

# Dynamic optimal income taxation with government commitment\*

Dagobert L. Brito

*Rice University, Houston, TX 77251, USA*

Jonathan H. Hamilton and Steven M. Slutsky

*University of Florida, Gainesville, FL 32611, USA*

Joseph E. Stiglitz

*Stanford University, Stanford, CA 94305, USA*

Received March 1989, revised version received May 1990

We analyze the optimal income tax problem when consumers work for many periods. Can information about abilities learned in one period be used later to attain more redistribution than is possible in a one-period world? When the government can commit to future policies and has the same discount rate as individuals, intertemporal nonstationarity of tax schedules can lead to Pareto improvements by relaxing lifetime self-selection constraints, analogous to randomization in a one-period world. More significant use of information is possible when the social discount rate is less than individuals' rates. The taxes converge over time to a nondistortionary schedule.

## 1. Introduction

Governments make some, but only limited, use of age-dependent income tax schedules. In the United States, the elderly benefit from a larger standard deduction in the Federal income tax, and the subsidies provided through the Social Security system (which depend on individual earnings histories) can be viewed as age-dependent transfers. At first glance, this limited use seems anomalous. Surely by the time an individual reaches age 50, the government has accumulated an enormous amount of information on him. The government could design tax schedules in which an individual's tax liability depends on his complete lifetime work history. This argument seems particularly

\*The authors thank David Sappington and two anonymous referees for valuable comments. Jonathan Hamilton thanks the Public Policy Research Center for financial support. Joseph Stiglitz thanks the Hoover Institution, the John M. Olin Foundation, and the National Science Foundation for financial support.

forceful under the now-widely-accepted view that limited information about individual abilities provides the reason for distortionary taxation. Lacking such information, the government must offer everyone the same tax schedule. The set of feasible consumption–leisure bundles must then satisfy self-selection constraints that limit the government’s redistribution capabilities.

Upon further reflection, this failure to use previous earnings to determine current taxes may not be so anomalous. Were the government to do so, individuals would recognize that their current income will affect future tax rates and would accordingly alter their current behavior. While the government might be able to use the information accumulated during an individual’s lifetime to reduce the distortion associated with progressive taxation in later years, doing so would increase the distortion in earlier years.

Indeed, what appears more anomalous is that, in the absence of an explicit commitment, individuals act as if they assume that the government will not base future taxes on current income. But this is also less of a paradox than it first appears – individuals save, even though the government has not committed itself not to confiscate wealth. The government’s desire to maintain a reputation can lead it to act as if it can make binding commitments not to use information learned in one period to engage in first-best lump-sum taxation in later periods. We analyze here the intertemporal structure of income taxation for a model with two ability classes when the government can commit itself to lifetime tax schedules. The government faces a single lifetime self-selection constraint for each type of individual which is essentially the same as the one-period self-selection constraint. It is easy to show that optimal tax schedules lead individuals immediately to reveal their ability levels. Learning individuals’ abilities is of no direct value to the government since it has committed itself not to use that information. This contrasts with Roberts (1984) who examines intertemporal optimal income taxation with an infinite horizon when the government cannot commit itself even implicitly not to use past incomes to determine tax schedules. In response, all individuals choose to earn the same income to prevent being heavily taxed in the future.

With preferences and productivity unchanging over time and with the usual assumption of strictly convex preferences, we might expect the mathematics simply to confirm the optimality of what we usually observe – tax schedules that do not vary with age. But this is not the case under all preference maps and social welfare functions. The possibility of nonstationary tax schedules should not have been unexpected. Even with concave utility functions, the self-selection constraints introduce a nonconvexity into the government’s optimization problem. Nonstationary tax schedules may arise in the optimum, just as the optimal solution in a one-period problem may entail random taxation [see, for example, Stiglitz (1982)]. By varying tax rates over time, the government can simulate the outcome that would have

resulted from random taxation. Age dependence is not really the issue, since we will see that the order in which the tax schedules are presented is of minimal importance.

To focus on the use of information over time, we concentrate our attention on the case where the government and individuals discount the future at the same rate. The government is able to commit itself to future tax schedules for individuals contingent upon past behavior. If the optimal one-period solution does not have random tax schedules, the government can do more redistribution in the multi-period problem than in the one-period problem. If the one-period solution entails randomization, a nonstationary dynamic solution can achieve similar benefits. A problem with random tax schedules is that horizontal equity may be violated – *ex post*, identical individuals receive different consumption bundles yielding different utilities. Intertemporal nonstationarity is an alternative method of implementing random taxes that does not require *ex ante* or *ex post* violations of horizontal equity.

These results follow from noting that a multi-period optimal tax model and a one-period model with random taxation are essentially the same. The analogy between random taxation and nonstationary taxation is suggestive, but there are differences between them. The random tax model has no restrictions on the relative frequencies with which different schedules are offered. With an infinite horizon and slow discounting, it is possible to duplicate exactly the random solution. However, with a finite number of periods or a high rate of discounting, there will be restrictions on the possible frequencies with which different schedules are offered.

When the government values the future more than individuals do, more systematic use of age-dependent tax schedules than just mimicking randomization is possible. As compared with the solution with identical discount rates, the government offers higher utility early and lower utility later to one of the types, while it offers the reverse to the other. Individuals weight the present relatively more in their self-selection constraints than does the government in its objective function. The government takes advantage of this difference in preferences in setting its optimal policy and causes the economy to approach over time the first-best allocation with nondistortionary redistribution.

To some readers, randomization may seem a mathematical curiosity of limited interest to policy design.<sup>1</sup> However, some of the arguments against implementing randomization are less serious criticisms of nonstationary tax schedules. With the latter, as noted above, it is still possible not to violate horizontal equity *ex post*. Intertemporal changes in the tax schedule only require keeping track of individuals' ability types (inferred from previous

<sup>1</sup>Examples of papers exploring the role of randomization in adverse selection and other problems include Weiss (1976), Stiglitz (1982), Fellingham, Kwon and Newman (1984), Maskin and Riley (1984), and Arnott and Stiglitz (1988).

income histories) which may be far simpler than implementing a randomization procedure that is perceived as fair and appropriate. Furthermore, the gains from using randomization may well be large; Brito et al. (1989) give an example in which these gains may be equivalent to as much as 12 percent of aggregate resources. Thus, one does not wish to dismiss lightly the use of nonstationary tax schedules.

Section 2 presents the basic model and summarizes both simple characterization results for efficient tax structures and necessary and sufficient conditions for randomization to be desirable. Section 3 presents our results for the model with equal discount rates. Section 4 considers the problem when the government has a different discount rate from individuals, and section 5 contains our conclusions.

## 2. The intertemporal model

Consider a society composed of two different classes of individuals, denoted A and B. The individuals within each class are identical, but the two classes differ either in tastes or abilities. The government is assumed initially not to know to which class any individual belongs but knows the numbers of individuals in each class, which we normalize to one. Let  $\rho$  be the discount factor for individuals in both classes. In each period, individuals consume a single good,  $C$ , and work to earn income,  $Y$ . Within any period, all members of each class have the same utility function over these bundles  $V^i(C, Y)$ ,  $i = a, b$ , with  $\partial V^i / \partial C \equiv V_c^i > 0$  and  $\partial V^i / \partial Y \equiv V_y^i < 0$ .<sup>2</sup> Utility declines with  $Y$  for a given  $C$  since the rise in  $Y$  implies a decrease in leisure. Individuals live for  $M$  periods and have a lifetime utility function  $\hat{V}^i = \sum_{t=1}^M \rho^{t-1} V^i$ ,  $i = a, b$ . The maximum income that individuals in each class can earn per period is bounded from above by  $K^i$ , so that  $Y^i \leq K^i$ ,  $i = a, b$ . The marginal rate of substitution for individual  $i$  is denoted  $MRS^i(C, Y) \equiv -V_y^i / V_c^i > 0$ ,  $i = a, b$ .

The following assumptions are made about  $V^i(C, Y)$ :

*Assumption 1.*  $V^i(C, Y)$ ,  $i = a, b$ , is twice continuously differentiable in  $C$  and  $Y$ .

*Assumption 2.*  $V^i(C, Y)$ ,  $i = a, b$ , is strictly concave in  $C$  and  $Y$ .

*Assumption 3.*  $MRS^a(C, Y) < MRS^b(C, Y)$ ; type A's indifference curves are flatter than B's.

Assumption 1 is made for convenience in exposition but can be relaxed

<sup>2</sup>The utility functions defined over  $\{C, Y\}$  can be derived from the more fundamental utility functions defined over goods and leisure. Let  $L^i$  be hours worked and  $w_i$  the wage rate of group  $i$ . Then  $L^i = Y/w_i$  and  $V^i(C, Y) = U^i(C, Y/w_i)$ , where  $U^i$  is  $i$ 's utility function over  $C$  and  $L$ .

without difficulty. Assumption 2 of concavity (instead of quasiconcavity) insures that lifetime discounted utility describes convex preferences. Assumption 3 is the single crossing assumption. Sadka (1976) has shown that common preferences over consumption and leisure, differences in earning abilities, and noninferiority of consumption are sufficient for single crossing. In Brito et al. (1990), optimality conditions analogous to those in Theorems 1–4 below are obtained without assuming single crossing. Here we assume it for ease in exposition.

Our major theme here is that nonstationary tax schedules may be used by the government to mimic the effects of randomization in the single period problem. Baron and Besanko (1984) and Laffont and Tirole (1988) report that, with commitment by the principal, nonstationary reward schedules are of no value. However, they use particular functional forms for which randomization is not desirable in one-shot settings. We consider general utility functions, including those for which randomization may be useful in one-period applications.

Brito et al. (1989) present necessary and sufficient conditions for randomization and characterize the optimal randomization in a single-period version of this model. Our results there include the following:

- (1) if consumers are sufficiently risk-averse, there exists a randomization scheme that preserves horizontal equity within types (defined as equal ex post utility, not equal consumption);
- (2) the optimal randomization scheme requires at most three distinct consumption bundles to be included in the lottery offered one type of consumer; and
- (3) if a local randomization is desirable with some probability vector over three bundles, then for any other probability triple there exist three bundles such that randomization over those bundles at the given probabilities is feasible and improves on the deterministic solution.<sup>3</sup>

Individuals are assumed to face  $M$  separate budget constraints that require that consumption in each period not exceed after-tax income in that period. We assume this to focus purely on the role of information transfer across periods in affecting taxation without complicating the analysis with the possibility of wealth or interest taxation. Assuming  $M$  budget constraints also eliminates the need to consider individual saving decisions. In the first period, every individual faces the same tax function,  $T^1(Y_1)$ , since the government has no basis upon which to distinguish individuals. Thereafter, the government can recall the incomes reported in previous periods by that individual and can condition the tax functions on previous periods' income.

<sup>3</sup>A local randomization scheme is one where the value of social welfare is increased by arbitrarily small deviations from the deterministic solution.

Thus, the tax function in period  $t > 1$  is written as  $T^t(Y_t | Y_1, \dots, Y_{t-1})$ . The government cannot condition an individual's current taxes on the current or past behavior of others.<sup>4</sup> Taking the sequence of tax functions as given, each individual chooses lifetime consumption and income to solve the following maximization:

$$\max_{(C_t, Y_t)} \sum_{t=1}^M \rho^{t-1} V^i(C_t^i, Y_t^i)$$

$$\text{s.t. } C_1^i \leq Y_1^i - T^1(Y_1^i),$$

$$C_t^i \leq Y_t^i - T^t(Y_t^i | Y_1^i, \dots, Y_{t-1}^i), \quad t = 2, \dots, M.$$

The solution gives lifetime consumption and income vectors as functions of the vector of tax functions,  $\hat{C}^i(T^1, \dots, T^M)$  and  $\hat{Y}^i(T^1, \dots, T^M)$ .

Given the choices by individuals in each class and subject to the budget constraint, the government chooses the set of tax functions to maximize social welfare (to be defined below). To analyse this, we carry out the standard transformation of this problem to the equivalent one of the government choosing lifetime allocations for each type which satisfy self-selection constraints and resource balance. Clearly, given the tax functions, the optimizing choices by individuals must satisfy the self-selection constraints since each person chooses the best allocation available. Conversely, for any set of bundles satisfying the self-selection constraints, there exist a series of tax functions that allow only those bundles to be chosen. For example, tax functions that lead to after-tax budget constraints that are step functions can do this. Individuals will choose to be only at bundles where consumption has jumped up.<sup>5</sup>

There is only one lifetime constraint for each class and no period-by-period self-selection constraints. In the first period, individuals decide

<sup>4</sup>This contrasts with Harris (1987). If the government knows that half the population are type A and half type B, it can 'force' truthful revelation in a Nash equilibrium by imposing heavy penalties on *all* individuals if more than half claim to be of a particular type. We find these Nash equilibria unpersuasive.

<sup>5</sup>When a self-selection constraint holds with equality, one of the groups is indifferent between the two bundles offered. The solution requires that all individuals in the group choose the bundle aimed at that group. This can be achieved by assuming that the government can assign indifferent individuals to whichever group it desires. Given that the government does not know to which group a particular individual belongs, this is not a reasonable assumption. An alternative view is that the solution is really an  $\epsilon$ -equilibrium. Although it cannot itself be achieved, a bundle arbitrarily close to that solution can be found which satisfies resource balance and which has the self-selection constraint hold with strict inequality. If the self-selection constraints must hold with strict inequality, then there may exist no solution to the maximization problem.

whether or not to reveal their type when choosing their current income based on the entire lifetime consequences that follow. If individuals reveal their type in the first period, the government knows thereafter who they are and can prevent them from acting as if they belong to a different class. Any attempted deviation could be punished by imposition of a large penalty. After revealing through first-period choices, individuals in later periods can no longer choose any bundle other than the one the government desires them to consume. Hence, after the first period, the government's choices are constrained only because of its initial commitment. It is well known that in the one-period problem, optimality requires separation, that is, the government chooses tax schedules such that individuals reveal their types. Because the government can commit itself, and separation is desirable in a one-period problem, separation will occur in every period.

The government has a single intertemporal budget constraint:  $\sum_{t=1}^M \delta^{t-1} [T_a^t + T_b^t] \geq 0$ , where  $T_i^t$  is tax revenue collected from group  $i$  in period  $t$  and  $\delta$  is the government's discount factor. This budget constraint is appropriate if the government can save or borrow using opportunities not available to individuals, such as a storage technology feasible only on a large scale or access to a world market closed to trade by individuals.<sup>6</sup>

The government maximizes the present discounted value of a weighted sum of lifetime utilities where the weights  $\alpha$  and  $(1-\alpha)$  are arbitrary and can vary to change the distribution between the groups. The weights on the two groups can be varied to trace out the entire utility possibility frontier. Even with different abilities and similar preferences, obtaining the complete frontier requires consideration of the case where the less able envy the bundle of the more able [see Stiglitz (1982)]. The government's discount factor,  $\delta$ , need not equal that of individuals. When they are equal ( $\delta = \rho$ ), the government's maximization corresponds to finding the multi-period Pareto frontier as  $\alpha$  varies from 0 to 1. When they differ ( $\delta \neq \rho$ ), the problem is no longer a Pareto problem since the government does not respect individuals' intertemporal preferences. While much literature analyzes why private and social discount rates could differ, these do not constitute our major reason for considering  $\delta \neq \rho$ . The major focus is on the case of equal discount rates. Allowing them to differ gives rise to a case which serves as a useful benchmark for comparison when discussing the uses of information in the optimal tax structure. The use of information across periods when  $\delta = \rho$  is much less systematic than when  $\delta \neq \rho$ .

Formally, the government's maximization problem is:

<sup>6</sup>Alternatively, the government faces a new cohort each period and is constrained to use the same tax structure for all cohorts. Such a framework gives rise to a single within-period budget constraint when redistribution across cohorts is possible. The independent problems in each period are identical to the problem with a single intertemporal constraint.

$$\begin{aligned}
\text{(P)} \quad & \max \sum_{t=1}^M \delta^{t-1} [\alpha V^a(C_t^a, Y_t^a) + (1-\alpha) V^b(C_t^b, Y_t^b)] \\
\text{s.t.} \quad & \sum_{t=1}^M \rho^{t-1} [V^i(C_t^i, Y_t^i) - V^i(C_t^j, Y_t^j)] \geq 0, \quad i = a, b, j \neq i: \lambda_i \\
& \sum_{t=1}^M \delta^{t-1} [C_t^a - Y_t^a + C_t^b - Y_t^b] \leq 0 \quad : \mu \\
& 0 \leq Y_t^i \leq K^i, \quad i = a, b, t = 1, \dots, M, \\
& C_t^i \geq 0, \quad i = a, b, t = 1, \dots, M.
\end{aligned}$$

The Lagrange multipliers on the self-selection constraints are  $\lambda_a$  and  $\lambda_b$ . The Lagrange multiplier on the budget constraint is  $\mu$ .

### 3. Optimal taxation when $\delta = \rho$

Let  $C_t^i(\alpha)$  and  $Y_t^i(\alpha)$ ,  $i = a, b$ , and  $t = 1, \dots, M$ , be the solutions to (P) as functions of  $\alpha$ . Define  $V^{ii}(\alpha) \equiv V^i(C_t^i(\alpha), Y_t^i(\alpha))$ .  $V^{i0}$ ,  $i = a, b$ , are the optimal one-period utilities when no taxes are imposed. Properties of the optimal solution when  $\delta = \rho$  are given in the following four theorems. All proofs are deferred to the appendix.

*Theorem 1. If individuals of some type are at least as well off as with no taxation, then that type's self-selection constraint is not binding. That is,*

$$\begin{aligned}
& \text{if } \sum_{t=1}^M \rho^{t-1} V^{ii}(\alpha) \geq V^{i0} \sum_{t=1}^M \rho^{t-1}, \\
& \text{then } \sum_{t=1}^M \rho^{t-1} V^i(C_t^j(\alpha), Y_t^j(\alpha)) < \sum_{t=1}^M \rho^{t-1} V^{ii}(\alpha), \\
& i = a, b \text{ and } j \neq i.
\end{aligned}$$

See fig. 1 for an illustration. If the government had to raise positive revenue, Theorem 1 holds with the modification that the  $V^{i0}$  are the one-period utilities when equal lump-sum taxes are imposed.

*Theorem 2. Only one type of individual can have a binding self-selection constraint. Formally,*



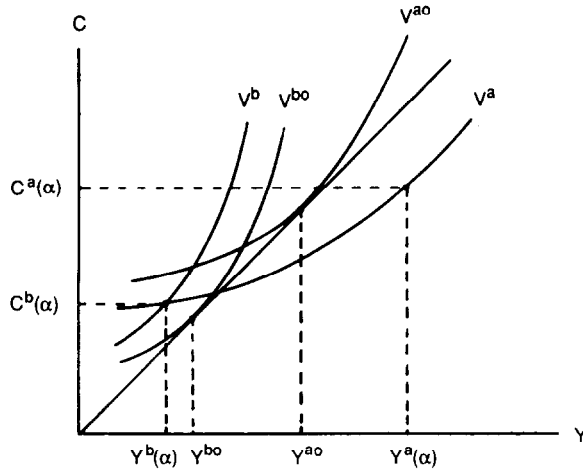


Fig. 1. Since type b is better off than at the no-tax outcome, b's self-selection constraint is not binding.

$$\text{if } \sum_{t=1}^M \rho^{t-1} V^{it}(\alpha) = \sum_{t=1}^M \rho^{t-1} V^i(C_t^j(\alpha), Y_t^j(\alpha)),$$

$$\text{then } \sum_{t=1}^M \rho^{t-1} V_t^j(\alpha) > \sum_{t=1}^M \rho^{t-1} V^j(C_t^i(\alpha), Y_t^i(\alpha)), \quad i = a, b \text{ and } j \neq i.$$

It follows immediately from this that any additional resources could be given to a group whose bundle is viewed as inferior by the other group without violating the self-selection constraints. Thus, production efficiency must hold, implying  $\mu > 0$ .

*Theorem 3.* If the self-selection constraint for one group does not bind, then for the other group, the consumption–income bundle is stationary and its MRS there equals the marginal rate of transformation (which equals one). That is, if  $\lambda_i = 0$ , then  $C_t^j(\alpha) = C_t^j(\alpha)$ ,  $Y_t^j(\alpha) = Y_t^j(\alpha)$  and  $MRS^j(C_t^j(\alpha), Y_t^j(\alpha)) = 1$ ,  $i = a, b$ ,  $j \neq i$ ,  $t = 1, \dots, M$ .

*Theorem 4.* Assume one group's self-selection constraint binds. In all periods, the other group's MRS at its own bundle lies between the MRT and the MRS of the group whose constraint binds. That is, at each  $t = 1, \dots, M$ , if  $\lambda_a > 0$ ,  $MRS^a(C_t^b(\alpha), Y_t^b(\alpha)) < MRS^b(C_t^b(\alpha), Y_t^b(\alpha)) < 1$ , while if  $\lambda_b > 0$ ,  $1 < MRS^a(C_t^a(\alpha), Y_t^a(\alpha)) < MRS^b(C_t^a(\alpha), Y_t^a(\alpha))$ .

Thus, if distortions exist in the one-period solution, then they exist in all periods of the multi-period solution. Information learned in the first period is not used to move to the first-best allocation in later periods. If the government uses the information it learns in specifying tax functions after the first period, it is committed to making only limited use of this information.

In our formal results above, we assume single crossing. Except for Theorem 4, this assumption plays no direct role in the proofs. In Theorem 4, without single crossing, it remains true that if  $\lambda_i > 0$ , then the *MRS* of group  $j$  at  $j$ 's bundle lies at least weakly between the *MRS* of the other group and the *MRT*. What differs is that either both *MRS* could be greater or less than the *MRT*, or all three could be equal. Furthermore, which of these possibilities occurs can change between periods.

From these results, it is not clear whether the government uses information gained in the first period to affect later period taxes. A simple repetition of the one-period nonrandom solution satisfies all the first-order conditions for (P). However, despite the apparent symmetry of the first-order conditions, multiple asymmetric solutions may arise in the form of nonstationarity of the optimal consumption–income vectors. Such nonstationary solutions arise from the same nonconvexity of the self-selection constraints that make random solutions in the one-period problem optimal. In fact, the following theorems show that there is an exact analogy between existence of a nondegenerate solution to the random tax problem and a nonstationary solution to (P).

Let

$$\bar{V}^i(\alpha) = \sum_{t=1}^M \rho^{t-1} V^i(C_t^i(\alpha), Y_t^i(\alpha)) \Big/ \sum_{t=1}^M \rho^{t-1}, \quad i = a, b,$$

be the average utility achieved by each group over its lifetime. Let the normalized utility possibility frontier be the utility possibility frontier in average utilities.

*Theorem 5.* *If  $\delta = \rho$  and the solution to (P) is nonstationary, then in the one-period problem some random taxation would improve on the deterministic optimum.*

*Theorem 6.* *If the solution to the one-period problem uses random taxation, if  $M = \infty$ , and if  $\delta = \rho \geq \frac{2}{3}$ , then the solution to (P) is nonstationary. The normalized Pareto frontier in (P) is identical to the Pareto frontier in expected utilities in the one-period problem.*

The random solution uses at most three bundles for the type whose

constraint does not bind. Since the probabilities associated with these bundles are not restricted, any set of probabilities summing to one could be the optimal solution. In the intertemporal problem, similarly only three different bundles need be assigned at different periods. A high discount factor ( $\delta \geq \frac{2}{3}$ ) and an infinite number of periods ensure that these three bundles can be assigned with the same discount weights as the bundles in the one-period problem have probability weights.<sup>7</sup> Thus, the intertemporal solution can exactly duplicate the random solution.

Even when exact duplication is not possible, nonstationarity may still be optimal.

*Theorem 7. If the optimal solution to the one-period problem uses local randomization, then even if  $\delta = \rho < \frac{2}{3}$  or if  $M$  is finite, the solution to (P) is nonstationary. The normalized Pareto frontier may be interior to that in the one-period random solution.*

In Brito et al. (1989) it is shown that if there exists a local randomization solution which improves on the deterministic solution, then an improving randomization can be found for any probability triple. Thus, even when the optimal random solution cannot be duplicated, another random solution can be duplicated which improves on the stationary solution. The necessary and sufficient condition for local randomization is given in that paper. Consider the deterministic solution with  $\lambda^i > 0$ . At that solution, let

$$\theta = [V_c^j(C^j, Y^j) | 1 - MRS^j(C^j, Y^j)] / [V_c^i(C^j, Y^j) | 1 - MRS^i(C^j, Y^j)]$$

and let  $H^i(C^j, Y^j)$  be the Hessian of  $i$ 's utility function at  $(C^j, Y^j)$ . Then some local randomization improves on the deterministic solution if and only if the matrix  $H^j(C^j, Y^j) - \theta H^i(C^j, Y^j)$  is not negative semidefinite. In essence, this condition states that local randomization is desirable if the utility function of the group whose self-selection constraint binds is more concave than that of the other group.

#### 4. Optimal taxation when $\delta \neq \rho$

When the government and individuals have different discount rates, systematic nonstationarity arises in the optimal solution. The first-order conditions in (P) are:

<sup>7</sup>Note that, if  $\rho = \delta < \frac{2}{3}$ , there exist triples,  $\Pi_i$ ,  $i = 1, 2, 3$ , which cannot be the sums of terms in the sequences  $(1 - \rho)\rho^{t-1}$ ,  $t = 1, \dots, \infty$ . For example, if  $\rho = \frac{1}{2}$ , the first term of the sequence equals  $\frac{1}{2}$  and the first-period bundle must have a discount weight of at least  $\frac{1}{2}$  in the lifetime utility function. A probability triple,  $\Pi_1 = \Pi_2 = \Pi_3 = \frac{1}{3}$ , could not be duplicated in the intertemporal problem.

$$[\delta^{t-1}\alpha + \lambda_a\rho^{t-1}](\partial V^a/\partial C_t^a) - \lambda_b\rho^{t-1}(\partial V^b/\partial C_t^a) - \mu\delta^{t-1} = 0, \quad t=1, \dots, M, \quad (1a)$$

$$[\delta^{t-1}\alpha + \lambda_a\rho^{t-1}](\partial V^a/\partial Y_t^a) - \lambda_b\rho^{t-1}(\partial V^b/\partial Y_t^a) + \mu\delta^{t-1} = 0, \quad t=1, \dots, M, \quad (1b)$$

$$[\delta^{t-1}(1-\alpha) + \lambda_b\rho^{t-1}](\partial V^b/\partial C_t^b) - \lambda_a\rho^{t-1}(\partial V^a/\partial C_t^b) - \mu\delta^{t-1} = 0, \quad t=1, \dots, M, \quad (1c)$$

$$[\delta^{t-1}(1-\alpha) + \lambda_b\rho^{t-1}](\partial V^b/\partial Y_t^b) - \lambda_a\rho^{t-1}(\partial V^a/\partial Y_t^b) + \mu\delta^{t-1} = 0, \quad t=1, \dots, M. \quad (1d)$$

When  $\delta > \rho$ , it follows from these conditions that as  $t$  increases, the economy approaches the one-period first-best *UPF* for any values of  $\lambda_a$  and  $\lambda_b$ . Because of the distortions in early periods, the normalized *UPF* based on average utilities is interior to the first-best *UPF*.

*Theorem 8.* Consider (P) when  $\delta > \rho$  and  $M = \infty$ . Assume that the utility functions satisfy the conditions that  $\lim_{C \rightarrow 0} (\partial V^a/\partial C)/(\partial V^b/\partial C)$  and  $\lim_{Y \rightarrow K^b} (\partial V^a/\partial Y)/(\partial V^b/\partial Y)$  are finite. Then, when  $\lambda_b = 0$  and  $\lambda_a > 0$ ,  $\lim_{t \rightarrow \infty} MRS^b(C_t^b, Y_t^b) = 1$ . Since  $MRS^a(C_t^a, Y_t^a) = 1$ , for all  $t$ , as  $t$  grows, the optimal solution approaches the one-period first-best Pareto frontier.

The following example shows that the assumptions of this theorem can be satisfied. With identical, additively separable utility functions ( $V^i(C, Y) = \psi(C) - \gamma(Y/w_i)$ ) and  $w_a > w_b$ ,  $(\partial V^a/\partial C)/(\partial V^b/\partial C) = 1$  at all  $C$ . The maximum income earned by the more able exceeds that earned by the less able ( $K^a > K^b$ ). Hence, even if  $\partial\gamma/\partial L$  goes to infinity as  $L$  approaches its maximum value of  $K^i/w_i$ , since  $K^b/w_a < K^a/w_a$ , then  $\partial\gamma(K^b/w_a)/\partial Y$  and  $\lim_{Y \rightarrow K^b} (\partial V^a/\partial Y)/(\partial V^b/\partial Y)$  are finite.

Different discount rates cause the government and individuals of type A to have different intertemporal preferences for income. The government places a higher value on the future than do individuals, so trade between them is possible. The government can offer type A individuals higher current utility and lower future utility while doing the reverse for type B (as compared with the solution when  $\delta = \rho$ ), while maintaining the self-selection constraints. Rapid discounting by individuals makes the early periods more important in the self-selection constraints; slower discounting by the government makes it place greater value on the lower distortions in the future. To see this, note that from eqs. (1a) and (1b),  $\partial V^a/\partial C_t^a = -\partial V^a/\partial Y_t^a = \mu/[\alpha + \lambda_b(\rho/\delta)^{t-1}]$ . It follows that the marginal utility of consumption rises over time indicating

that consumption declines. As  $t$  goes to infinity,  $\alpha \partial V^a / \partial C_t^a$  goes to  $\mu$ . For group B consumption,  $(1-\alpha)(\partial V^b / \partial C_t^b) = \mu + (\lambda_a)(\rho/\delta)^{t-1}(\partial V^a / \partial C_t^b)$ . Since  $\liminf C_t^b > 0$ , eventually  $\partial V^b / \partial C_t^b$  declines with  $C_t^b$  rising. As  $t$  goes to infinity,  $(1-\alpha)(\partial V^b / \partial C_t^b)$  goes to  $\mu$ . Thus, in the limit, the solution is not only Pareto optimal, but it is first-best in the sense of being the same as the solution to the one-period problem without self-selection constraints.

If  $\delta < \rho$ , the government values the early periods relatively more than do individuals. The optimal solution tends to have distortions on the less able's bundle increasing over time. Consider the constant path along which both groups receive the same bundles as in the one-period deterministic optimum every period. Even with different discount rates, this is feasible in (P). If the government adjusts the period  $t$  bundles in the direction of the first best, the value of its objective function increases, but the self-selection constraint on the more able would be violated. However, by making B's bundle in period  $t+1$  less attractive for A, the self-selection constraint could still be satisfied. Given the difference in discount rates, the value of the government's objective function grows with this procedure. While it is thus straightforward to show that the distortions increase over time, an exact characterization of the optimum is difficult since several possibilities exist.

Consider the identical separable utility functions in  $C$  and  $L$  above. Assume  $\lambda_a > 0$ . The left-hand side of (1c) can be written as  $[(\delta/\rho)^{t-1}(1-\alpha) - \lambda_a]\psi'(C_t^b) - \mu(\delta/\rho)^{t-1}$ . For sufficiently large  $t$ , both terms become negative and a corner solution results with  $C^b$  going to zero in finite time. For some  $\psi(\cdot)$ , this may be optimal. However, if  $\lim_{C \rightarrow 0} \psi(C) = -\infty$ , it cannot be an optimal solution. If  $C_t^b = 0$  at any finite  $t$ , the value of the government's objective function would be  $-\infty$ . Instead,  $\lambda_a$ , which is constant over time for fixed  $M$ , will be such that for all  $t \leq M$ ,  $(\delta/\rho)^{t-1}(1-\alpha) - \lambda_a > 0$ . Clearly,  $\lambda_a$  must then decline with  $M$ . For large  $M$ ,  $\lambda_a$  will be close to zero. In this case, in the early periods the solution will be approximately first best. Type A individuals get an undistorted bundle at each  $t$  with  $C_t^a$  rising and  $Y_t^a$  declining as  $t$  increases. B's bundle starts at a point with almost no distortion and  $MRS^b(C_t^b, Y_t^b)$  diverges from 1.

Our formulation assumed identical discount rates for the two types. If  $\delta > \rho_a > \rho_b$  or  $\delta \geq \rho_b > \rho_a$ , the solution for  $\delta > \rho$  still applies, since with  $\lambda_a > 0$ , only A's discount factor appears in the first-order conditions. If  $\rho_a > \delta$ , the solution would be similar to that with  $\rho_a = \rho_b > \delta$ .

## 5. Conclusions

Our results indicate that, when the government respects individual discount rates, only in a weak sense does the optimal tax system incorporate any information about individuals learned from their responses over time. First, if the government is able to randomize in each period, then no benefit

is gained by keeping track of what individuals have earned in past periods. A lottery can be offered in each period, independent of other periods, satisfying self-selection constraints and yielding the best possible outcome. Second, if the government cannot randomize directly, then it can duplicate randomization by intertemporal nonstationarity. Such nonstationarity requires that the government keep track of individuals' past behavior since, after the first period, some individuals would like to choose different bundles than those assigned to them. Third, the information on past earning experience does not yield systematic increases in the value of the government's objective function. The time pattern of tax schedules is not motivated by attempts to gather information about abilities, but is designed to provide lifetime utilities consistent with the self-selection constraints. Only the first-period choice by individuals is relevant information; after that, individuals in each period are in effect assigned a particular bundle.

By contrast, if the government discounts the future at a different rate from individuals, then there is a systematic change in the bundles given to individuals over their lifetimes. In the limit, the distortions may be eliminated. This arises because the different intertemporal preferences of the government and individuals leaves room for 'trade' between them.

In our formal results, we have only considered a single intertemporal government budget constraint. Another possibility is that the government cannot borrow or lend, just as individuals are constrained in our model. In this case, nonstationary tax policies may still be desirable, but the conditions are more restrictive. It remains true that nonstationarity is only desirable if randomization is desirable in the one-period problem. However, it may also be necessary to give the type whose bundle is not envied different bundles over time (in contrast to Theorem 3). If that type has constant (or nearly constant) marginal utility of consumption, its welfare is only slightly reduced by the time-varying consumption path. The nonstationarity for that group takes the place of storage or of trade with the rest of the world. The period-by-period budget constraints in general prevent the normalized Pareto frontier from coinciding with the one-period frontier with randomization. If the government cannot borrow but can save resources until later periods, there would be restrictions on the order in which different bundles could be assigned. In the limiting case of no discounting and a sufficiently long horizon, the temporal order in which the bundles are given could be changed to satisfy a no-borrowing constraint without affecting lifetime utility.

Our results show that nonstationarity over time and randomization within each period can substitute for each other in the optimal intertemporal income tax. It is not clear which approach is preferable since each has some advantages.

First, randomization and nonstationarity are not perfect substitutes. Even if the government has a single intertemporal budget constraint, nonstationarity

is guaranteed to do as well as randomization only with an infinite horizon and a sufficiently large discount factor. If the government has a separate budget constraint in each period, the optimal randomization cannot be completely duplicated by intertemporal variability, so that randomization along with nonstationarity would be needed to reach this Pareto frontier.

Second, political and administrative difficulties could prevent implementation of either method. On the one hand, the government may be reluctant to incorporate randomization explicitly in the tax code. This is especially true since optimal randomization requires individuals to declare their type and then receive at random a tax schedule before choosing their labor supplies. The optimal randomization generally cannot be implemented by random collection or enforcement after labor supply decisions. On the other hand, intertemporal nonstationarity requires keeping track of past incomes to determine individuals' current tax payments. However, this is simplified since the government needs only to recall each individual's type as revealed by past decisions.

Third, to increase acceptance of the tax system by society, it is desirable that the system be fair and also be perceived as fair. A standard notion of fairness is horizontal equity, i.e. that individuals in the same circumstances be treated the same. Randomization satisfies horizontal equity *ex ante* but not *ex post*. Before the random selection, all individuals of the same type face the same lottery. After receiving a random draw of tax functions, individuals of the same type will be induced to choose bundles which need not yield the same utility. Intertemporal nonstationarity achieves horizontal equity both *ex ante* and *ex post* in each period. Individuals of the same type are induced to choose the same bundles as each other in every period even though the choice varies over time.

Fourth, both procedures induce differences in the bundles chosen by individuals of a type either within a period in an expected sense under randomization or over time under nonstationarity. Since the utility functions are strictly concave, individuals would desire to reduce these differences. Under randomization, individuals might gain by purchasing insurance counteracting the randomness in the tax system. If such policies were forbidden, then similar effects could be achieved by trades with other individuals of the same type. For the same reason, under nonstationarity, individuals would desire to smooth consumption and leisure over time by saving or borrowing. Saving or insurance serves to counteract the weakening of self-selection constraints which motivated the asymmetry of bundles in the first place. The ability to save or buy insurance will be a factor in the decision to reveal one's type truthfully. The choice between nonstationarity or randomness may depend upon whether it is easier to prevent saving or insurance. If saving and insurance are desirable for other reasons or cannot be prevented, then simple repetition of the solution to the nonrandom one-period problem may

be the best feasible solution. However, the opposite problem arises if only symmetric solutions are allowed when individuals have nonconvex opportunity sets. Individuals might randomize consumption bundles to convexify budget sets; to attain self-selection, more distortionary taxation would have to be imposed. Prohibiting gambling would, under these circumstances, increase welfare.

To focus on the informational aspects of the dynamic structure, the model was intentionally kept simple. Preferences and production possibilities were nonstochastic and identical in every period. Thus, no motivation for nonstationary taxation other than relaxing self-selection constraints exists. Nonstationarity of the tax schedules is the only incentive to save or borrow. Several important features that could influence the design of optimal dynamic tax schedules were left out. A more complete model could incorporate life-cycle features. If wages first rose and then fell with age, nonstationary tax schedules could relax borrowing constraints on individuals. In addition, future income might be subject to exogenous uncertainty. In our framework, after learning an individual's type in the first period, the government could pin an individual down to a particular bundle in any later period by punishing deviations. If income were random, these punishments would be significantly harder to implement.

### Appendix: Proofs of theorems

*Proof of Theorem 1.* Let

$$\bar{C}^i(\alpha) = \sum_{t=1}^M \delta^{t-1} C_t^i(\alpha) \Big/ \sum_{t=1}^M \delta^{t-1}$$

and

$$\bar{Y}^i(\alpha) = \sum_{t=1}^M \delta^{t-1} Y_t^i(\alpha) \Big/ \sum_{t=1}^M \delta^{t-1}, \quad i = a, b.$$

Since  $V^i(C^i, Y^i)$  is strictly concave,  $V^i(\bar{C}^i(\alpha), \bar{Y}^i(\alpha)) \geq \bar{V}^i(\alpha)$ , with strict inequality if the bundles are not stationary. If

$$\sum_{t=1}^M \rho^{t-1} V^i(C_t^j(\alpha), Y_t^j(\alpha)) \Big/ \sum_{t=1}^M \rho^{t-1} = \bar{V}^i(\alpha), \text{ then } V^i(\bar{C}^j(\alpha), \bar{Y}^j(\alpha)) \geq \bar{V}^i(\alpha),$$

with strict inequality if  $(C_t^j(\alpha), Y_t^j(\alpha))$  is nonstationary, also follows from strict concavity. Assume, without loss of generality, that  $\bar{V}^a(\alpha) > V^{a0}$ , so that



$V^a(\bar{C}^a(\alpha), \bar{Y}^a(\alpha)) \geq V^{a0}$  and  $V^a(\bar{C}^b(\alpha), \bar{Y}^b(\alpha)) > V^{a0}$  must hold. Since the indifference curve  $V^{a0}$  is tangent to the 45° line ( $C = Y$ ),  $\bar{C}^a(\alpha) > \bar{Y}^a(\alpha)$  and  $\bar{C}^b(\alpha) > \bar{Y}^b(\alpha)$ . Thus,  $\bar{C}^a(\alpha) - \bar{Y}^a(\alpha) + \bar{C}^b(\alpha) - \bar{Y}^b(\alpha) > 0$  must hold. But,

$$\begin{aligned} & \sum_{t=1}^M \rho^{t-1} \{C_t^a(\alpha) - Y_t^a(\alpha) + C_t^b(\alpha) - Y_t^b(\alpha)\} \Big/ \sum_{t=1}^M \rho^{t-1} \\ &= \bar{C}^a(\alpha) - \bar{Y}^a(\alpha) + \bar{C}^b(\alpha) - \bar{Y}^b(\alpha), \end{aligned}$$

so the budget constraint is violated unless A's self-selection constraint is satisfied with strict inequality. Q.E.D.

*Proof of Theorem 2.* There exists  $\alpha_0$  such that  $V^i(\alpha_0) = V^{i0}$ ,  $i = a, b$ . For any  $\alpha \neq \alpha_0$ , either  $V^a(\alpha) > V^{a0}$  or  $V^b(\alpha) > V^{b0}$ . Then from Theorem 1, one type cannot have a binding self-selection constraint. Q.E.D.

*Proof of Theorem 3.* Since type  $i$  does not envy type  $j$ 's sequence of bundles, if  $MRS^j(C_t^j(\alpha), Y_t^j(\alpha)) \neq 1$ , we can adjust  $j$ 's bundle and free up additional resources without harming  $j$ . Thus,  $MRS^j(C_t^j(\alpha), Y_t^j(\alpha)) = 1$ ,  $t = 1, \dots, M$ . If the sequence of bundles assigned to  $j$  were nonstationary, then given that  $V^j$  is strictly concave,  $j$  would prefer a move toward a stationary bundle with the same present discounted utility. A small move in that direction would maintain budget balance and would not violate  $i$ 's self-selection constraint. Thus, assigning a nonstationary bundle could not have been optimal. Q.E.D.

*Proof of Theorem 4.* Since  $\delta = \rho$ , these terms cancel from the first-order conditions in (1). Assume that  $\lambda_a > 0$ . For any  $(C_t^b, Y_t^b)$ , the FOC are:

$$(1 - \alpha)V_c^b(C_t^b, Y_t^b) - \lambda_a V_c^a(C_t^b, Y_t^b) - \mu = 0,$$

$$(1 - \alpha)V_y^b(C_t^b, Y_t^b) - \lambda_a V_y^a(C_t^b, Y_t^b) + \mu = 0.$$

Eliminating  $\mu$  and combining terms yields:

$$(1 - \alpha)V_c^b(C_t^b, Y_t^b)[1 - MRS(C_t^b, Y_t^b)] = \lambda_a V_c^a(C_t^b, Y_t^b)[1 - MRS^a(C_t^b, Y_t^b)].$$

Adding  $\lambda_a V_c^a(C_t^b, Y_t^b)MRS^b(C_t^b, Y_t^b)$  to both sides, substituting the first-order condition for  $C_t^b$ , and combining terms yields:

$$\mu[1 - MRS^b(C_t^b, Y_t^b)] = \lambda_a V_c^a(C_t^b, Y_t^b)[MRS^b(C_t^b, Y_t^b) - MRS^a(C_t^b, Y_t^b)].$$

Since  $\lambda_a > 0$  (by assumption) and  $\mu > 0$ , the result follows immediately. Q.E.D.

*Proof of Theorem 5.* A deterministic solution to the one-period problem can be duplicated as a stationary solution in the dynamic problem by assigning the solution in the one-period problem in every period of (P). A nonstationary solution can always be duplicated by a randomized solution in the one-period problem. Assign any bundle with a probability equal to the sum of the discount weights which that bundle received in the nonstationary solution. If a nonstationary solution dominates the optimal stationary one, then it dominates any feasible stationary solution. It then follows that the random analogue of the nonstationary solution must dominate the one-period problem which dominates the optimal deterministic solution. Q.E.D.

*Proof of Theorem 6.* The proof requires the following lemma.

*Lemma 1.* Consider an infinite sequence defined by  $(1-x)x^{t-1}$ ,  $t=1, \dots, \infty$ . Consider three numbers  $\Pi_1, \Pi_2, \Pi_3$  with  $\Pi_i > 0$  and  $\Pi_1 + \Pi_2 + \Pi_3 = 1$ . If  $\frac{2}{3} \leq x < 1$ , then there exists a partition,  $I_i$ , of the positive integers indexing the terms of the sequence such that

$$(1-x) \sum_{t \in I_i} x^{t-1} = \Pi_i, \quad i=1, 2, 3.$$

*Proof.* Construct the desired partition by the following procedure. Define  $\Omega_1^r, \Omega_2^r$  and  $\Omega_3^r$  as the sums of the terms in each partition using the first  $r$  terms of the sequence. For  $r=1$ , one of  $\Omega_i^1 = 1-x$  and the rest are zero. Let  $S_r \equiv \Omega_1^r + \Omega_2^r + \Omega_3^r = \sum_{t=1}^r (1-x)x^{t-1}$ . By summing, it follows that  $1-S_r = x^r$ . If  $x \geq \frac{2}{3}$  then for all  $r$ ,  $(1-x)x^{r-1} \leq \frac{1}{3}x^{r-1} = \frac{1}{3}(1-S_{r-1})$ . That is, the  $r$ th term is less than one-third the sum of the remaining terms, including itself. Thus, the  $r$ th term can always be put into at least one of the partial sums without exceeding that  $\Pi_i$ . Then for all  $r$ , the partitions can be formed with  $\Pi_i - \Omega_i^{r-1} \geq 0$ ,  $i=1, 2, 3$ . As  $r$  goes to infinity,  $\sum_{t=1}^r (1-x)x^{t-1}$  goes to one so that  $\Omega_1^r + \Omega_2^r + \Omega_3^r$  goes to 1. It then follows that  $\Omega_i^r$  goes to  $\Pi_i$  as  $r$  goes to infinity. Q.E.D.

*Proof of Theorem.* From the lemma, if three or fewer distinct bundles arise in the optimal solution, a discount factor greater than or equal to  $\frac{2}{3}$  insures that any probabilities for these bundles are attainable. The optimal random solution to the one-period problem requires at most three distinct bundles. The maximization problem for the government is a linear programming problem in the probabilities, viewing the bundles as fixed. There are at most

three binding constraints: one of the self-selection constraints; the government budget constraint; and the constraint that the probabilities sum to one. There exists a solution with the number of probabilities not equal to zero no greater than the number of binding constraints.

First, for any  $\alpha$ , the random one-period solution can be duplicated in (P). Let  $k(j)$  equal the number of bundles in the lottery offered to type  $j$ . For some  $j$ ,  $k(j)=1$  with  $(C^{j1}, Y^{j1})$  received by  $j$  with certainty, while for  $i \neq j$ ,  $k(j) \leq 3$  can hold at an optimal solution. Hence, there exist  $(C^{ih}, Y^{ih})$  and  $\pi_{ih}$ ,  $h=1, 2, 3$ , in an optimal solution. Consider (P) and multiply the objective function, the self-selection constraints, and the budget constraint by the constant  $(1/\sum_{t=1}^{\infty} \rho^{t-1})=1-\rho$ . This leaves the solution unchanged. From Lemma 1, there is a partition of the integers such that  $\sum_{t \in I_h} (1-\rho)^{t-1} = \pi_{ih}$ ,  $h=1, 2, 3$ . Then assign  $(C^{ih}, Y^{ih})$  to  $i$  in all periods  $t \in I_h$ ,  $h=1, 2, 3$ , and assign  $(C^{j1}, Y^{j1})$  to  $j$  in all periods. By construction, this is feasible in (P) since it is feasible in the one-period problem.

Second, this solution is optimal in (P). If it were not optimal, then there would exist a feasible lifetime bundle  $(\hat{C}^i, \hat{Y}^i)$ ,  $i=a, b$ , such that

$$\begin{aligned} & \left[ \alpha_i \sum_{t=1}^{\infty} \rho^{t-1} V^i(\hat{C}_t^i, \hat{Y}_t^i) + \alpha_j \sum_{t=1}^{\infty} \rho^{t-1} V^j(\hat{C}_t^j, \hat{Y}_t^j) \right] (1-\rho) \\ & > \alpha_i \sum_{h=1}^3 \sum_{t \in I_h} \rho^{t-1} (1-\rho) V^i(C^{ih}, Y^{ih}) + \alpha_j V^j(C^{j1}, Y^{j1}) \\ & = \alpha_i \sum_{h=1}^3 \pi_{ih} V^i(C^{ih}, Y^{ih}) + \alpha_j V^j(C^{j1}, Y^{j1}), \quad i=a, b, \text{ and } j \neq i. \end{aligned}$$

Since  $\rho < 1$  and since  $V^a(C^a, Y^a)$  and  $V^b(C^b, Y^b)$  have upper bounds, given the bounds on  $C^a$  and  $Y^a$ , there would exist a finite  $T$  such that

$$\begin{aligned} & \left[ \alpha_i \sum_{t=1}^T \rho^{t-1} V^i(\hat{C}_t^i, \hat{Y}_t^i) + \alpha_j \sum_{t=1}^T \rho^{t-1} V^j(\hat{C}_t^j, \hat{Y}_t^j) \right] (1-\rho) \\ & > \alpha_i \sum_{h=1}^3 \pi_{ih} V^i(C^{ih}, Y^{ih}) + \alpha_j V^j(C^{j1}, Y^{j1}). \end{aligned} \tag{A.1}$$

Since  $\sum_{h=1}^3 \sum_{t \in I_h} \rho^{t-1} < 1/(1-\rho)$ , the inequality (A.1) will still hold if the left-hand side is multiplied by  $1/[(1-\rho) \sum_{t=1}^T \rho^{t-1}]$ . Similarly, for  $T$  large enough, a self-selection constraint which held with inequality will still hold

with inequality, while one that held with equality will be violated by, at most, an arbitrarily small  $\varepsilon$ . Similarly, the resource balance will be violated by, at most, an arbitrarily small  $\varepsilon$ . If the self-selection constraint which is violated is that for group  $i$ , then there exists an  $\eta > 0$ , such that substituting  $\hat{Y}_t^j = \hat{Y}_t^j + \eta$ , for  $\hat{Y}_t^j$ , will have all constraints satisfied and still leave (A.1) satisfied. Then the lotteries  $((\hat{C}_t^a, \hat{Y}_t^i), \pi_{it})$  and  $((\hat{C}_t^j, \hat{Y}_t^j), \pi_{jt})$  with  $\pi_{it} = \pi_{jt} = \rho^{t-1} / \sum_{t=1}^T \rho^{t-1}$  are feasible in the one-period problem and yield a higher value than the optimum, which is a contradiction.

This shows existence by construction of an optimal solution. Since the solution duplicates that of the one-period problem, the remainder of the theorem follows immediately. Q.E.D.

*Proof of Theorem 7.* When  $M$  is finite or  $\delta < \frac{2}{3}$ , only a subset of the probability simplex  $\sum_{h=1}^3 \pi_{th} = 1$  can be achieved by intertemporal variation in the optimal bundles. If the optimal lotteries in the one-period problem involve probabilities in the attainable subset, then the optimal solution can be duplicated. If not, but a local randomization does improve the deterministic outcome, then, as shown in Brito et al. (1989), a local randomization with any probabilities improves on the deterministic solution. Thus, randomization with probabilities from the attainable part of the simplex will improve on the deterministic solution and these can be duplicated by nonstationary solutions in (P). For  $M$  sufficiently large but finite, a nonstationary solution can come sufficiently close to the nonrandom solution to improve on the best nonstationary solution. When some nonstationarity is desirable, but the optimal random solution cannot be duplicated, the outcome on the normalized Pareto frontier will be interior to that in the one-period problem. Q.E.D.

*Proof of Theorem 8.* Given  $\lambda_b = 0$ , eqs. (1a) and (1b) yield  $MRS^a(C_t^a, Y_t^a) = 1$ , for all  $t$ . Divide eq. (1c) by  $\delta^{t-1}(\partial V^b / \partial C_t^b)$  and (1d) by  $\delta^{t-1}(\partial V^b / \partial Y_t^b)$ .

$$(1 - \alpha) - \lambda_a(\rho/\delta)^{t-1} [(\partial V^a / \partial C_t^b) / (\partial V^b / \partial C_t^b)] - \mu / (\partial V^b / \partial C_t^b) = 0, \\ t = 1, \dots, M, \quad (\text{A.2a})$$

$$(1 - \alpha) - \lambda_b(\rho/\delta)^{t-1} [(\partial V^a / \partial Y_t^b) / (\partial V^b / \partial Y_t^b)] + \mu / (\partial V^b / \partial Y_t^b) = 0, \\ t = 1, \dots, M. \quad (\text{A.2b})$$

Since  $\partial V^i / \partial C$  is finite for  $C > 0$  and  $\lim_{C \rightarrow 0} (\partial V^a / \partial C) / (\partial V^b / \partial C)$  is finite,  $\lim_{t \rightarrow \infty} \sup (\rho/\delta)^{t-1} [(\partial V^a / \partial C_t^b) / (\partial V^b / \partial C_t^b)] = 0$ . Hence, should  $\lim_{t \rightarrow \infty} \sup (\partial V^b / \partial C) = \infty$ , then after some  $\bar{t}$ , the left-hand side of (A.2a) would be strictly positive, violating the first-order condition. Thus, no  $C_t^b$  sequence goes to zero,

guaranteeing that  $\lim_{t \rightarrow \infty} \inf C_t^b > 0$ . Similarly, from (A.2b),  $\lim_{t \rightarrow \infty} \sup Y_t^b < K^b$ . Therefore,

$$\lim_{t \rightarrow \infty} \sup (\rho/\delta)^{t-1} (\partial V^b / \partial C_t^b) = \lim_{t \rightarrow \infty} \sup (\rho/\delta)^{t-1} (\partial V^b / \partial Y_t^b) = 0.$$

Given this, divide (1c) and (1d) by  $\delta^{t-1}$  and solve for

$$MRS^b(C_t^b, Y_t^b) = -(\partial V^b / \partial Y_t^b) / (\partial V^b / \partial C_t^b).$$

$$MRS^b(C_t^b, Y_t^b) = \frac{\mu - \lambda_a (\rho/\delta)^{t-1} (\partial V^a / \partial Y_t^b)}{\mu + \lambda_a (\rho/\delta)^{t-1} (\partial V^a / \partial C_t^b)}, \quad (\text{A.3})$$

then  $\lim_{t \rightarrow \infty} MRS^b(C_t^b, Y_t^b) = 1$ . Q.E.D.

## References

- Arnott, R.J. and J.E. Stiglitz, 1988, Randomization with asymmetric information: A simplified exposition, *Rand Journal of Economics* 19, 344–362.
- Baron, D. and D. Besanko, 1984, Regulation and information in a continuing relationship, *Information Economics and Policy* 1, 267–302.
- Brito, D., J. Hamilton, S. Slutsky and J. Stiglitz, 1989, Randomization in optimal income tax schedules, University of Florida Working Paper in Economics 89-6.
- Brito, D., J. Hamilton, S. Slutsky and J. Stiglitz, 1990, Pareto efficient tax structures, *Oxford Economic Papers* 42, 61–77.
- Fellingham, J.C., Y.K. Kwon and D.P. Newman, 1984, Ex ante randomization in agency models, *Rand Journal of Economics* 15, 290–301.
- Harris, C., 1987, On strategic taxation, *Nuffield College Discussion Papers in Economics*, no. 7.
- Laffont, J.-J. and J. Tirole, 1988, The dynamics of incentive contracts, *Econometrica* 56, 1153–1175.
- Maskin, E. and J. Riley, 1984, Optimal auctions with risk-averse buyers, *Econometrica* 52, 1473–1518.
- Mirrlees, J.A., 1971, An exploration in the theory of optimum income taxation, *Review of Economic Studies* 38, 175–208.
- Roberts, K.W., 1984, The theoretical limits to redistribution, *Review of Economic Studies* 51, 177–195.
- Sadka, F., 1976, On income distribution, incentive effects and optimal income taxation, *Review of Economic Studies* 43, 261–268.
- Stiglitz, J.E., 1982, Self-selection and Pareto efficient taxation, *Journal of Public Economics* 17, 213–240.
- Weiss, L.W., 1976, The desirability of cheating incentives and randomness in the optimal income tax, *Journal of Political Economy* 89, 1343–1352.