Yet Another Reason to Tax Goods*

Carlos E. da Costa[†]
Getulio Vargas Foundation

October 31, 2003

Abstract

When, in a dynamic model, choices by an agent : i) are not observed, and; ii) affect preferences conditional on the realization of types, new and unexpected features come up in Mirrlees' (1971) optimal taxation framework. In the simplest possible model where a non-trivial filtration may be incorporated, we show how these two characteristics make it necessary for IC constraints to be defined in terms of strategies rather than pure announcements. Tax prescriptions are derived, and we are able to show that uniform taxation prescription of Atkinson and Stiglitz fails to hold, in general. Clean results regarding capital income taxation are not easy to come about because usual assumption on preferences do not allow for determining which constraints bind at the optimum. However, in the most 'natural' cases, we show that return on capital ought to be taxed.
Keywords: Optimal Taxation; Non-observability; Dynamic Contracts.
JEL Classification: H21, D82.

1 Introduction

Despite its enormous impact, not only in the field of public economics but in all economic theory, the optimal income taxation framework of Mirrlees (1971) has lead to few fiscal policy prescriptions in all but one area: that of supplementary commodity taxation.

The main result concerning the role of commodity taxation when an optimally designed non-linear income tax schedule is available is the uniform tax prescription of Atkinson and Stiglitz (1976) - henceforth AS. It says that, if preferences are separable between leisure and the other goods, there is no need for taxing goods - the nonlinear tax schedule on income will do the job.

---

[*] A longer, more ambitious version of this paper circulated under the name 'Redistribution with Ex-ante Unobservable Choices'. I thank Luis Braido and participants in seminars at PUC-RJ and IBMEC for their helpful comments. The usual disclaimer applies.

[†] Address: Graduate School of Economics at the Getulio Vargas Foundation - Praia de Botafogo 190/1105. Rio de Janeiro/RJ-BRAZIL 22450-900; E-mail: cdacosta@fgv.br. Phone: 55-21-2559.5493

This result has been expressed in many different ways in Mirrlees (1976), Cooter (1978), Christiansen (1984) and Konishi (1995) and challenged in many others - e.g., Naito (1999), Saez (2001), Cremer et al. (2002), etc.

In this paper, we investigate AS in a dynamic - two periods - version of Stiglitz's (1982) two type representation of Mirrlees' (1971) optimal income taxation model.

Contrary to other papers that introduce dynamics in this framework - e.g. da Costa and Werning (2000), Golosov et al. (2003) -, we assume that tax rates on savings are linear and must be measurable with respect to the information set available at the point where investment is made.

This is not, as one might think, a technical curiosity. On one hand we incorporate to the model the undisputable fact that though governments can tax labor income and define net income as a function of gross income it cannot control how agents distribute expenditures through time.[1] On the other hand, we take the requirement that marginal tax rates on capital income be known at the moment an investment to be a reasonable restriction on the set of instruments available to the government.

As we shall see, this simple and compelling modification on tax instruments is enough to break down AS, which does not happen in the previous dynamic taxation models.

It is also worth noticing that, with the exception of Naito (1999), who considers a technology where labor inputs are not perfect substitutes, all papers generate a violation of AS with the introduction of another dimension of heterogeneity among agents. Here, however, heterogeneity is still unidimensional, in the form of differences in productivity.

More to the point, in a recent contribution to the literature, Cremer et al. (2001b) show how uniform taxation is usually not optimal, and how income effects become important, when another dimension of unobserved heterogeneity - namely, endowment of other goods - is present.[2] What is surprising in our result is that no other dimension of heterogeneity is introduced in the model, yet we generate tax prescriptions akin to theirs.

There is, however, a subtle way in which heterogeneity appears in our framework. As we hope to make clear, an agent off the equilibrium path have a different 'endowment' from a second period perspective - like in Cremer et al. (2001b) - for she has a different level of savings. This is enough to make their result arise in a model with essentially no other source of heterogeneity but the one found in Mirrlees (1971). The AS uniform taxation result breaks down in our setup and we show that homotheticity is needed for uniform taxation to obtain.

Another important set of results in our paper is related to the taxation of capital income. As already discussed, all papers that explore this issue in a Mirrlees' setup allow for taxes on income to depend on the realized type of each agent. With separability, an inverse Euler equation is found that implies that marginal utility of today's income is smaller than expected marginal utility of tomorrow's income.[3]

---

[1] We emphasize non-observability of savings, but other forms of smoothing consumption - like purchase of durables - render the control of consumption virtually infeasible.

[2] See their example in page 790, second paragraph.

One might interpret this inverse euler equation as a prescription of positive tax on capital income, if marginal tax rates were constant. However, Albanesi and Sleet (2003) show that expected marginal tax rates on capital income is zero - and, consequently, no revenue is raised with these taxes. The inverse Euler equation is a consequence of taxes that induce a positive correlation between after tax returns on asset and labor income.

In our paper, since the tax rate is constant across types, we are able to go unambiguously from the relationship between marginal utility of income today ant tomorrow to the sign of marginal tax rates.

Because off-equilibrium savings become important in defining incentive compatibility constraints - henceforth, IC constraints - and these are non-standard, one may not tell 'a priori' which ones will bind at the optimum. However, we offer some arguments for defending that some are more likely to bind at the optimum than others and call these the 'normal' cases. These are exactly the ones where the IC constraints bind in the usual direction - a high type announcing to be a low type.

We, then, show that, in 'normal' cases, if leisure is a normal good, savings ought to be taxed. This result, that goes along with the findings of Cremer and Gahvari (1995, 1999), da Costa and Werning (2001) and Golosov et al. (2003) is due to the fact that agents who 'intend to announce' falsely their types save more, under normality of leisure, then agents who intend to abide by the rules.

The remainder of this paper is organized as follows. The model economy is presented in section 2. Then, in section 2.1 the concept of equilibrium and the approach we adopt for tackling the problem is described. Optimal taxation is characterized in section 3, where Atkinson and Stiglitz (1976) uniform taxation result is discussed. Section 4 concludes. All results are proved in the appendices.

2 The Environment

The economy is populated by a continuum of 'ex-ante' identical agents who live for two periods and have preferences represented by

$$\upsilon\left(\mathbf{x}^0, l^0\right) + E[\upsilon(\tilde{\mathbf{x}}^1, \tilde{l}^1)], \tag{1}$$

where $\mathbf{x}^0, \mathbf{x}^1 \in \mathbb{R}^n$ are, respectively, first and second period consumptions and $l^0, l^1 \in \mathbb{R}$ are, respectively, first and second period labor supplies. We shall be using bold to represent vectors, and following the convention that prices are row and quantities are column vectors.

Uncertainty arises in this problem because at the first period agents do not know their 'adult' productivities: which we shall call their types. Once uncertainty is realized, preferences are the same irrespective of one's type; productivity is the only dimension of heterogeneity 'ex-post'. An agent of productivity

---

[3] In da Costa and Werning (2000) an expected marginal rate of substitution condition that generalizes the inverse Euler equation to the case of intertemporal non-separability is found.

w needs $l = Y/w$ hours to produce output Y. Hence, the higher the productivity, the more leisure one agent gets for the same output she produces. In first period, however, w is identical for all agents and normalized to 1.

Output can be transformed in any consumption good with a linear technology. Units are chosen in such a way that production prices are normalized to 1.

To make the problem as simple as possible we follow Stiglitz (1982) in considering only two possible types: H (for high productivity) and L (for low). We also assume that they are in equal proportion in the population for notational simplicity.

Finally, given our explicit intent to investigate Atkinson and Stiglitz's uniform taxation result, we take temporary utility to be separable between consumption and leisure/labor.

$$\upsilon\left(\mathbf{x}, l\right) \equiv u\left(\mathbf{x}\right) - \zeta\left(l\right),$$

with $u'\left(\cdot\right), \zeta'\left(\cdot\right), \zeta''\left(\cdot\right) > 0$ and $u''\left(\cdot\right) < 0$.

With this assumptions, (1) becomes

$$u\left(\mathbf{x}^0\right) - \zeta\left(l^0\right) + \frac{1}{2}\left[u\left(\mathbf{x}^H\right) - \zeta\left(l^H\right) + u\left(\mathbf{x}^L\right) - \zeta\left(l^L\right)\right]. \tag{2}$$

We let shocks be independent across agents so that we may take the 'ex-post' distribution of types to coincide with the 'ex-ante' one.

The economy is divided in two sectors: a production and a consumption sectors. The production sector transforms labor inputs lw in goods $\mathbf{x}$ with a linear technology. Output may also be transferred inter-temporally using a linear technology. That is, the production sector transforms one unit of consumption today in one unit of consumption tomorrow.

We assume throughout the paper that savings are positive in equilibrium. This assumption, which is immaterial for the conclusions we reach, is made to guarantee that consumption does not take place before production.[4]

Finally, there is a benevolent government who inhabits this economy and maximizes the agents' expected utilities.

Following the tradition founded by Mirrlees (1971), the set of instruments available for the government to pursue its goals is not imposed in an 'ad hoc' fashion but is derived from the informational structure of the problem. So, we continue the description of the environment with its informational structure.

First, as is standard in optimal taxation, we assume that once uncertainty is realized each agent's productivity is only observed by the agent herself. Though productivity is not directly observable output produced by each agent is observed by everyone, which makes it negotiable.

---

[4] One way to avoid introducing this assumption would be to model an overlapping generations economy. In this case, the young would trade with the 'middle-aged' (we don't have the elderly, yet.) and there would be no problem of this kind. However this would just create too much distraction from the main points we wish to emphasize.

Transactions between sectors are observable by the government but transactions within sectors are not. Labor is assumed to be traded only at a prohibitively high cost outside the production sector, however. Therefore, because labor can only be traded - at reasonable cost - with the production sector, and because these transactions are observed by the government, the use of a non-linear income tax schedule is possible.

As for the trade of goods - that including the trade of consumption across time - the assumption is that it can be done between agents - outside the production sector - at no transaction costs. This rules out the possibility of having the government directly control savings, and that of it defining a non-linear tax schedule for other goods.

2.1 The Direct Mechanism

To find the optimal tax schedule we proceed as usual and define a truthful direct mechanism. Given the optimal allocations, we invoke the taxation principle to map it into an optimal tax schedule.

The problem here is non-standard, however, because some choices are made before the revelation of types. Hence, we dedicate the next few pages to discuss the characterization of optimal allocations.

Before, however, we need to invest in some notation.

If we define s as savings and $\mathbf{p}$ as consumer prices for first period goods, we get the following (conditional) indirect utility function

$$v\left(\mathbf{p}, y-s\right) \equiv \left\{ \begin{array}{c} \max u\left(\mathbf{x}\right) \\ \text{s.t. } \mathbf{px} \leq y-s \end{array} \right. ,$$

with $\mathbf{x}\left(\mathbf{p}, y-s\right)$ as the corresponding (conditional) Marshallian demand.

Analogously, for second period prices, $\mathbf{q}$, and disposable income y^i

$$v\left(\mathbf{q}, y^i+s\right) \equiv \left\{ \begin{array}{c} \max u\left(\mathbf{x}\right) \\ \text{s.t. } \mathbf{qx} \leq y^i+s \end{array} \right. \qquad i = H, L.$$

with $\mathbf{x}\left(\mathbf{q}, y^i+s\right)$ $(i = H, L)$.

Notice that at different periods agents face different price sets. Later we shall come back to this discussion, since the most reasonable interpretation for this model should not allow for this distinction.

2.1.1 The Nature of the Game

The game played by the government and the agents is a Stackelberg game, where the Government, the Stackelberg leader, moves first by choosing (y, Y), a budget set $\left\{\left(y^L, Y^L\right), \left(y^H, Y^H\right)\right\}$, and tax rates $\tau^1 = (\mathbf{p}-\mathbf{1})$ and $\tau^2 = (\mathbf{q}-\mathbf{1})$.

The agent follows, with the next move, by deciding what and how much to consume (or, equivalently, how much to save), given (y, Y). This decision is made before nature defines the agent's type. Once uncertainty is realized the agent chooses her bundle from the budget set offered by the principal. That is the timing of the game is as follows

$$\begin{array}{c}\xrightarrow{\hspace{30em}}\\ \underbrace{\begin{array}{c}\text{gov}'\text{t defines}\\ \mathbf{p},\mathbf{q},\ \text{and}\\ \left\{\left(y^i,Y^i\right)\right\}_{i=H,L,0}\end{array}}_{t=-1}\ \underbrace{\begin{array}{c}\text{agents choose}\\ \mathbf{x}\ \text{and}\ s\\ \text{given}\ y,Y\end{array}}_{t=0}\ \underbrace{\begin{array}{c}\text{nature}\\ \text{chooses}\\ \text{types}\end{array}}_{t=0.5}\ \underbrace{\begin{array}{c}\text{agents}\\ \text{choose from}\\ \left\{\left(y^i,Y^i\right)\right\}_{i=H,L}\end{array}\ \begin{array}{c}\text{agents}\\ \text{choose}\ \mathbf{x}\\ \text{given}\ \left(y^i,Y^i\right)\end{array}}_{t=1}\end{array}$$

The solution for this problem is potentially very complex. The fact that savings affect preferences, means that whether IC constraints are violated or not depend on the level of savings. On the other hand, it is the expected marginal utility of income at the incentive compatible allocations that will determine the optimal level of savings, both on and off the equilibrium path.

The way we deal with this simultaneity is by first defining strategies as mappings from types to announcements $\sigma : \Omega \rightarrow \Omega$. We argue that, if an agent decides to adopt strategy σ, we need only to consider the incentive compatibility constraints for this strategy at the appropriate - that is, expected utility maximizing - level of savings.

A strategy is a rule that associates to each type a specific action. In this case, an announcement. That is, if we introduce a player, called nature, the strategy of an agent defines a response for each move by nature in the form of an announcement of one's type.

The choice of an optimal strategy is not done in isolation. When deciding her strategy, at the first period, the agent is also choosing a level of savings. In fact, because each strategy defines a strictly concave savings problem, we associate to each announcement a unique optimal level of savings.

Formally, let $\sigma^k(j)$ be the announcement made in strategy k if one realizes type j. There are, in this case, four possible strategies:

$$\begin{array}{rr}\sigma^1(H)=H; & \sigma^1(L)=L\\ \sigma^2(H)=L; & \sigma^2(L)=L\\ \sigma^3(H)=H; & \sigma^3(L)=H\end{array}$$

and

$$\sigma^4(H)=L; \quad \sigma^4(L)=H$$

We shall use a special notation σ^* for strategy σ^1, the truthful announcement. Since this is the strategy we want to induce, we shall be comparing it with all other strategies.

The point we emphasize here is that, when looking for implementable allocations, we need only to compare the expected payoff for pairs of announcement and savings - $\left(\sigma^k, s^k\right)$ for $k = 2, 3, 4$ - with the expected payoff for the pair (σ^*, s^*) we want to induce.

We let y^i and Y^i be, respectively, the income available and the output produced by agent of type i (for $i = H, L$) and y and Y the same variables in the first period of her life, that is, before she realizes her 'adult' productivity.

Under our separability assumption, labor supply plays no role, whatsoever in determining the optimal level of savings.

Thus, we define s^* by

$$s^* \equiv \arg\max_s \left\{ 2v\left(\mathbf{p}, y-s\right) + \textstyle\sum_{i=H,L} v\left(\mathbf{q}, y^i+s\right) \right\}, \tag{3}$$

for savings along the equilibrium path, and s^k by the analogous expressions at the allocations corresponding to the prescriptions of each strategy.

We shall also work with the simplified notation

$$\begin{aligned} v^* &\equiv v\left(\mathbf{p}, y-s^*\right) \\ v^*\left(i\right) &\equiv v\left(\mathbf{q}, y^i+s^*\right) \quad i=H,L. \end{aligned}$$

and

$$\begin{aligned} v^2 &\equiv v\left(\mathbf{p}, y-s^2\right) \\ v^2\left(L\right) &\equiv v\left(\mathbf{q}, y^L+s^2\right) \end{aligned}$$

At this point it should already be clear that it does not suffice to consider IC constraints at the equilibrium level of savings. We have to consider off-equilibrium savings choices because, though truthful announcement may be the optimal strategy at the equilibrium level of savings, there might be another level of savings that makes some other strategy's expected payoff higher than the equilibrium one.

2.1.2 The Redundancy of Second Period IC constraints

In principle, to set up the program to be solved by the government we should include not only all the IC constraints that guarantee that truthful announcement for all realizations is the chosen strategy but also the traditional second period constraints. However, as we will see, only two first period IC constraints need to be considered.

To get to this point we start by proving the redundancy of second period IC constraints. To do that, we shall first introduce some notation that will allow us to economize on space and improve aesthetically the paper.

Let

$$U\left(\mathbf{q}, I, Y, w\right) \equiv v\left(\mathbf{q}, I\right) - \zeta\left(Y/w\right).$$

Then, we define

$$U^k\left(i\right) \equiv U\left(\mathbf{q}, y^i+s^k, Y^i, w^i\right)$$

$$U^k\left(i|j\right) \equiv U\left(\mathbf{q}, y^i+s^k, Y^i, w^j\right)$$

$$U^k \equiv U\left(\mathbf{p}, y^i-s^k, Y, 1\right)$$

for $k = 1, 2, 3$, and 4, though we shall always use star instead of 1 for the equilibrium savings. It should also be clear in this case, that $T\left(Y\right) \equiv Y - y$.

What we show next is that, if the first period constraints are satisfied then, the second period ones, will also be.

In fact, consider the first period IC constraints:

$$\begin{aligned} 2U^{*}+U^{*}(H)+U^{*}(L) &\geq 2U^{2}+U^{2}(L|H)+U^{2}(L) \qquad (4)\\ 2U^{*}+U^{*}(H)+U^{*}(L) &\geq 2U^{3}+U^{3}(H)+U^{3}(H|L) \qquad (5)\\ 2U^{*}+U^{*}(H)+U^{*}(L) &\geq 2U^{4}+U^{4}(L|H)+U^{4}(H|L) \qquad (6)\end{aligned}$$

With the notation we have adopted the second period constraints can be written

$$\begin{aligned} U^{*}(H) &\geq U^{*}(L|H) \text{ and} \qquad (7)\\ U^{*}(L) &\geq U^{*}(H|L). \qquad (8)\end{aligned}$$

If one of these constraints is violated, there is a pair strategy/savings that is better than truthfully reporting one's type and choosing the corresponding maximizing level of savings. Hence, necessity.

As for sufficiency, because savings are optimally chosen in (4) the right hand side of this equation is greater than the same expression evaluated at any level of savings. In particular at $s=s^{*}$, that is,

$$\begin{aligned} 2U^{*}+U^{*}(H)+U^{*}(L) &> 2U^{*}+U^{*}(L|H)+U^{*}(L)\\ &\therefore\ U^{*}(H)>U^{*}(L|H)\end{aligned}$$

And constraint (7) is satisfied with strict inequality. By the same argument, (5) implies that

$$\begin{aligned} 2U^{*}+U^{*}(H)+U^{*}(L) &> 2U^{*}+U^{*}(H)+U^{*}(H|L)\\ &\therefore\ U^{*}(L)>U^{*}(H|L)\end{aligned}$$

Which shows that (8) is also satisfied strictly.

Hence, what we have shown is that, if the first period constraints are satisfied, so are the second period ones. Also it is readily seen that the constraints are satisfied with strict inequalities. As a consequence, it is apparent that any tax schedule will be interim inefficient, in the sense that once the saving decision is made, agents would want the government to redesign the tax schedule.

Conversely, considering the case of (4), if we required

$$U^{*}(H)=U^{*}(L|H) \qquad (9)$$

to be binding, we would have,

$$2U^{*}+U^{*}(H)+U^{*}(L)=2U^{*}+U^{*}(L)+U^{*}(L|H)$$

But then, generically,

$$2U^{*}+U^{*}(H)+U^{*}(L)<2U^{2}+U^{2}(L)+U^{2}(L|H).$$

Which means that the allocation would not be implementable. The agent would choose to save an amount s^2 (not s^*) and always pick the allocation intended for type L.

This results are akin to the ones found in the repeated moral hazard literature.[5] Once savings choices are made agents would be better off if the government could redesign the tax schedule using (9). The 'contract' - for that matter, any deterministic implementable contract - is not renegotiation-proof in the sense of Dewatripont (1988).

2.1.3 The Relevant First Period IC Constraints.

Before writing down the Lagrangian for this problem we introduce one more assumption that will guarantee that strategy 4 can be ruled out on the grounds that there are no allocations that justify adopting it and that still satisfy the other constraints.

We should also notice is that if we write $m(\cdot) \equiv -U_Y/U_y$, with subscripts denoting partial derivatives, it is easy to see that our assumption on preferences imply

$$\frac{\partial m(\cdot)}{\partial w} < 0, \text{ and } \frac{\partial m(\cdot)}{\partial s} < 0.$$

The first consequence is single-crossing (or Spence-Mirrlees condition)results from normality of consumption. The second one is due to normality of leisure and will turn out to be very important to many of our proofs.

One last piece of notation. Let

$$m^k(i|j) \equiv -\frac{U_Y\left(y^i + s^k, Y^i, w^j\right)}{U_y\left(y^i + s^k, Y^i, w^j\right)}$$

denote the marginal rate of substitution for an agent of type j who announces to be of type i, given that she follows strategy k. We shall also use $m^k(\cdot|j)$ for a generic allocation, for type j, at the level of savings corresponding to strategy k and $m^k(j) \equiv m^k(j|j)$.

Notice the dependence of $m(\cdot)$ on s. In fact, if leisure is normal, it is easy to see that $\partial m(\cdot)/\partial s < 0$. The slope of indifference curves in the $y \times Y$ space are affected by the choice of savings. In this sense, the ranking of bundles is not independent on a agents' first period choices.

Back to our problem, we are down to three first period IC constraints. What we show next is that, if there is a level of savings that, at the same time, makes strategy 4 optimal, and gives the same expected utility that the truthful announcement, then one of the other two constraints is violated.

For doing so, however, we need the following claim, proved in the appendix.

Claim 1 *Monotonicity is necessary for an allocation to be implementable*

---

[5] See Chiappori et Alli (1995), for example.

Assume that constraint (6) is binding at the optimal allocation. That is, there is a level of savings s^4 such that

$$2U^* + U^*(H) + U^*(L) = 2U^4 + U^4(H|L) + U^4(L|H).$$

Notice that, because this is the optimal strategy, it must be the case that

$$U^4(H|L) \geq U^4(L) \text{ and } U^4(L|H) \geq U^4(H).$$

From, (19) and (20) it is rather straightforward to see that this requires $\left(y^H, Y^H\right) < \left(y^L, Y^L\right)$. But in this case, claim 1 implies that the allocation is not implementable.

This shows that, whenever the first two constraints are satisfied, constraint (6) is satisfied as a strict inequality. Hence, we may always leave it aside.

The maximization problem, thus, includes IC constraints (4) and (5). Anyone used to dealing with this type of model would be tempted to rule out constraint (5), and maintain work only with the most appealing case, where one is only to be concerned with the high type pretending to be a low type. In fact, were we in the usual Mirrlees' setup and single crossing would be invoked to show that constraint (5) is slack. Unfortunately the structure imposed so far does not allow for ruling out the possibility that (5) binds at the optimum.

This being said, we could go on and derive results for the case when both constraints bind. However, most expressions would be very cumbersome, and ambiguity would emerge in the case of policy recommendation regarding capital income tax. We shall, then, deal with the case we view as the most likely to emerge for most reasonable specifications of our model.

In this case, we owe the reader an explanation as of why we think that the case where only constraint (4) binds at the optimum is the most likely to emerge.

The argument goes as follows. As we have seen, what makes the 'renegotiation-proof' solution unfeasible is the fact that, on the one hand, there is no loss in utility from moving from the allocation intended to a high type to the allocation intended to the low type, given that one turns out to be of a high type, while on the other, there are gains from re-optimizing savings from the choice of this different strategy.

Let us now concentrate in the no intervention regime - i.e., assuming that there is no government. We know in this case that, then different realization of productivities lead to different choices of Y (which, in this case, is identical to y.) Standard revealed preferences arguments may then be used to show that the gains from re-optimization are smaller then the losses of a high type who chooses the equilibrium Y of a low type and those of a low type who chooses the equilibrium Y of a high type.

Because increasing expected utility requires transferring resources from the high type to the low type, the loss of utility of a high type picking the allocation of a low type tends to decrease, while that of a low type picking the allocation intended to the high type tends to increase.

If IC constraint (4) binds at the optimum, the gains from re-optimization exactly compensate for the loss of the type of deviation prescribed by this strategy. As for the low type, assuming that the loss in utility from picking the wrong allocation has really increased, it is very unlikely that the gains of optimization will compensate for it given that they did not in the no intervention regime. Even more so if the difference in marginal utility of income is reduced with intervention.

This is not a proof, of course. It just gives us reasons to believe that we need not worry about strategy 3 for most reasonable specifications of the model. Therefore, from now on, we shall deal with the case where only constraint (4) binds at the optimum but we shall remark the results that depend on this assumption.

3 Optimal Taxation

Bearing the all this notation in mind, we are finally in a position to spell out the government's program in a (almost) standard way.

$$\max_{\mathbf{p},\mathbf{q},\{(y^i,Y^i)\}_{i=H,L,0}} 2v^* - \zeta(Y) + \sum_{i=H,L} \left[v^*(i) - \zeta\left(Y^i/w^i\right)\right]$$

subject to

$$2v^* + \sum_{i=H,L} \left[v^*(i) - \zeta\left(Y^i/w^i\right)\right] \geq 2v^2 + 2v^2(L) - \sum_{i=H,L} \zeta\left(\frac{Y^L}{w^i}\right), \qquad [\,\mu\,]$$

and

$$2\left[(Y-y) + (\mathbf{p}-\mathbf{1})\,\mathbf{x}^*\right] + \sum_{i=H,L} \left[\left(Y^i - y^i\right) + (\mathbf{q}-\mathbf{1})\,\mathbf{x}^*(i)\right] \geq 0 \qquad [\,\lambda\,]$$

where $\mathbf{x}^*$, $\mathbf{x}^*(H)$, $\mathbf{x}^*(L)$, denote Marshallian demands. Notice also the Lagrange multipliers inside brackets at the right of each constraint.

There are some important differences here with respect to a standard optimal taxation problem. First, there is an extra term in the utility function which is the first period utility. Second, the s term that appears not only in the objective function but also in the IC constraints. Most important, however, is the fact that the IC constraints are not there to guarantee that the agent chooses a certain *action*, but that she chooses a certain *strategy* (and corresponding savings).

One might also have noticed that we allow for taxes to differ across ages - i.e., $\mathbf{p} \neq \mathbf{q}$. If one thinks of this model - as we do - as describing a situation where two generations inhabit the economy at the same time, this is not a reasonable specification, given our assumptions on the underlying informational structure of the economy. We adopt this assumption because it provides cleaner formulae and facilitates the discussion, at the same time that it does not alter the conclusions regarding AS or optimal capital income taxes.

Nonetheless, for the sake of completeness we derive in appendix B the optimal tax formulae for the case in which prices are obliged to be identical in both periods. Conclusions there are identical to the ones herein, though tax formulae need not coincide for the general case, as expected.

Finally, it is worth remarking that first period labor supply does not appear in the IC constraint. This does not mean that labor supply is not distorted in the first period. In fact it is easy to verify that, at the optimum, $v_y(\mathbf{p}, y) \neq \zeta'(Y)$.

The first step towards evaluating AS is the derivation of optimal commodity tax formulae. We recall the fact that we are assuming separability.

This been said, we start by writing the Lagrangian,

$$\begin{aligned}\mathcal{L} &= 2U^* + U^*(H) + U^*(L) - \lambda\Big[2(y - Y) + \textstyle\sum_{i=H,L}(y^i - Y^i)\Big] \\ &\quad + \mu^2\left[[2U^* + U^*(H) + U^*(L)] - \left[2U^2 + U^2(L) + U^2(L|H)\right]\right].\end{aligned}$$

The first order condition with respect to q^i $(\partial\mathcal{L}/\partial q^i = 0)$ is, then:

$$\begin{aligned}&\left[v_y^*(H)\,x^{*i}(H) + v_y^*(L)\,x^{*i}(H)\right](\mathbf{1} - \mu) - \\ &2\mu v_y^2(L)\,x^{2i}(L) + \lambda\big\{2(\mathbf{1} - \mathbf{p})\,\mathbf{x}_y^* s_i + x^{*i}(H) + x^{*i}(L) + \\ &(\mathbf{q} - \mathbf{1})\left[\mathbf{x}_i^*(H) + \mathbf{x}_i^*(L) + \left(\mathbf{x}_y^*(H) + \mathbf{x}_y^*(L)\right)s_i\right]\big\} = 0 \qquad (10)\end{aligned}$$

where we used Roy's identity to obtain $v_i^*(k) = v_y^*(k)\,x^{*i}(k)$ for $k = H, L$.

Now, we write the first order conditions with respect to y^H, i.e., $\partial\mathcal{L}/\partial y^H = 0$:

$$\begin{aligned}&v_y^*(H)(1 + \mu) + \lambda\Big[-1 - 2(\mathbf{p} - \mathbf{1})\,\mathbf{x}_y^* s_{y^H}^* + \\ &(\mathbf{q} - \mathbf{1})\Big(\left[\mathbf{x}_y^*(H) + \mathbf{x}_y^*(L)\right]s_{y^H}^* + \mathbf{x}_y^*(H)\Big) = 0, \qquad (11)\end{aligned}$$

and y^L, i.e. $\partial\mathcal{L}/\partial y^L = 0$:

$$\begin{aligned}&v_y^*(L)(1 + \mu) - \mu 2 v_y^2(L) + \lambda\Big[-1 - 2(\mathbf{p} - \mathbf{1})\,\mathbf{x}_y^* s_{y^L}^* + \\ &(\mathbf{q} - \mathbf{1})\Big(\left[\mathbf{x}_y^*(H) + \mathbf{x}_y^*(L)\right]s_{y^L}^* + \mathbf{x}_y^*(L)\Big) = 0. \qquad (12)\end{aligned}$$

Next, we multiply (11) by the vector $\mathbf{x}_i^*(H)$ and (12) by $\mathbf{x}_i^*(L)$ add the two, and replace the expressions we obtain in (10). This yields

$$\begin{aligned}\frac{2\mu v_y^2(L)}{\lambda}\left[x^{*i}(L) - x^{2i}(L)\right] &= -2\hat{s}_i[(\mathbf{p} - \mathbf{1})\,\mathbf{x}_y^* + \\ &(\mathbf{q} - \mathbf{1})\left(\mathbf{x}_y^*(H) + \mathbf{x}_y^*(L)\right)] - (\mathbf{q} - \mathbf{1})\left[\hat{\mathbf{x}}_i^*(H) + \hat{\mathbf{x}}_i^*(L)\right], \qquad (13)\end{aligned}$$

where $\hat{s}_i \equiv s_i - \left(s_{y^H} x^{*i}(H) + s_{y^L} x^{*i}(L)\right)$, $\hat{\mathbf{x}}_i^*(H) \equiv \mathbf{x}_i^*(H) - \mathbf{x}_y^*(H)\,x^{*i}(H)$, and $x^{2i}(L)$ is the marshallian demand for good i at level of savings s^2 by agent of type L.[6]

The discouragement of consumption of good i has two components: the direct effect on compensated demands in each state of the world,[7] and the indirect effect, through savings.

---

[6] In fact, as we shall see, this represents the (identical) demand of both a high type mimicking a low type and of a low type at off-equilibrium level of savings.

[7] To see that this is in fact the discouragement of consumption of good i just notice that the vector $\mathbf{x}_i^*(H) - \mathbf{x}_y^*(H)\,x^{i*}(H)$ is the derivative of the compensated (holding state utility constant) demands for all goods with respect to price q^i, which is equivalent to the gradient of compensated demand for good i.

As for first period, a similar procedure is used. We start by deriving the first order condition with respect to p^j,

$$\begin{aligned}&2v_y^* x^{*j}(1+\mu) - 2v^2 x^{2j} + \lambda \left\{ 2x^{*j} + 2(\mathbf{p}-\mathbf{1})\,\mathbf{x}_j^* - \right. \\ &\left. \left[2(\mathbf{p}-\mathbf{1})\,\mathbf{x}_y^* - (\mathbf{q}-\mathbf{1})\left(\mathbf{x}_y^*(H) + \mathbf{x}_y^*(L) \right) \right] \right\} s_j^* = 0\end{aligned}$$

where, once again, we used the facts that $v_j^* = -v_y^* x^{*j}$ and $v^2 = -v^2 x^{2j}$.

From the agent's first order condition we know that $2v_y^* = v_y^*(H) + v_y^*(L)$ and $v^2 = v^2(L)$. Therefore, we may write the above expression as

$$\begin{aligned}&2\mu v_y^2 \left[x^{*j} - x^{2j} \right] + \lambda 2(\mathbf{p}-\mathbf{1})\,\mathbf{x}_y^* \left[s_j - \left(s_{y^H} + s_{y^L} \right) x^{*j} \right] + \\ &\lambda(\mathbf{q}-\mathbf{1})\left(\mathbf{x}_y^*(H) + \mathbf{x}_y^*(L) \right) \left[s_j - \left(s_{y^H} + s_{y^L} \right) x^{*j} \right] + \\ &2\lambda(\mathbf{p}-\mathbf{1})\left(\mathbf{x}_j^* - \mathbf{x}_y^* x^{*j} \right) = 0\end{aligned}$$

Using the analogous definitions, for $\hat{s}_j, x^{2j}$ and $\hat{\mathbf{x}}_j^*$, this simplifies to

$$\frac{\mu v_y^2}{\lambda}\left(x^{2j} - x^{*j} \right) = \left[(\mathbf{p}-\mathbf{1})\,\mathbf{x}_y^* + (\mathbf{q}-\mathbf{1}) \frac{\mathbf{x}_y^*(H) + \mathbf{x}_y^*(L)}{2} \right] \hat{s}_j^* + (\mathbf{p}-\mathbf{1})\,\hat{\mathbf{x}}_j^*. \quad (14)$$

Now it is the deviation in the consumption of x^j in the first period that determines the discouragement of this good. The intuition is identical.

From now on, we search for conditions that will deliver AS. The strategy is to suppose that it holds, that is that prices are uniform, and verify what is needed for the derived expressions to be satisfied.

We shall consider two levels of AS. First we explore the possibility of $\mathbf{p} = \mathbf{q}$. If this is the case, we may write $\mathbf{p} = \mathbf{q} = \mathbf{1}$. The right hand side of expressions (13) and (14) vanish and they become, respectively

$$\frac{2\mu v_y^2(L)}{\lambda}\left[x^{*i}(L) - x^{2i}(L) \right] = 0 \ \forall i, \text{ and}$$

$$\frac{\mu v_y^2}{\lambda}\left(x^{2j} - x^{*j} \right) = 0 \ \forall j.$$

Adding across goods, the conditions are equivalent to $s^* = s^2$, which we have seen to be generically untrue. This is easy to understand. For the consumption choice not to give any information as of an agent's true type, we need both $x^{*i}(L) - x^{2i}(L)$ and $x^{2j} - x^{*j}$ to be 0 for all goods. That is the demand for all goods must vary by the same *amount* with $s^* - s^2$. This is only satisfied in the case where $s^* = s^2$.

It is well known that our assumption on preferences guarantees the optimality of uniform taxation in Atkinson and Stiglitz (1976). The same is true for the setup of all papers with dynamics added to which we previously referred.

However, when dynamics is added to the problem, it is no longer the case that prices have to be identical across periods. All papers, so far have arrived at the weaker form of AS. So, let us see it this also obtains in our setup.

So let us consider a weaker form of AS. Namely, we shall require uniform taxation of goods only within each period, i.e., $\mathbf{q} = \mathbf{1}$ but $\mathbf{p} = \rho\mathbf{1}$. This means that we will be able to implement the optimal tax scheme with uniform taxation on goods and linear taxation of capital income.

Then, equation (13) becomes

$$\frac{\mu v_y^2(L)}{\lambda}\left[x^{*i}(L)-x^{2i}(L)\right]=\mathbf{1}\mathbf{x}_y^{*}\hat{s}_i(\rho-1) \tag{15}$$

while (14) becomes

$$\frac{\mu v_y^2}{\lambda}\left[x^{*j}-x^{2j}\right]=-\mathbf{1}\mathbf{x}_y^{*}\hat{s}_j(\rho-1) \tag{16}$$

In this case, we get proposition 2, which is, in some sense, the most important result in the paper, in that it shows that separability alone is not sufficient to deliver the uniform tax prescription of Atkinson and Stiglitz.

Proposition 2 *There is a ρ that satisfies equations (15) and (22) only if*

$$\frac{x^{*i}(L)-x^{2i}(L)}{x_y^{*i}(H)+x_y^{*i}(L)}, \tag{17}$$

and

$$\frac{x^{*j}-x^{2j}}{x_y^{*i}(H)+x_y^{*i}(L)} \tag{18}$$

are constant across goods.

The most important thing to realize is that the condition in Proposition 2 is *in addition* to separability. Were we in a traditional Mirrlees' setup and only separability would be needed. However, preferences must be such that marshallian demands satisfy constancy across goods of (17) and (18).

If elasticities of demand differ across goods this expression will also differ, in general.[8] When an agent decides that she will announce a low type no matter what, she increases her savings - see claim 4.

When uncertainty is finally revealed, she will announce L, but will have more income than a type L who saved the amount compatible with a truthful announcement strategy. If income effects differ across goods, the pattern of consumption will be altered by her savings decision, and will signal deviant behavior.

One might be wondering why, it does not suffice the fact that if she has more savings her consumption will necessarily increase for some goods and this will signal her lying. The point here is that capital income taxation takes care of this part of the effects of deviant behavior.

---

[8] An example where this is necessarily violated is when a subset of 3 goods have constant but different income elasticities of demand, over the relevant range of income. This is but one example where it is easy to prove the result. The important point is that the non-constancy across goods of (17) and (18) is the rule rather than the exception.

It is also interesting to compare our results to the ones in Cremer et al. (2001b, p. 709). In that case it is assumed that agents have different endowments of a certain good k. This is not observed by the government. However, because agents that have higher endowment are richer, they increase the consumption of other normal goods when compared to agents that have a lower endowment. This helps the government in identifying deviant behavior.

In our case, the 'higher endowment' only appears off the equilibrium path. Yet, it generates the same type of prescriptions that arise in their model.

Next, we show that homotheticity restores uniform taxation as the optimal prescription.

If preferences are homothetic, $x^i = \omega^i y$, (where ω^i is a constant share) and $x^i_y = \omega^i$. That is, homotheticity implies that $x^{*i}(L)\,p^i = \omega^i\left(y^L + s^*\right)$ and $x^{2i}(L)\,p^i = \omega^i\left(y^L + s^2\right)$. Hence,

$$\left(x^{*i}(L) - x^{2i}(L)\right) = \omega^i\left(s^* - s^2\right)/p^i.$$

Similarly,

$$x^{*i}_y(H) + x^{*i}_y(L) = \omega^i\left(y^H + y^L + 2s^*\right)/p^i$$

Expression (17), thus, becomes

$$\frac{s^* - s^2}{y^H + y^L + 2s^*}.$$

By the same token, assuming that preferences are homothetic, it is simple to see from (21) that expression (18) becomes

$$\frac{s^2 - s^*}{y^H + y^L + 2s^*}$$

which is also independent of j.

Notice also that, in the homothetic case, (15) and (16) deliver

$$\frac{\mu v^2_y(L)}{\lambda}\frac{s^* - s^2}{y^H + y^L + 2s^*} = \frac{\mu v^2_y}{\lambda}\frac{s^* - s^2}{y^H + y^L + 2s^*} = \rho - 1,$$

which may be used to prove the following.

Corollary 3 *If $s^* < s^2$ (resp. $>$) first period consumption ought to be subsidized (resp. taxed) with respect to second period.*

To understand what it means notice that this is equivalent to taxing the return on capital income. In fact, $\rho = 1 + r$, where r is the riskless rate of return on capital income. But the marginal rate of transformation of consumption today in consumption tomorrow is 1. Therefore, r is the subsidy on future consumption.

Notice also that, even though, we are working in a Mirrlees setup, uniform taxation is only optimal if the conditions for the result in a Ramsey problem are present.

Finally, we shall use the following fact to prove our main proposition.

Claim 4 *If leisure is normal, then* $s^2 > s^*$.

Putting together corollary 3 and claim 4, we get proposition 5.

Proposition 5 *If preferences are separable and homothetic:*
i) goods are not taxed, and;
ii) capital income is taxed.

Cremer et al. also show that AS breaks down in a model where agents have different (non-observed) endowments. This extra dimension of heterogeneity is not needed in our model. In fact agents are heterogeneous here in only one dimension, as in Atkinson and Stiglitz (1976) and Mirrlees (1976). It must also be said that it is not the linearity restriction on commodity taxation that drives the result, since uniform taxes *are linear*.

There is however one sense in which one may argue that we have introduced another dimension of heterogeneity. Along the equilibrium path agents only differ in their productivities. However, off-equilibrium agents add another dimension of heterogeneity in their second period 'endowments' very much like in Cremer et al. (2001). This might explain why we get about the same prescriptions in a model where, unlike in their case, no extra dimension of heterogeneity was ever introduced.

4 Conclusion

In this paper, the simultaneous appearance of three issues, never considered in a same optimal taxation framework, to the best of our knowledge, results in the problem being quite non-standard. First, agents do not have all the information at the time when some of the decisions are made. Second, savings are not observable, so we are restricted to anonymous taxes on capital income as in Cremer et al. (2001a). Unlike them, however, we add the third aspect we referred to. We let savings affect 'ex-post' preferences.

These latter two elements: non-observability of first period choices and choices affecting preferences, render the problem a much harder one to characterize, since the set of implementable allocations in the second period becomes endogenous.

This is simple to understand. Once preferences are changed, the ranking of two bundles may be inverted, and what was incentive compatible for one specific 'ex-ante' choice may not to be for another. Seen from a different perspective, the possibility of anticipating the pattern of announcements, conditional on realized types, by agents, will make off-equilibrium (from the perspective of second period choices) savings crucial in defining the set of feasible allocations.

Implementation requires a different set of incentive constraints, which we impose in order to derive policy prescriptions. We show that this form of non-observability generates a violation of Atkinson and Stiglitz' (1976) uniform taxation prescription. Homotheticity is also needed for the result to be valid. A similar result is obtained in Cremer et al. (2001), where another dimension of heterogeneity, in the form of unknown endowments is introduced.

What is surprising here is that we do not have another dimension of heterogeneity. At least not along the equilibrium path. The subtlety here is that agents *do* differ in another dimension - consumption endowments, as in Cremer et al. (2001) - from their off-equilibrium counterparts. This is enough for us to obtain the type of prescriptions they generate in a model with exogenous heterogeneity.

When capital income taxation is investigated sharp results are harder to come about. The problem here is that the assumptions usually adopted in the optimal taxation literature do not allow for determining which constraints bind at the optimum in this model. Still, in the 'normal cases', taxation of savings (or subsidization of early consumption) is shown to be optimal.

From a purely theoretical perspective, our results are akin to the ones found in Golosov et al. (2003) and da Costa and Werning (2000), though we consider linear taxes on capital return, only.

The model is admittedly simplistic and, though simplicity is important for highlighting the key issues at stake, it would be very interesting to explore in further detail the robustness of some of the results we found. However, this is no easy task, considering the non-trivial conceptual problems that arise in such framework.

A Proofs

Proof of Claim 1. At the equilibrium level of savings, define the following set, for each allocation (y, Y),

$$\mathbb{Z}_+^H(y, Y) \equiv \left\{(y', Y') \in R_+^2; (x', y') \succsim_H (y, Y)\right\}$$

That is the set of bundles preferred to (y, Y) by agent type H.
Similarly,

$$\mathbb{Z}_+^L(y, Y) \equiv \left\{(y', Y') \in R_+^2; (x', y') \succsim_L (y, Y)\right\}$$

Define also

$$\mathbb{Z}_+(y, Y) \equiv \left\{(y', Y') \in R_+^2; (x', y') > (y, Y)\right\},$$

the set of bundles for which both quantities are at least as great as (y, Y), with at least one strictly greater. We can see that

$$\mathbb{Z}_+^L(y, Y) \cap \mathbb{Z}_+(y, Y) \subset \mathbb{Z}_+^H(y, Y) \cap \mathbb{Z}_+(y, Y) \tag{19}$$

In fact, without loss, take a path starting at (y, Y)such that

$$\begin{aligned} dV(H) &= V_y(\cdot|H)\left[dy - m(\cdot|H)\,dY\right] \geq 0 \Rightarrow \\ dy &= m(\cdot|H)\,dY \end{aligned}$$

In this case,

$$dV(L) = V_y(\cdot|L)\left[m(\cdot|H) - m(\cdot|L)\right] dY < 0$$

for $dY > 0$. If we define $\mathbb{Z}_-(y,Y) \equiv \left\{(y',Y') \in R_+^2; (x',y') < (y,Y)\right\}$, analogously, the same procedure shows that,

$$\mathbb{Z}_+^H(y,Y) \cap \mathbb{Z}_-(y,Y) \subset \mathbb{Z}_+^L(y,Y) \cap \mathbb{Z}_-(y,Y) \tag{20}$$

Now take a pair of allocations $\left(y^L, Y^L\right)$ and $\left(y^H, Y^H\right)$ such that $\left(y^H, Y^H\right) < \left(y^L, Y^L\right)$. From, (19) and (20), either $\left(y^L, Y^L\right) \in \mathbb{Z}_+^H(y,Y)$, or $\left(y^H, Y^H\right) \in \mathbb{Z}_+^H(y,Y)$. In either case, the allocation is not implementable. ■

Proof of Claim 4. Let

$$\mathbb{Z}_+^{H^*}\left(y^H, Y^H\right) \equiv \left\{(y',Y') \in R_+^2; (x',y') \succsim_{H^*} \left(y^H, Y^H\right)\right\},$$

at the equilibrium level of savings, s^*, and

$$\mathbb{Z}_+^{H^2}\left(y^H, Y^H\right) \equiv \left\{(y',Y') \in R_+^2; (x',y') \succsim_{H^2} \left(y^H, Y^H\right)\right\},$$

at the optimal level of savings for strategy 2, s^2.

Because the allocation is implementable we know form claim 1 that $\left(y^L, Y^L\right) \in \mathbb{Z}_-\left(y^H, Y^H\right)$ and $\left(y^L, Y^L\right) \notin \mathbb{Z}_+^{H^*}\left(y^H, Y^H\right)$. On the other hand, from the definition of strategy 2 it must be the case that $\left(y^L, Y^L\right) \in \mathbb{Z}_+^{H^2}\left(y^H, Y^H\right)$. However, if $s^2 < s^*$ normality of leisure implies that

$$\mathbb{Z}_+^{H^2}\left(y^H, Y^H\right) \cap \mathbb{Z}_-\left(y^H, Y^H\right) \subset \mathbb{Z}_+^{H^*}\left(y^H, Y^H\right) \cap \mathbb{Z}_-\left(y^H, Y^H\right).$$

A contradiction. ■

Proof of Proposition 2. First notice that

$$\kappa s_i = v_{yy}^*(H)\, x^i(H) + v_y^*(H)\, x^i(H) + v_{yy}^*(L)\, x^i(L) + v_y^*(L)\, x^i(L)$$

where κ is a negative constant that corresponds to the second order condition for the agent's savings problem. Also notice that

$$\kappa s_{y^H} = -V_{yy}(H) \quad \text{and} \quad \kappa s_{y^l} = -V_{yy}(L)$$

Hence, from the definition of $\hat{s}_i$ we have that

$$\kappa \hat{s}_i = -2 v_y^* \left[x_y^i(H) + x_y^i(L)\right]$$

But, then,

$$\rho - 1 = \Phi \frac{x^{*i}(L) - x^{2i}(L)}{x_y^i(H) + x_y^i(L)} \quad \forall i \tag{21}$$

with

$$\Phi = -\kappa \frac{v_y^2 \mu}{v_y^* \lambda},$$

where we also used the fact that, at $\mathbf{q} = \mathbf{1}$, $\mathbf{1x}_y^* = 1$.

Now just note that (21) can only be satisfied for all i if the right hand side is independent of i.

Analogously, for the case of first period goods, Considering our previous discussion, we want to find conditions under which $\mathbf{q} = \mathbf{1}$ and $\mathbf{p} = \rho \mathbf{1}$. Because the last term is a compensated demand, it vanishes at $\mathbf{p} = \rho \mathbf{1}$. Then, the expression becomes

$$-\frac{2\mu v_y^2}{\lambda} \frac{x^{*j} - x^{2j}}{\hat{s}_j} = \rho - 1. \tag{22}$$

Whence, we need the left hand side to be independent of j. ■

B Identical Prices in the Two Periods

In this appendix we consider the case where prices cannot be made time dependent. That is an important restriction, if we want to think of this framework as a reduced form of an overlapping generations model.

Before writing the program a word is needed regarding the possibility of taxing capital income.

We have been using the unity as the production price of future consumption in terms of today's consumption as a normalization. If the overlapping generations - OLG - interpretation were to be taken seriously, then, taxing capital income would not be feasible. The point is that under the assumptions used so far, the young and the elderly would transfer resources, and no aggregate savings would be observed in equilibrium.

We sidestep this discussion - which in its most general form hinges only in the specific normalization we have adopted - and write the government's program as

$$\begin{aligned} &\max_{\mathbf{p}, R, \{(Y^i, y^i)\}_{i=H,L,0}} && 2U^* + U^*(H) + U^*(L) && \\ &\text{s.t. } 2U^* + U^*(H) + U^*(L) \geq 2U^2 + U^2(L|H) + U^2(L) && && [\ \mu\ ] \\ &\text{s.t. } 2Y + Y^H + Y^L \geq 2\textstyle\sum_k x^k + \sum_k \sum_{i=H,L} x^k(i), && && [\ \lambda\ ] \end{aligned}$$

where R is the (inverse of) consumer price of future income in terms of today's consumption, i.e. gross interest rates.

As done in section 3, we start by differentiating the Lagrangian with respect to p_j and setting it equal to 0.

$$\begin{aligned} &\left[2U_j^* + U_j^*(H) + U_j^*(L)\right](1+\mu) - \mu\left[2U_j^2 + U_j^2(L|H) + U_j^2(L)\right] - \\ &\lambda \textstyle\sum_k \left[x_j^k + \sum_{i=H,L} x_j^k(i) + \left(-x_y^k + \sum_{i=H,L} R x_y^k(i)\right) s_j\right] = 0. \end{aligned} \tag{23}$$

To interpret this result, we rewrite the left hand side of this expression by using the (conditional) Roy identity, $U_j = U_y x^j$.

That is, we note that

$$\begin{aligned}\left[2U_j^* + U_j^*(H) + U_j^*(L)\right](1+\mu) - \mu\left[2U_j^2 + U_j^2(L|H) + U_j^2(L)\right] =& \\ (1+\mu)\left[2U_y^* x^{*j} + U_y^*(H)\,x^{*j}(H) + U_y^*(L)\,x^{*j}(L)\right] -& \\ \mu\left[2U_y^2 x^{2j} + U_y^2(L|H)\,x^{2j}(L|H) + U_y^2(L)\,x^{2j}(L)\right]& \qquad (24)\end{aligned}$$

Now, consider the first order conditions, with respect to y,

$$2U_y^*(1+\mu) - \mu 2U_y^2 = \lambda \textstyle\sum_k \left[x_y^k(1-s_y) + R\sum_{i=H,L} x_y^k(i)\,s_y\right], \qquad (25)$$

with respect to y^H,

$$2U_y^*(H)(1+\mu) = \lambda \textstyle\sum_k \left[x_y^k s_{y^H} + x_y^k(H)\left(1+Rs_{y^H}\right) + Rx_y^k(L)\,s_{y^H}\right], \qquad (26)$$

and with respect to y^H,

$$\begin{aligned}2U_y^*(L)(1+\mu) - \mu\left[U_y^2(L|H) + U_y^2(L)\right] =& \\ \lambda \textstyle\sum_k \left[x_y^k s_{y^L} + x_y^k(H)\,s_{y^L} + Rx_y^k(L)\left(1+Rs_{y^L}\right)\right].& \qquad (27)\end{aligned}$$

Multiply (25) by x^{*j}, (26) by $x^{*j}(H)$ and (27) by $x^{*j}(L)$ to get, respectively,

$$2x^{*j}\left[U_y^*(1+\mu) - \mu U_y^2\right] = \textstyle\sum_k x_y^k x^{*j} + \lambda\sum_k \left[-x_y^{*k} + R\sum_{i=H,L} x_y^{*k}(i)\right] s_y^* x^{*j},$$

$$\begin{aligned}x^{*j}(H)\,U_y^*(H)(1+\mu) &= \lambda \textstyle\sum_k x_y^k(H)\,x^j(H) + \\ &\quad \lambda \textstyle\sum_k \left[-x_y^k + R\sum_{i=H,L} x_y^k(i)\right] s_{y^H} x^j(H),\end{aligned}$$

and

$$\begin{aligned}x^{*j}(L)\left[2U_y^*(L)(1+\mu) - \mu\left[U_y^2(L|H) + U_y^2(L)\right]\right] =& \\ \lambda \textstyle\sum_k x_y^k(L)\,x^{*j}(L) + \lambda\sum_k \left[-x_y^k + R\sum_{i=H,L} x_y^k(i)\right] s_{y^L} x^{*j}(L),&\end{aligned}$$

Now, add the three to obtain

$$\begin{aligned}&(1+\mu)\left[2U_y^* x^{*j} + U_y^*(H)\,x^{*j}(H) + U_y^*(L)\,x^{*j}(L)\right] - \\ &\mu\left[2U_y^2 x^{*j} + \left(U_y^2(L|H) + U_y^2(L)\right)x^{*j}(L)\right] = \\ &\quad \lambda\left(s_y x^j + s_{y^H} x^j(H) + s_{y^L} x^j(L)\right)\textstyle\sum_k \left[-x_y^k + R\sum_{i=H,L} x_y^k(i)\right] + \\ &\quad \lambda \textstyle\sum_k \left(x_y^k x^j + \sum_{i=H,L} x_y^k(i)\,x^j(i)\right) \qquad (28)\end{aligned}$$

We can now substitute (28) in (23), using also (24) to get

$$\begin{aligned}&\mu[2U_y^2\left(x^{2j} - x^{*j}\right) + U_y^2(L|H)\left(x^{2j}(L|H) - x^{*j}(L)\right) + \\ &U_y^2(L)\left(x^{2j}(L) - x^{*j}(L)\right) = \lambda \textstyle\sum_k \left(h_j^k + \sum_{i=H,L} h^k(i)\right) \\ &\qquad + \lambda \hat{s}_j \textstyle\sum_k \left[-x_y^k + R\sum_{i=H,L} x_y^k(i)\right] \qquad (29)\end{aligned}$$

where $\hat{s}_j = s_j - \left(s_y x^j + s_{y^H} x^j(H) + s_{y^L} x^j(L)\right)$, $h^k_j = x^k_j - x^k_y x^j$, and $h^k_j(i) = x^k_j(i) - x^k_y(i) x^j(i)$ (for $i = H, L$.).

Finally we use the fact that separability implies $U^2_y(L|H) = U^2_y(L)$ to write (29) as

$$\begin{aligned}
&2\mu\left[U^2_y\left(x^{2j} - x^{*j}\right) + U^2_y(L)\left(x^{2j}(L) - x^{*j}(L)\right)\right] = \\
&\lambda \textstyle\sum_k \left(h^k_j + \sum_{i=H,L} h^k(i)\right) + \lambda \sum_k \left[-x^k_y + R\sum_{i=H,L} x^k_y(i)\right]\hat{s}_j.
\end{aligned}$$

Let us consider the expression above. The right hand side is just the discouragement of consumption of good j decomposed in its two parts - the direct effect and the indirect effect through savings, exactly like in (13) and (14). As for the right hand side, it is the change in consumption of good j, both in youth and in adulthood as caused by the choice of an optimal off-equilibrium level of savings.

To investigate AS, assume that $\mathbf{p} = \mathbf{1}$, then

$$\begin{aligned}
&2\mu\left[U^2_y\left(x^{2j} - x^{*j}\right) + U^2_y(L)\left(x^{2j}(L) - x^{*j}(L)\right)\right] = \\
&\lambda \textstyle\sum_k \left(h^k_j + \sum_{i=H,L} h^k(i)\right) + \lambda \sum_k \left[-x^k_y + R\sum_{i=H,L} x^k_y(i)\right]\hat{s}_j
\end{aligned} \quad (30)$$

Then, because $\sum_k x^k_y = \sum_k x^k_y(H) = \sum_k x^k_y(L) = 1$, we have that

$$2\mu\left[U^2_y\left(x^{2j} - x^{*j}\right) + U^2_y(L)\left(x^{2j}(L) - x^{*j}(L)\right)\right] = \lambda(R-1)\hat{s}_j$$

Notice also that $U^2_y = RU^2_y(L)$ implies

$$\left[R\left(x^{2j} - x^{*j}\right) + \left(x^{2j}(L) - x^{*j}(L)\right)\right] = \frac{\lambda(R-1)}{2\mu U^2_y(L)}\hat{s}_j. \quad (31)$$

Therefore, for the expression to hold true, we need

$$\frac{R\left(x^{2j} - x^{*j}\right) + \left(x^{2j}(L) - x^{*j}(L)\right)}{\hat{s}_j}$$

to be independent of j.

We cannot understand what is necessary for this condition to be satisfied without, first, understanding $\hat{s}_j$. Let us, then, explore the its meaning.

Consider the following problem:

$$\max_s \left\{ 2U(\mathbf{p}, y - s, Y, w) + \sum_{i=H,L} U\left(\mathbf{p}, y^i + Rs, Y^i, w^i\right) \right\} \quad (32)$$

The first and second order conditions for this problem are, respectively

$$2U_y - R\left[U^*_y(H) + U^*_y(L)\right] = 0,$$

and

$$\Theta \equiv 2U_{yy} + R^2\left[U^*_{yy}(H) + U^*_{yy}(L)\right] < 0.$$

We may then use this to find s_j :

$$\begin{aligned}
\frac{ds}{dp_j} &= -\frac{-2U_{yj} + R\left[U^*_{yj}\left(H\right) + U^*_{yj}\left(L\right)\right]}{\Theta}, \\
&= -\frac{2U_{yy}x^j - R\left[U_{yy}\left(H\right)x^j\left(H\right) + U_{yy}\left(L\right)x^j\left(L\right)\right]}{\Theta} \\
& \quad -\frac{U_y x^j_y - R\left[U_y\left(H\right)x^j_y\left(H\right) + U_y\left(L\right)x^j_y\left(L\right)\right]}{\Theta}
\end{aligned}$$

where, once again, Roy's identity was used.

The same procedure for s_y, s_{y^H} and s_{y^L}, yields

$$\begin{aligned}
\frac{ds}{dy} &= -\frac{-2U_{yy}}{\Theta}, \\
\frac{ds}{dy^H} &= -\frac{RU_{yy}\left(H\right)}{\Theta}, \text{ and} \\
\frac{ds}{dy^L} &= -\frac{RU_{yy}\left(L\right)}{\Theta}, \text{ respectively.}
\end{aligned}$$

Combining these, it is readily seen that

$$\begin{aligned}
\frac{ds}{dp_j} &= -\frac{ds}{dy}x^j - \frac{ds}{dy^H}x^j\left(H\right) - \frac{ds}{dy^L}x^j\left(L\right) \\
& \quad -\frac{2U_y x^j_y - R\left[U_y\left(H\right)x^j_y\left(H\right) + U_y\left(L\right)x^j_y\left(L\right)\right]}{\Theta}.
\end{aligned}$$

Hence,

$$\hat{s}_j = -\frac{2U_y x^j_y - R\left[U_y\left(H\right)x^j_y\left(H\right) + U_y\left(L\right)x^j_y\left(L\right)\right]}{\Theta}$$

Because $2U_y = R\left[U_y\left(H\right) + U_y\left(L\right)\right]$, we can finally write

$$\hat{s}_j = R\frac{U_y\left(H\right)\left(x^j_y\left(H\right) - x^j_y\right) + U_y\left(L\right)\left(x^j_y\left(L\right) - x^j_y\right)}{\Theta} \tag{33}$$

The condition, therefore, is equivalent to having

$$\frac{R\left(x^{2j} - x^{*j}\right) + \left(x^{2j}\left(L\right) - x^{*j}\left(L\right)\right)}{U_y\left(H\right)\left(x^j_y\left(H\right) - x^j_y\right) + U_y\left(L\right)\left(x^j_y\left(L\right) - x^j_y\right)} \tag{34}$$

is independent of j.

The important thing to realize here is that, if goods have different income elasticity of demand this condition is (generically) not satisfied, despite our having imposed separability. Hence, AS breaks down.

We should expect homotheticity to do the job. In fact, what we have shown in section 3 was that, with homotheticity, optimal taxes are uniform in both periods. In this case, the constraint that taxes - hence prices - do not differ across time is not binding.

For completeness, however we show that it is really the case that homotheticity delivers uniform taxation as the optimal policy.

If preferences are homothetic,

$$\begin{aligned}
p^j x^{2j} &= \omega^j \left(y - s^2\right), \\
p^j x^{*j} &= \omega^j \left(y - s^*\right), \\
p^j x^{2j}(L) &= \omega^j \left(y^L + Rs^2\right), \\
p^j x^{*j}(L) &= \omega^j \left(y^L + Rs^*\right), \\
p^j x_y^j(i) &= \omega^j, \; i = H, L,
\end{aligned}$$

where ω^j is the (constant) proportion of income spent on good j.

In this case, $\hat{s}_j$ and the left hand side of (31) both vanish for all j, which guarantees that the condition is satisfied.

Unfortunately, because the two sides vanish at the optimum we cannot use expression (31) to evaluate the sign of $R - 1$. Hence, we derive first order conditions with respect to R, assume homotheticity, and show that $R < 1$.

If preferences are homothetic, $\mathbf{p} = \mathbf{1}$ and the first order condition with respect to R is

$$\left(U_y^*(H) + U_y^*(L)\right) s^* (1 + \mu) - \mu \left(U_y^2(L|H) + U_y^2(L)\right) s^2 - \\ \textstyle\sum_k \left[x_y^k s_R + \sum_{i=H,L} x_y^k(i) \left(s^* - Rs_R^*\right)\right]$$

Back to program (32), we adopt the same procedure we have been adopting so far to get

$$\frac{ds}{dR} = \frac{\left[U_y^*(H) + U_y^*(L)\right] + R\left[U_{yy}^*(H) + U_{yy}^*(L)\right] s}{\Theta}.$$

Next, we define $\hat{s}_R \equiv s_R - s_{y^H} s - s_{y^L} s$.

$$\hat{s}_R = \frac{U_y^*(H) + U_y^*(L)}{\Theta} < 0$$

Multiply (26) and (27) by s^* and consider the fact that $U_y^2(L|H) = U_y^2(L)$, under separability, to get:

$$\mu U_y^o(L) \left[s^* - s^2\right] = \lambda \textstyle\sum_k \left[-x_y^k + R \sum_{i=H,L} x_y^k(i)\right] \hat{s}_R^*$$

We have, then, seen that $\mathbf{p} = \mathbf{1}$, $x_y^k = x_y^k(H) = x_y^k(L) = 1$. Hence,

$$\mu U_y^2(L) \left[s^* - s^2\right] = \lambda [R - 1] \hat{s}_R^*$$

So, $R < 1$ if and only if $s^* < s^2$.

This just proves the obvious, given what we have seen in section 3. When, Proposition 5 is valid, the restriction of equal taxes across time is not binding.

References

[1] Atkinson A.B., Stiglitz, J. E. (1976) ”The Design of Tax Structure: Direct versus Indirect Taxation.” *Journal of Public Economics* 6, 55-75.

[2] Brito, D., Hamilton, J., Slutsky, S. and Stiglitz, J. (1991) ”Dynamic Optimal Income Taxation with Government Commitment” *Journal of Public Economics, 44: 15-35*

[3] Chamley, C. (1986) ”Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives” *Econometrica*, 58: 69-76.

[4] Chiappori, P.-A., Macho, I., Rey P. and Salanié, B. (1994) ”Repeated Moral Hazard: the Role of Memory, Commitment and the Access to Credit Markets”, *European Economic Review*, **38**: 1527-53.

[5] Christiansen, V. (1984) ”Which Commodity Taxes Should Supplement the Income Tax?” *Journal of Public Economics*, 24: 195-210.

[6] Cooter, R. (1978) ”Optimal Tax Schedules and Rater: Mirrlees and Ramsey”. *American Economic Review,* **68:** 756-768.

[7] Cremer, H and Gahvari, F. (1995) ”Uncertainty, Optimal Taxation and the Direct Versus Indirect Tax Controversy.” *The Economic Journal* **105**, 1165-1179

[8] Cremer, H and Gahvari, F. (1999) ”Uncertainty, Commitment and Optimal Taxation.” *Journal of Public Economic Theory*, 1: 51-70

[9] Cremer, H., Pestieau, P. and Rochet, J.-C. (2001a) ”Capital Income Taxation when Inherited Wealth is not Observable” mimeo, IDEI.

[10] Cremer, H., Pestieau, P. and Rochet, J.-C. (2001b) ”Direct versus Indirect Taxation: the Design of the Tax Structure Revisited” *International Economic Review*, **42**: 781-799.

[11] da Costa, C. (2001) ”Two Essays on Supplementary Commodity Taxation” PhD Thesis, University of Chicago.

[12] da Costa, C. and Werning, I. (2000). ’A Note on Pre-Committed Goods’. mimeo., University of Chicago.

[13] da Costa, C. and Werning, I. (2001) ”On The Optimality of the Friedman Rule with Heterogeneous Agents and Non-linear Taxation” mimeo., University of Chicago.

[14] Diamond, P. and Mirrlees, J. (1978) ”A Model of Social Insurance with Variable Retirement” *Journal of Public Economics* 10: 295-336.

[15] Deaton, A. and Muellbauer, J. (1980) ”Economics and Consumer Behavior” Cambridge: Cambridge University Press

[16] Dewatripont, M. (1988) "Commitment through Renegotiation-proof Contracts with Third Parties" *Review of Economic Studies* **55**, 377-90.

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

[18] Golosov, M., Kocherlakota, N. and Tsyvinski, A. (2003) "Optimal Indirect and Capital Taxation" *Review of Economic Studies* **70**, 569-587.

[19] Golosov, M., and Tsyvinski, A. (2003) "Designing Optimal Disability Insurance" *Federal Reserve Bank of Minneapolis Research Department Working Paper* 293.

[20] Judd, K. (1987) "The Welfare Cost of Factor Taxation in a Perfect Foresight Model" *Journal of Political Economy*, **95**:675-709

[21] Kocherlakota, N. (2003) "A Mirrlees Approach to Dynamic Optimal Taxation: Implication for Wealth Taxes and Asset Prices" *Federal Reserve Bank of Minneapolis Research Department Staff Report* ???.

[22] Konishi, H. (1995) "A Pareto-improving Commodity Tax Reform Under a Smooth Nonlinear Income Tax". *Journal of Public Economics*, **56**: 413-446.

[23] Lucas, R. E. (1990) "Supply-side Economics: an Analytical Review". *Oxford Economic Papers*, 42: 293-316

[24] Mirrlees, J. (1971) "An Exploration on the Theory of Optimum Income Taxation" *Review of Economic Studies* **38**, 175-208.

[25] Mirrlees, J. (1976) "Optimal Tax Theory: A Synthesis," *Journal of Public Economics*, **6**: 327-58.

[26] Naito, H. (1999) "Reexamination of Uniform Commodity Taxes Under a Non-Linear Tax System and Its Implications For Production Efficiency" *Journal of Public Economics*, **71**: 165-188.

[27] Rogerson, W. (1985) "Repeated Moral Hazard" *Econometrica*, 58: 69-76.

[28] Saez, E. (2002) "The Desirability of Commodity Taxes with Non-linear Income Taxation and Heterogeneous Tastes" *Journal of Public Economics*, **83**: 217-230.

[29] Stiglitz, J. E. (1982) "Self-selection and Pareto Efficient Taxation," *Journal of Public Economics*, **17**: 213-240.

[30] Stiglitz, J. E. (1987) "Pareto Efficient and Optimal Taxation and the New New Welfare Economics" in *Handbook of Public Economics*, vol. II, Chapter 15, p. 991-1042.