2) - \theta \psi(e_2), (c + t)x + (\alpha + \tau_1)e_1 + (\alpha + \tau_2)e_2 = N \}
\]

So that

\[
\partial_{\sigma^2} w = \partial_{\sigma^2} \psi(e_2) \partial_N e_2
\]

\[
\partial_{\sigma^2} w \leq 0 \iff \partial_N e_2 \geq 0
\]

So that \( \partial_{\sigma^2} w \leq 0 \) if and only if \( e_2 \) is a normal good (i.e. if the consumption of \( e_2 \) increases with disposable income \( N \)). If \( u \) is homogeneous of degree one, we have that \( \partial_{\sigma^2} w = 0 \), so that \( V'' > 0 \).

**Theorem 1.** Let \( F(\theta/\sigma) \) the probability that the energy transaction cost be smaller than \( \theta \) for given cost of effort \( \sigma \). If this probability is decreasing with \( \sigma \) (that is if low productivity people have high transaction costs) then optimal taxation requires only income taxation and no indirect taxes.

## 4 Optimal taxation with externality

We assume now that consume good \( E_1 \) generates the environmental damage:

\[
-dE_1
\]
We choose a constant marginal damage of pollution because emissions from a country at some date only impact marginally the worldwide stock.

We want to show that the standard Atkinson-Stiglitz result does not hold in general. The tax on the dirty good is not equal, in general, to the pigovian tax, and the tax on the clean good is not equal, in general, to zero. As before, we assume that $u(x,e)$ is homogeneous of degree one. We also assume that the government has redistributive tastes, so that he maximizes a sum of $W(U)$ where $W$ is increasing and concave. As $u$ is homogeneous of degree one, $z$ only depends on $\sigma$.

Denoting $E_1$ the sum of emissions, the social surplus is thus:

$$
\int_{\Theta} \int_{\Theta} W(U(\theta,\sigma,\tau,N(\theta)))f(\theta,\sigma)d\theta d\sigma - dE_1
$$

$$
+ \lambda \int_{\Theta} \int_{\Theta} [z^*(\theta,\sigma) - c\theta^*(\theta,\sigma) - \alpha e^*(\theta,\sigma)]f(\theta,\sigma)d\theta d\sigma
$$

If the Atkinson-Stiglitz theorem was to hold, then indirect taxes $\tau_1$ on good 1 would be equal to the social marginal externality cost of consuming good $e_1$, and the tax $\tau_2$ on good 2 would be zero. Assume that this tax system satisfies:

$$
\tau_1 = \frac{d}{\lambda}, \tau_2 = 0
$$

We assume that the income tax $T(z)$ is optimal given this indirect tax system. We look at the effect of a reform $(0,d\tau_1,d\tau_2)$ on goods $(x,e_1,e_2)$. To study whether this tax reform is profitable, we proceed as in Saez (2002). The indirect utility of consumption is again denoted $w(\theta,N)$ where $N$ is disposable income:

$$
w(\theta, N) = \max_{x,e_1,e_2} \{u(x,e_1 + e_2) - \theta \psi(e_2)\}
$$

subject to:

$$
(\alpha + \tau_1)e_1 + \alpha e_2 + cx \leq N
$$

We can rewrite the social planner program. He seeks $(\tau_1,\tau_2,N())$ which maximize:

$$
\int_{\Theta} \int_{\Theta} W(U(\theta,\sigma))f(\theta,\sigma)d\theta d\sigma - d(\int_{\Theta} \int_{\Theta} e_1^*(\theta,\sigma)f(\theta,\sigma)d\theta d\sigma)
$$

$$
+ \lambda \left( \int_{\Theta} \int_{\Theta} (z^* - N(z^*) + \tau_1 e_1^*(\theta,\sigma) + \tau_2 e_2^*(\theta,\sigma)f(\theta,\sigma)d\theta d\sigma) - B \right)
$$

**Tax reform $d\tau_1$** We start from a case where $\tau_1 = \frac{d}{\lambda}$ and $\tau_2 = 0$. We look at the effect of a tax reform $d\tau_1$.

$$
U(\theta,\sigma) = \max_z \{w(\theta,N(z)) - \sigma \varphi(z)\}
$$

So that:

$$
\partial_{\tau_1} U(\theta,\sigma) = \partial_{\tau_1} w(\theta,N)
$$

$$
= -(\partial_N w)e_1 d\tau_1
$$
So that the total effect on welfare can be written:

\[-\left(\int_{\Sigma} \int_{\Theta} W'(U(\theta, \sigma))(\partial_N w)e_1 f(\theta, \sigma)d\theta d\sigma\right) d\tau_1 - dE_1\]

(5)

The tax change has two effects on public fund:

- It raises mechanically the revenue:

  \[\lambda (E_1 d\tau_1 + \tau_1 dE_1)\]

  (6)

- It has a behavioral effect \(dz_{\tau_1}(\theta, \sigma)\), via change in \(z(\theta, \sigma)\):

  \[\lambda \int T'(z(\theta, \sigma)) dz_{\tau_1}(\theta, \sigma)f(\theta, \sigma)d\theta d\sigma\]

  (7)

We see that the \(d\) term in Eq.5 simplifies with the \(\tau_1\) term in Eq.6. So that the total effect is the sum of the three terms:

\[-\left(\int_{\Sigma} \int_{\Theta} W'(U(\theta, \sigma))(\partial_N w)e_1 f(\theta, \sigma)d\theta d\sigma\right) d\tau_1 -\]

\[\lambda E_1 d\tau_1\]

(8)

\[\lambda \int T'(z(\theta, \sigma)) dz_{\tau_1}(\theta, \sigma)f(\theta, \sigma)d\theta d\sigma\]

(9)

(10)

To check the sign of the sum of Eq.8 + Eq.9 + Eq.10, we use the fact that any small income tax reform has no first order effect on welfare because the income tax is optimal. We look at the effect of a small income tax change \(dT\), such that \(dT(z) = E_1(z)d\tau_1\), where \(E_1(z)\) denotes average consumption of \(e_1\) for individuals earning \(z\). The effect of this tax change can also be decomposed into mechanical, welfare and behavioral effect.

**Tax reform \(dT(z)\)**  Denoting \(f(\theta, \sigma|z)\) the probability density of types \((\theta, \sigma)\) knowing \(z\), and \(f(z)\) the probability density of income \(z\), the effect on welfare of this income tax reform is the following:

\[-\int_{z} \int_{\theta, \sigma} W'(U(\theta, \sigma))(\partial_N w)E_1(z)f(\theta, \sigma|z)d\theta d\sigma f(z)dzd\tau_1 - ddE_1^T\]

As before, there is no first order effect via a change in \(z^*\) because the agent had chosen \(z^*\) optimally (given the income tax schedule and the indirect taxes and before the tax reform).

The reform raises revenues:

\[\lambda \int_{z} \left(\int_{\theta, \sigma} (dT(z) + \tau_1 dE_1^T(z)) f(\theta, \sigma|z)d\theta d\sigma\right) f(z)dz\]

(11)
And has behavioral effect:

\[
\lambda \int_z \int_{\theta, \sigma} T'(z(\theta, \sigma)) dz T(\theta, \sigma) f(\theta, \sigma|z) d\theta d\sigma f(z) dz 
\]  

(12)

We see that the \(d\) simplifies with the term with \(\tau_1\). So that the total effect is the sum of the three terms:

\[
- \int_z \int_{\theta, \sigma} W'(U(\theta, \sigma))(\partial_N w) E_1(z)f(\theta, \sigma|z) d\theta d\sigma f(z) dz d\tau_1
\]  

(13)

\[
\lambda \int_z \int_{\theta, \sigma} dT(z) f(\theta, \sigma|z) d\theta d\sigma f(z) dz
\]  

(14)

\[
\lambda \int_z \int_{\theta, \sigma} T'(z(\theta, \sigma)) dz T(\theta, \sigma) f(\theta, \sigma|z) d\theta d\sigma f(z) dz
\]  

(15)

**Comparison between the two reforms** We compare the effects, term by term, of the two tax reforms. The difference in the effect on welfare (Eq.8-Eq.13) is the following:

\[
- \int_z \int_{\theta, \sigma} W'(U(\theta, \sigma))(\partial_N w) E_1(z)f(\theta, \sigma|z) d\theta d\sigma f(z) dz \pi(z) dz d\tau_1
\]  

(16)

If \(\partial_N w\) does not depend on \(\theta\), *conditionnal on* \(z\), then this term can be rewritten:

\[
- \int_z (\partial_N w) \int_{\theta, \sigma} W'(U(\theta, \sigma))(e_1(\theta, \sigma) - E_1(z)) f(\theta, \sigma|z) d\theta d\sigma f(z) dz d\tau_1
\]  

(17)

This is the case if \(u\) is homogeneous of degree one (proof in appendix). Let look at \(U(\theta, \sigma)\), conditional on \(z\). It is easy to see that if \(u\) is homogeneous of degree one, then \(z\) only depends on \(\sigma\). So that, all agents earning \(z\) have the same productivity \(\sigma\), so that they also have the same disutility of labor \(\sigma\phi(z)\). For a given level \(z\), and the associated after tax income \(N\), agents of type \(\theta\) maximize:

\[
w(\theta, N) = \max_{x, e_1, e_2} \{u(x, e_1 + e_2) - \theta \psi(e_2)\}
\]

\[s.t. \quad (\alpha + \tau_1)e_1 + \alpha e_2 + cx \leq N\]

A revealed preference argument gives immediately that, at \(z\) given (and thus at \(\sigma\) given), higher \(\theta\) have lower \(U(\theta, \sigma)\) and thus higher \(W'(U)\). Moreover, amongst agents earning \(z\), the consumption of \(e_1\) increases with \(\theta\) (see proof in appendix), and then increases with \(W'(U)\). As a result, \(\int_{\theta, \sigma} W'(U(\theta, \sigma))(e_1(\theta, \sigma) - E_1(z)) f(\theta, \sigma|z) d\theta d\sigma f(z) dz d\tau_1 > 0\), and thus

\[
- \int_z (\partial_N w) \int_{\theta, \sigma} W'(U(\theta, \sigma))(e_1(\theta, \sigma) - E_1(z)) f(\theta, \sigma|z) d\theta d\sigma f(z) dz d\tau_1 < 0
\]
The two remaining terms are:
\[ \lambda \int \int_{\theta, \sigma} (e_1(z)d\tau_1 - dT(z))f(\theta, \sigma|z)d\theta d\sigma f(z)dz = 0 \]
by definition and
\[ \lambda \int \int_{\theta, \sigma} T'(z(\theta, \sigma))(dz_\tau(\theta, \sigma) - dz_T(\theta, \sigma))f(\theta, \sigma|z)d\theta d\sigma f(z)dz \quad (18) \]

In order to compare \( dz_\tau(\theta, \sigma) \) and \( dz_T(\theta, \sigma) \), we use the fact that the reform \( d\tau_1 \) induces the same behavioral response as a reform of the income tax schedule, specific to each agent of type \((\theta, \sigma)\), with \( dT(\theta, \sigma) = e_1^*(\theta, \sigma)d\tau_1 \) (see Saez (2002)).

Agent \((\theta, \sigma)\) chooses his amount of hours worked as if he was subject to a linear budget constraint \( R + (1 - t)z \), where \( t \) is the marginal tax rate at \( z = z^*(\theta, \sigma) \). He chooses thus \( z^* \) that maximizes: \( U(\theta, \sigma, R + (1 - t)z, z) \) (with \( R = z - T(z) - (1 - t)z \) and \( t = T'(z) \)). Using Slutsky and denoting \(-z^c\) the compensated demand for leisure, we get that:
\[ \frac{\partial z(\theta, \sigma)}{\partial (1 - t)} = \frac{\partial z^c(\theta, \sigma)}{\partial (1 - t)} + \frac{\partial z(\theta, \sigma)}{\partial R}z \]
The behavioral effect of the reform \( dT(\theta, \sigma) \) is then:
\[ dz_\tau_1 = \frac{\partial z(\theta, \sigma)}{\partial R}dR - \frac{\partial z(\theta, \sigma)}{\partial (1 - t)}dt \]
Using that \( R = z - T(z) - (1 - t)z \):
\[ dR = (1 - T'(z))dz - (1 - t)dz - dT(z) + zdt = -dT(z) + zdt \]
and
\[ dt = dT'(z) + T''(z)dz = \frac{\partial e_1}{\partial z}(\theta, \sigma)d\tau_1 + T''(z)dz \]
\[ dz_\tau_1 = -\frac{\partial z(\theta, \sigma)}{\partial R}d\tau_1 + (\frac{\partial z}{\partial R}(\theta, \sigma)z(\theta, \sigma) - \frac{\partial z}{\partial (1 - t)}(\theta, \sigma))dt \]
\[ = -\frac{\partial z}{\partial R}(\theta, \sigma)e_1(\theta, \sigma, z, T(z))d\tau_1 - \frac{\partial z}{\partial (1 - t)}(\theta, \sigma)dt \]
So that the behavioral effect of this reform is:
\[ dz_\tau_1(1 + \frac{\partial z}{\partial (1 - t)}T''(z)) = -\frac{\partial z}{\partial R}e_1d\tau_1 - \frac{\partial z}{\partial (1 - t)}\frac{\partial e_1}{\partial z}d\tau_1 \]

On the other hand, the reform \( dT(z) \) yields to a behavioral effect on agent \((\theta, \sigma)\)
\[ dz_T(1 + \frac{\partial z}{\partial (1 - t)}T''(z)) = -\frac{\partial z}{\partial R}E_1(z)d\tau_1 - \frac{\partial z}{\partial (1 - t)}\frac{\partial E_1}{\partial z}d\tau_1 \]
This term is negative if agents who earn \( \tilde{z} \) (see proof in appendix).

\[
\int_z \int_{(\theta, \sigma)} \frac{T'(z)}{(1 + \frac{\partial z}{\partial (1-t)} T''(z))} \left( -\frac{\partial E_1}{\partial \tilde{z}} d\tau_1 - \frac{\partial z}{\partial (1-t)} \frac{\partial E_1}{\partial \tilde{z}} d\tau_1 \right) f(\theta, \sigma|z) d\theta d\sigma f(z) dz
\]

So that the overall difference between the behavioral terms (Eq18) is:

\[
\int_z \int_{(\theta, \sigma)} \frac{T'(z)}{(1 + \frac{\partial z}{\partial (1-t)} T''(z))} \left( \frac{\partial z}{\partial (1-t)} \frac{\partial E_1}{\partial \tilde{z}} d\tau_1 \right) f(\theta, \sigma|z) d\theta d\sigma f(z) dz
\]

If \( \frac{\partial z}{\partial (1-t)} \) and \( \frac{\partial E_1}{\partial \tilde{z}} \) do not depend on the agent's type, conditionnal on \( z \), then Eq.18 boils down to:

\[
\int_z \int_{(\theta, \sigma)} \frac{T'(z)}{(1 + \frac{\partial z}{\partial (1-t)} T''(z))} \frac{\partial z}{\partial (1-t)} f(\theta, \sigma|z) d\theta d\sigma f(z) dz
\]

In particular, this is the case if \( u(x, e_1 + e_2) \) is homogenous of degree one (see the proof in appendix).

Let look at the remaining term (from Eq.19):

\[
\int_z \int_{(\theta, \sigma)} \frac{T'(z)}{(1 + \frac{\partial z}{\partial (1-t)} T''(z))} \frac{\partial z}{\partial (1-t)} f(\theta, \sigma|z) d\theta d\sigma f(z) dz
\]

This term is negative if agents who earn \( z + dz \) would consume relatively less \( e_1 \) than agent earning \( z \) if they were forced to earn \( z \). We need the distribution of \( (\theta|z) \). But if \( u \) is homogeneous of degree one, then all agents choosing the same \( z \) have the same \( \sigma \) (from the FOC on \( z \), see proof in appendix). So that the distribution \( f(\theta|z) \) can be rewritten \( f(\theta|\sigma = g(z)) \) with \( g < 0 \).

\[
E_1(z) = \int_\theta e_1(\theta, g(z)) f(\theta|\sigma = g(z)) d\theta \tag{20}
\]

\[
\frac{dE_1(z)}{dz} = \int_\theta \frac{\partial e_1(\theta, g(z))}{\partial z} f(\theta|\sigma = g(z)) d\theta + \int_\theta e_1(\theta, g(z)) \frac{\partial e_1(\theta, g(z))}{\partial \sigma} f(\theta|\sigma = g(z)) g'(z) d\theta
\]

So that:

\[
\int_{(\theta, \sigma)} \left( \frac{\partial E_1}{\partial \tilde{z}} - \frac{\partial e_1}{\partial \tilde{z}} \right) f(\theta, \sigma|z) d\theta d\sigma
\]

\[
= -\int_\theta \frac{\partial e_1(\theta, g(z))}{\partial \theta} \frac{\partial e_1(\theta, g(z))}{\partial \sigma} F(\theta|\sigma = g(z)) g'(z) d\theta
\]
But, we know that $F_{\sigma}(\theta|\sigma = g(z)) \leq 0$ and $g' < 0$. Moreover, $\frac{\partial e_1(\theta,g(z))}{\partial \theta} > 0$ (see proof in appendix). We assume here that $T'(z) \geq 0$ for every $z$ (this result is valid when leisure is a non-inferior good and when the government has redistributive tastes, see Seade (1982)). So that increasing $\tau_1$ yields negative marginal surplus.

Exactly the same reasoning applies with $\tau_2$, we can show that increasing $\tau_2$ from 0 to $d\tau_2$ increases total welfare (because, at $z$ given, $\partial_\theta e_2 < 0$).

**Theorem 2.** In the presence of externality, with redistributive tastes and with heterogeneous preferences:

1. In general, the tax on the clean good should not be zero and the tax on the dirty good should not be the pigovian tax.

2. If $\text{cov}(\theta,\sigma) > 0$, the tax on the dirty good should be less than the pigovian tax and the tax on the clean good should be strictly positive.
A Proofs

Lemma 4. if $u$ is homogeneous of degree one, then

1. $\partial^2_{\theta,N}w = 0$
2. At $N$ given, $\partial_\theta e_1 \geq 0$
3. At $N$ given, $\partial_\theta e_2 \leq 0$

Proof. The program of agent $(\theta, \sigma)$ reads:

$$u(x, e_1 + e_2) - \theta \psi(e_2) + (\partial_N w)(N - cx - (\alpha + \tau) e_1 - \alpha e_2)$$

Where $(\partial_N w)$ is the lagrangian multiplier for type $(\theta, \sigma)$ associated with his budget constraint. The solution reads:

$$u_x = (\partial_N w)c$$
$$u_e = (\partial_N w)(\alpha + \tau)$$
$$\theta \psi'(e_2) = (\partial_N w)\tau$$
$$cx + (\alpha + \tau_1)e_1 + \alpha e_2 = N$$

We look for $\partial^2_{\theta,N}w$. At $N$ given, we compute the variation of $w_N$ induced by $d\theta$. We get:

$$u_{xx} dx + u_{xe} d e = d(\partial_N w)c$$
$$u_{ee} de + u_{xe} dx = d(\partial_N w)(\alpha + \tau_1)$$
$$\theta \psi''(e_2) de_2 + d\theta \psi'(e_2) = d(\partial_N w)\tau_1$$
$$cdx + (\alpha + \tau)de - \tau de_2 = 0$$

As $u$ is homogeneous of degree one, $u_x$ and $u_y$ are homogeneous of degree zero and using Euler theorem:

$$u_{xx}x + u_{xe}e = 0$$
$$u_{yy}x + u_{ee}e = 0$$

So that $(x, y)$ is a eigenvector of matrix $\begin{pmatrix} u_{xx} & u_{xy} \\ u_{yx} & u_{yy} \end{pmatrix}$ So that the Hessian $H(u)$ of $u$ is zero. We find that:

$$d(\partial_N w) = 0$$
$$de_2 = -\frac{\phi'(e_2)}{\theta \phi''(e_2)} d\theta < 0$$
$$de_1 = -c^2 u_{ee} - (\alpha + \tau_1)\alpha u_{xx} + (2\alpha + \tau_1)cu_{xx} - \frac{\phi'(e_2)}{(-c^2 u_{ee} - (\alpha + \tau_1)^2 u_{xx} + 2(\alpha + \tau_1)cu_{ee}) \theta \phi''(e_2)} d\theta$$

$\square$
Lemma 5. if $u$ is homogeneous of degree one, then $\frac{\partial z}{\partial (1-t)}$ and $\frac{\partial z}{\partial R}$ do not depend on the agent’s type, conditionnal on $z$.

Proof. The program of agent $(\theta, \sigma)$ reads:

$$u(x, e_1 + e_2) - \theta \psi(e_2) - \sigma \varphi(z) + (\partial_N w)(N(z) - cx - (\alpha + \tau)e_1 - \alpha e_2)$$

Where $(\partial_N w)$ is the lagrangier multiplier for type $(\theta, \sigma)$ associated with his budget constraint. The solution reads:

$$\begin{align*}
    u_x &= (\partial_N w)c \\
    u_e &= (\partial_N w)(\alpha + \tau) \\
    \theta \psi'(e_2) &= (\partial_N w)\tau \\
    \sigma \varphi'(z) &= (\partial_N w)(1-t) \\
    cx + (\alpha + \tau)e_1 + \alpha e_2 &= R + (1-t)z
\end{align*}$$

We look for $\frac{\partial z}{\partial (1-t)}$. We compute first orders effect of a compensated variation of $(1-t)$, we get

$$\begin{align*}
    u_{xx} dx + u_{xe} de &= d(\partial_N w)c \\
    u_{ee} de + u_{xe} dx &= d(\partial_N w)(\alpha + \tau) \\
    \theta \psi''(e_2)de_2 &= d(\partial_N w)\tau \\
    \sigma \varphi''(z)dz &= d(\partial_N w)(1-t) + (\partial_N w)d(1-t)
\end{align*}$$

We find that

$$d\tilde{z} = \frac{(\partial_N w)(\theta \psi''(e_2)(-c^2 u_{ee} - (\alpha + \tau)u_{xx} + 2(\alpha + \tau)cu_{xe}) + \tau^2 H(u)) d(1-t)}{\sigma \varphi''(z)(\theta \psi''(e_2)(-c^2 u_{ee} - (\alpha + \tau)u_{xx} + 2(\alpha + \tau)cu_{xe}))}$$

$$\begin{align*}
    &= (\partial_N w)\frac{(\theta \psi''(e_2)(-c^2 u_{ee} - (\alpha + \tau)u_{xx} + 2(\alpha + \tau)cu_{xe}))}{(\partial_N w)\sigma \varphi''(z)(\theta \psi''(e_2)(-c^2 u_{ee} - (\alpha + \tau)u_{xx} + 2(\alpha + \tau)cu_{xe}))}d(1-t) \\
    &= \frac{\varphi'(z)}{(1-t)\varphi''(z)}d(1-t)
\end{align*}$$

Similarly, we compute $\frac{\partial z}{\partial R}$. We compute first orders effect of a variation of $R$, we get

$$\begin{align*}
    u_{xx} dx + u_{xe} de &= d(\partial_N w)c \\
    u_{ee} de + u_{xe} dx &= d(\partial_N w)(\alpha + \tau) \\
    \theta \psi''(e_2)de_2 &= d(\partial_N w)\tau \\
    \sigma \varphi''(z)dz &= d(\partial_N w)(1-t) \\
    cdx + (\alpha + \tau)de - \tau de_2 &= dR + (1-t)dz
\end{align*}$$
We get that:

\[ dz = \frac{(\partial_N w) \left( -(1-t)\theta \psi''(e_2)H(u) \right) d(1-t)}{\sigma \phi''(z)\theta \psi''(e_2)\left( -c^2 u_{ee} - (\alpha + \tau)u_{xx} + 2(\alpha + \tau)cu_{xe} \right) + ((1-t)^2 \theta \psi''(e_2) + \sigma \phi''(z)\tau^2)H(u)} \]

There is no first order effect of the variation of R on z.

\[ \square \]
References


**URL:** [http://restud.oxfordjournals.org/content/49/4/637.abstract](http://restud.oxfordjournals.org/content/49/4/637.abstract)