Journal of Public Economics 12 (1979) 1–26. © North-Holland Publishing Company

THE CONCEPT OF OPTIMAL TAXATION IN THE OVERLAPPING-GENERATIONS MODEL OF CAPITAL AND WEALTH

J. A. ORDOVER*

New York University, New York, NY 10003, USA

E. S. PHELPS*

Columbia University, New York, NY 10027, USA

Received March 1979, revised version received April 1979

The optimal fiscal program of the present generation may be conceived as maximizing the social welfare of that generation subject to a capital-accumulation quota and a wealth-accumulation limit (the difference being the stock of public debt). The standard methods of optimal tax analysis are shown to be applicable to the multiperiod generation if (1) the capital-wealth target is Pareto efficient in the sense that no other target would permit a mutual welfare gain for the present generation and the next, and (2) the government 'budget constraint' specifies an algebraic budgetary deficit consistent with the attainment of the target. Important results on interest-income taxation are derived.

1. Introduction

Every meaningful welfare analysis of the fiscal choices before the current generation of workers and savers must somehow take into account the impact that each tax structure would have upon the productive capacity made available to the next generation and upon the after-tax wealth-claims to retirement consumption which are to be set against that future capacity. The 'maximin' tax analysis by Ordover and Phelps (1975) of a simple overlapping-generations model of heterogeneous workers, compounded from the settings in Diamond (1965) and Mirrlees (1971), and the corresponding numerical simulation analysis by Jha (1978), met this problem by requiring that the present generation endow the next one with at least the same capital stock per worker as it enjoys and burden the next generation with at most the same wealth per worker to be paid to the retired as it must pay.[1] Then

*Ordover acknowledges a grant from the Sloan Foundation and Phelps a fellowship from the Guggenheim Foundation. We are also grateful to the National Science Foundation for financial support. We would like to thank the editor of the *Journal of Public Economics* for his very helpful comments.

[1]A modification of this requirement was considered by Ordover (1976) and additionally in our joint paper; the initial per worker quantity of wealth claims, and hence the per worker public debt held by the retired population, could be chosen by the present working generation subject to the condition that this quantity as well as the per worker quantity of capital be replicated for the next generation.

the next generation, not being damaged, could not complain – or so we thought. Yet the next generation might well complain if the current generation in maintaining both capital and wealth at their initial levels thereby missed opportunities for a mutual welfare gain.

A traditional solution to the problem treats the choice of capital and wealth as an exercise in dynamic programming to achieve the optimal growth of social welfare from generation to generation. (Of course, the optimal path of capital and wealth will leave no unexploited opportunity for a mutual welfare gain if the intergeneration optimality criterion is of the Paretian type.) Two papers representing this classical approach, both on fiscal planning for 'maximin growth' in the overlapping-generations setting, found a Pareto-like condition that capital and wealth will necessarily satisfy on the intergenerationally optimal path. In their study of homogeneous populations of workers, Phelps and Riley (1978) showed that the marginal rate of substitution for the *next* generation between the capital left it by the retired generation, k, and the wealth-claim to consumption by the retired generation, x, is equal to the gross marginal product of that capital; to this number the *present* generation must equate *its* marginal rate of substitution between those same two variables for the growth of capital and wealth to be maximin-optimal over generations. In his subsequent study of heterogeneous worker populations, Phelps (1976) confirmed that this same marginal product provides the parametric shadow rate of interest, r^*, with which the fiscal planners in the present generation must discount its future consumption claim, x, in relation to the capital, k, it deposits at the end of its labors – and discount its future tax revenue from interest-income levies in relation to its present tax revenue from current wage-income levies. The present generation optimally taxes as if it were maximizing its own (maximin) social welfare subject to a single constraint on capital and wealth $(1+r^*)^{-1}x-k=\Delta^*$ $\equiv$constant. Here the second parameter Δ^* may be interpreted as the discounted value of the public debt to be sold to the present generation's savers along the maximin-optimal path; together with initial conditions k_{-1} and x_{-1}, it yields the optimal algebraic budgetary deficit to be allowed the current generation.

The present paper is a further contribution to the welfare economics of taxation in a dynamic setting – namely the overlapping-generations models with heterogeneous labor recalled in section 2. Our concept of optimal taxation is first developed in section 3 where we attribute to each generation any intragenerational social welfare function of the Bergson–Samuelson type; the same concept is redeveloped using the 'maximin' social welfare function in section 4. Two key elements of our concept of optimality may be noted. We do not require that the path of capital and wealth be maximin-optimal, or anything-optimal, over generations, nor do we return to our original requirement that capital and wealth per head be maintained; instead, we permit the present generation to set (within the limits of feasibility) an arbitrary target for the capital and wealth levels that its fiscal plan is constrained to meet; but we often require that the present generation confine its target to those consistent with the achievement of *Pareto efficiency* between it and the next generation – always taking as given whatever the next generation is (known to be) going to do toward its successors. We also permit each generation to employ *graduated* (nonlinear) taxation of its wage and interest incomes, making use of the Hamiltonian methods developed by Mirrlees (1971) and first applied to life-cycle worker–savers by Atkinson and Stiglitz (1976).

A methodological implication of our concept is the possibility of representing any pair of constraints, one on capital and the other on wealth, by a single linear constraint containing a shadow rate of interest and prescribed public debt; but that shadow interest rate is not generally the market interest rate. Those analyses of the taxation of life-cycle saving and interest that purport to hold for an arbitrary budgetary deficit and fixed rate of interest, such as the otherwise exemplary analysis by Atkinson and Stiglitz, are valid provided the algebraic deficit allowed and interest rate imposed are both associated with a target outcome (k, x) that is Pareto efficient; at any other target the shadow interest rate in the associated linear constraint, $(1+r^*)^{-1}x - k = \Delta^*$, will differ from the market interest rate with the consequence that the Atkinson–Stiglitz tax schedules will be inoptimal.

In section 5 we use the preceding conceptual apparatus to engage (again) in the mounting controversy over the taxation of interest income. It is first shown that the Phelps–Sadka injunction against a positive marginal tax rate 'at the top' extends to interest and wealth as well as wages provided that our Pareto-efficiency condition is met. The other issue is the optimality of interest-income taxation in general. Our 1973 and 1976 'maximin' tax analyses reduced the issue to the question: would some taxation of interest or wealth make the current generation's discounted tax revenue higher than otherwise possible? If so, no tax exemption of interest or wealth would be maximin-optimal for the present generation, since maximin taxation maximizes discounted tax revenues. This result is reaffirmed. But we also confirm a proposition of Atkinson and Stiglitz when properly qualified: under our Pareto-efficiency condition but not otherwise, it is Bergson–Samuelson optimal and maximin-optimal to exempt all interest income from tax if each worker's utility function is everywhere weakly separable between the two consumption goods (early and late) and leisure.

Yet section 6 concludes with a warning. Despite separability, exemption of interest income from tax may reduce the potential welfare of the next generation if unaccompanied by other tax-transfer changes; it may reduce the potential welfare of the present generation if the latter binds itself to keep the budget balanced.

2. The overlapping-generations model

The labor heterogeneity in the model is taken from Mirrlees (1971). The effectiveness of a person's labor is graded by an index m. The effective labor supplied by a worker of grade m, denoted l^m, is the worker's effort or labor-time, e^m, augmented by the factor m, i.e. $l^m = me^m$. Let us denote by $\Phi_v(m)$ the proportion of persons in generation v whose labor is m or less, $0 \leqq m \leqq M_v$, agreeing henceforth to leave the subscript v implicit. Then the postulated distributions of m all have the following properties:

$$0 \leqq \Phi(0) < 1 = \Phi(M), \qquad 0 < M \leqq \infty,$$

$$0 \leqq \Phi(a) \leqq \Phi(b), \quad \text{if } 0 \leqq a < b \leqq M.$$

The two-period life-cycle of overlapping generations is taken from Diamond (1965). In any generation v, $v \geqq 0$, every worker retires from paid work after the first period of economic life and no one bequeathes. The amount worked in the first period by a person born into generation v with labor of grade m is denoted by e_v^m, the consumption at the end of the first period is denoted by c_v^m, and the amount consumed at the end of the second and last period is called x_v^m. The corresponding amounts *per person* in generation v are

$$e_v = \int_0^M e_v^m \, \mathrm{d}\Phi(m); \qquad c_v = \int_0^M c_v^m \, \mathrm{d}\Phi(m); \qquad x_v = \int_0^M x_v^m \, \mathrm{d}\Phi(m).$$

The total effective labor supplied per person of generation v is then

$$l_v = \int_0^M l_v^m \, \mathrm{d}\Phi(m) = \int_0^M m e_v^m \, \mathrm{d}\Phi(m).$$

A neoclassical, one-product, constant-return-to-scale production function, F, is posited that makes output per person of any generation v depend on effective labor per person, l_v, and the predetermined per person quantity of capital, k_{v-1}, left by the previous generation, $v = 1, 2, 3, \ldots$. With absolutely no loss of generality we take the generations to be of unchanging size. Then the capital stock which generation v leaves per person of generation $v+1$ is the unconsumed output per person of generation v:

$$k_v = F(k_{v-1}, l_v) - c_v - x_{v-1} \geqq 0, \qquad v = 1, 2, \ldots. \tag{1}$$

It is assumed that both partial derivatives of the net production function, $Y(k_{v-1}, l_v) \equiv F(k_{v-1}, l_v) - k$, exist with $Y_k \geqq -1$ and $Y_l \geqq 0$.

All members of generation v receive a first-period demogrant, g_v, and a further installment, h_v, upon retirement. We assume that first-period consumption is untaxed: its consumer price is equal to the producer price and both are equal to one. All 'direct' taxation falls only on wage income, and indirect taxation falls only on second-period consumption; the latter taxation operates like a levy on saving or on interest income. For simplicity we require households to prepay their indirect taxes: a household that consumes x_v^m in its second period must pay $T_s(x_v^m - h_v)$ in indirect taxes in the first period; then there are no second-period revenue collections which need discounting back to the first period. The budget equation for household of type m in the extensive form is

$$x_v^m = h_v + (1 + r_v)\sigma_v^m, \tag{2a}$$

$$c_v^m = g_v + w_v l_v^m - T_l(w_v l_v^m) - T_s(x_v^m - h_v) - \sigma_v^m. \tag{2b}$$

The function $T_l(w_v l_v^m)$ gives the tax paid on earned income, and σ_v^m is the amount saved. Furthermore, w_v is the before-tax wage rate paid for one unit of effort by workers of the standard ($m = 1$) type, and r_v is the rate of return to private saving. For simplicity we suppose that w_v and r_v are equal to the corresponding marginal products:

$$w_v = F_l(k_{v-1}, l_v) > 0; \qquad 1 + r_v = F_k(k_v, l_{v+1}) \geqq 0. \tag{3}$$

There is a common utility function $u(c^m, x^m, e^m)$ which does not depend on m. Each household selects the triplet $\{c^m, x^m, e^m\}$ to maximize its utility subject to the individualized budget constraint (2) corresponding to its m. Formally, the household's problem is:

$$\max_{\{c^m, x^m, e^m\}} \; u(c^m, x^m, e^m) \tag{4}$$

subject to

$$wme^m + g + (h/F_k) - c^m - (x^m/F_k) - T_l(wme^m) - T_s(x^m - h) \geqq 0,$$

where we have omitted the generational subscript v. The first-order necessary conditions for this maximization program are:

$$e^m[u_e + \lambda wm(1 - T'(wme^m))] = 0, \tag{5a}$$

$$u_c = \frac{-u_e}{wm(1 - T_l')}, \tag{5b}$$

$$u_x = \frac{-u_e(1/F_k + T_s')}{wm(1-T_l')}, \tag{5c}$$

$$wme^m + g + (h/F_k) - c^m - (x^m/F_k) - T_l(\cdot) - T_s(\cdot) = 0. \tag{5d}$$

Conditions (5b) and (5c) are satisfied only for those who work. For those who strictly prefer full-time leisure to work, the term $-u_e/wm(1-T')$ must be replaced by λ. A simple envelope argument establishes that the *maximum direct utility function,*

$$u^*(m) \equiv \max u(wme^m + g + (h - x^m)/F_k - T_l(\cdot) - T_s(\cdot), x^m, e^m), \tag{6}$$

must satisfy the differential equation

$$\frac{\mathrm{d}u^*(m)}{\mathrm{d}m} = \frac{-u_e e^m}{m} = \frac{-u_e l^m}{m^2} \geqq 0. \tag{7}$$

If we take $u^*(m)$, x^m, and e^m as given, we can invert eq. (6) to obtain

$$c^m = z(u^*(m), x^m, e^m). \tag{8}$$

Eqs. (7) and (8) completely summarize the household sector.

The income statement of the v-government – the government that taxes generation v, that is – is given by

$$g_v + \frac{h_v}{F_k(k_v, l_{v+1})} = \int_0^M [T_l(w_v m e_v^m) + T_s(x_v^m - h_v)]\,\mathrm{d}\Phi(m) + (\Delta_v - d_{v-1}) + [F_k k_{v-1} - (x_{v-1} - d_{v-1})]. \tag{9}$$

Here d_{v-1} is the *face value* of the one-period discount-bonds issued by the government in period $v-1$ and Δ_v is the *market value* of the bonds issued in period v. If a bond entitles its owner to one unit of second-period consumption after tax (i.e. upon the payment of the appropriate tax), then, by arbitrage, its market value upon issue at the end of period v is $1/F_k(k_v, l_{v+1})$. This result reveals an interdependence between generations missing from the static Atkinson–Stiglitz model: the market value of a given debt issue by the v-government is a (decreasing) function of the effective labor supplied by generation $v+1$; and the v-generation supplies of labor, l_v^m, since they are not generally independent of $1+r_v$ given tax schedules, will also be some function of effective labor supplied by generation $v+1$. The presence of the last term in square brackets of the right-hand side reflects the fact that the v-government collects a surtax (positive or negative) from the $v-1$ generation equal to the excess of total earnings from capital over the claim that the $v-1$ generation has to those earnings. This surtax siphons off any windfall gain to that generation arising from the effects on earnings of unanticipated tax reform by generation v.

3. Optimal taxation: Bergson–Samuelson

Here we posit, for purposes of comparison with Atkinson and Stiglitz, that the v-government maximizes the Bergson–Samuelson social welfare function

$$\mathscr{W}^v = \int_0^M G^v[u(c_v^m, x_v^m, e_v^m)]\,\mathrm{d}\Phi(m).$$

What intergenerational constraints shall we place on this maximization? One is the ethical restraint that the government must fulfil the second-period consumption expectations of workers now retired from generation $v-1$. Hence x_{v-1} is a datum for the v-government.

The second restraint we impose is that the subsequence $\{k_{v+i}, x_{v+i} | i = 1,\ldots\}$ is predesignated. While the vector (k_v, x_v) is open to choice, all the subsequent vectors are fixed – perhaps optimally by some criterion of intergenerational justice – and the previous vectors are predetermined by history. The analytical rationale here is a familiar one: by the principles of Bellman and Euler, if the *whole* sequence including $i=0$ is to be a Pareto-efficient one, it is necessary that any particular vector $\{k_v, x_v\}$ be chosen efficiently when the other vectors are taken as given.

This latter restraint permits us to solve for the intragenerationally optimal l_{v+i}, beginning with $i=1$, as a function of the vectors $\{k_{v+i-1}, x_{v+i-1}\}$ and $\{k_{v+i}, x_{v+i}\}$, given the ith generation's social welfare function and all subsequent welfare functions. Hence the labor supply of generation $v+1$ is a function of the vector (k_v, x_v) chosen by the v-government, which function we denote by $\mathscr{L}(k_v, x_v)$, given the subsequent vectors $\{k_{v+i}, x_{v+i} | i = 1,2,\ldots\}$ and social welfare functions $\{\mathscr{W}^{v+i} | i = 1,2,\ldots\}$.

We can now formulate the welfare-maximization problem facing the government of generation v. For convenience of notation, we focus on the first generation, setting $v=1$. Following Atkinson and Stiglitz, we take x_1^m and e_1^m to be control variables, treat $u^*(m)$ as the state variable, and hence regard c_1^m as a function $z(\cdot)$ of $u^*(m)$, x_1^m, and e_1^m. In addition to the differential equation $\mathrm{d}u^*(m)/\mathrm{d}m$, the $v=1$ government faces two additional constraints. First, total second-period claims by its citizens cannot exceed some $\bar{x}_1$ and, secondly, the capital deposited cannot be less than some $\bar{k}_1$, i.e.

$$\bar{x}_1 - \int_0^M x_1^m \mathrm{d}\Phi(m) \geqq 0, \tag{10}$$

$$-\bar{k}_1 + F(k_0, l_1) - \int_0^M c_1^m \mathrm{d}\Phi(m) - x_0 \geqq 0. \tag{11}$$

(Of course, we are free to require that $(\bar{k}_1, \bar{x}_1)$ be chosen to yield Pareto-efficiency between generations 1 and 2.) The corresponding Hamiltonian function to be maximized is

$$\begin{aligned}\mathscr{H}^1(m) = {} & u^*(m)\mathrm{d}\Phi(m) \\ & + \left[\lambda^1(\bar{x}_1 - x_1^m) + \gamma^1\left(-\bar{k}_1 + F\left(k_0, \int_0^M l_1^m \mathrm{d}\Phi\right) - x_0 \right.\right. \\ & \left.\left. - c_1^m\right)\right]\mathrm{d}\Phi(m) \\ & + \mu^1(m)\left[\frac{-e_1^m u_e^*}{m}\right],\end{aligned} \tag{12}$$

where λ^1 and γ^1 are the Lagrange multipliers associated with the two isoperimetric constraints on the control variables in (10) and (11), respectively, and are therefore independent of m, i.e. $\mathrm{d}\lambda^1/\mathrm{d}m = \mathrm{d}\gamma^1/\mathrm{d}m = 0$; $\mu^1(m)$ is the co-state variable associated with the constraint (7).

We can give an interpretation of the multipliers λ^1 and γ^1 that will prove useful in the derivation of the optimal tax formulas. Using a well-known envelope theorem in control theory[2] we note that $\partial \mathscr{W}^{*1}/\partial \bar{x}_1 = -\lambda^1$ and $\partial \mathscr{W}^{*1}/\partial \bar{k}_1 = \gamma^1$ where $\mathscr{W}^{*1}(k_1, x_1)$ is the maximum social welfare that generation 1 can attain given its target vector $(\bar{k}_1, \bar{x}_1)$. If both of our constraints are binding at this maximum, then $\lambda^1 < 0$ and $\gamma^1 < 0$. Hence,

$$\left(\frac{\mathrm{d}k_1}{\mathrm{d}x_1}\right)_{\mathscr{W}^{*1}} \equiv -\frac{\partial \mathscr{W}^{*1}/\partial \bar{x}_1}{\partial \mathscr{W}^{*2}/\partial \bar{k}_1} = \frac{\lambda^1}{\gamma^1} > 0 \tag{13}$$

along any predetermined isowelfare contour. Two such contours for generation 1 are depicted in fig. 1.

One could also find from eq. (12) the marginal rate of substitution for generation 1 between k_0 and x_0, making another application of the envelope argument. By the same analysis we find that *for generation 2* the marginal

[2]Bryson and Ho (1969, pp. 90–91).

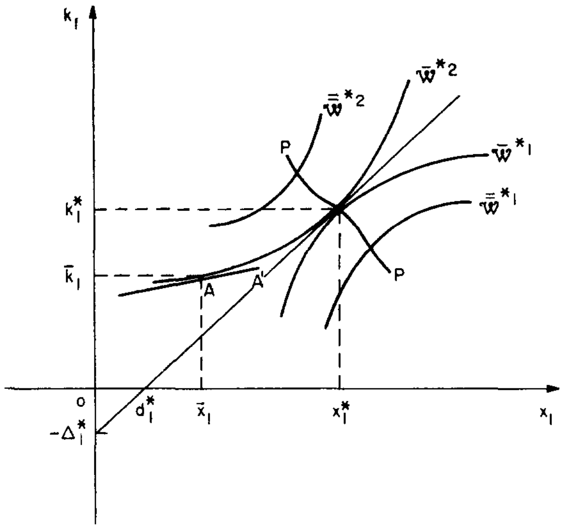

Fig. 1

rate of substitution between k_1 and x_1 is the reciprocal of one-plus-the-before-tax-rate-of-interest:

$$\left(\frac{dk_1}{dx_1}\right)_{\mathscr{W}^{*2}} = \frac{1}{F_k(k_1, \mathscr{L}(k_1, x_1))}. \tag{14}$$

This result, first obtained by Phelps and Riley (1978), is of wide generality, not depending on the form of the u^2 and $\mathscr{W}^2$ functions. Fig. 1 depicts two isowelfare contours for generation 2.

As shown in fig. 1, the isowelfare contours $\mathscr{W}^{*1}$ and $\mathscr{W}^{*2}$ are strictly quasiconcave in (k_1, x_1).[3] It follows from these respective quasiconcavities that there exists an interior locus, labelled PP in fig. 1, of intergenerationally Pareto-efficient vectors along which the two marginal rates of substitution are equal.[4]

[3]The requisite proof for $\mathscr{W}^{*1}$ is given in Phelps and Riley (1978, pp. 105–106). The proof for $\mathscr{W}^{*2}$ is as follows. Consider a vector $(\bar{k}_1, \bar{x}_1)$ as at point A in Fig. 1. The slope of $\bar{\mathscr{W}}^{*2}$ at that point is $1/F_k(\bar{k}_1, \mathscr{L}(\bar{k}_1, \bar{x}_1))$. Now assume that generation is faced with a different vector (k_1', x_1') lying on the line segment AA'. If generation 2, behaving passively, does not adjust its tax rates and consequently the values of c_2, x_2, and l_2, it will violate its own investment constraint analogous to eq. (11) of the text. The reason that this constraint would be violated at A' is that $F_{kk} < 0$. Since the constraint on k_2 must be met, c_2 must be reduced, or l_2 must be increased, or both. In either case aggregate social welfare must decline.

[4]To show that there must be a stretch of the PP locus in the positive quadrant note that $\partial \mathscr{W}^{*1}/\partial x_1 \to \infty$ as $x_1 \to 0$ whenever $\partial u/\partial x^m \to \infty$ as $x^m \to 0$. On the other hand F_k is bounded above zero, at least for some values of k_1, since l_2 is finite. That is, F_k does not tend to zero as $x_1 \to 0$.

The Hamiltonian formulation in (12) is well suited as it stands to the analysis of intragenerationally optimal taxation in an intergeneration setting in those instances when no particular constraints are placed on the target vector $(\bar{k}_1, \bar{x}_1)$. When we speak of our concept of optimal taxation we mean the above formulation in its full generality. Yet there is surely special interest in those cases where the target vector meets the condition of Pareto-efficiency between the generations. For that special case a somewhat different development of the optimization problem, as follows, is a little more expeditious.

We noted earlier that at any interior Pareto-efficient vector (k_1^*, x_1^*) the marginal rate of substitution between k_1 and x_1 must be the same for the two adjacent generations. Therefore, we have $\lambda^1 = \gamma^1 / F_k(k_1^*, \mathscr{L}(k_1^*, x_1^*))$ and consequently the Hamiltonian in (12) simplifies to

$$\begin{aligned}\mathscr{H}^1(m) = {} & u^*(m)\mathrm{d}\Phi(m) \\ & + \gamma^1 \Bigg\{ [F_k(k_1^*, \mathscr{L}(\cdot))^{-1} x_1^* - k_1^*] - \Bigg[F_k(k_1^*, \mathscr{L}(\cdot))^{-1} x_1^m \\ & - F\left(k_0, \int_0^M l_1^m \mathrm{d}\Phi(m)\right) + c_1^m + x_0 \Bigg] \Bigg\} \mathrm{d}\Phi(m) \\ & + \mu^1(m) \left[\frac{-e_1^m u_e^*}{m} \right].\end{aligned}$$

If we then define r_1^* and Δ_1^* by

$$1 + r_1^* = F_k(k_1^*, \mathscr{L}(k_1^*, x_1^*))$$

and

$$(1 + r_1^*)^{-1} x_1^* - k_1^* = \Delta_1^*,$$

we obtain a new Hamiltonian formulation of the maximization problem:

$$\begin{aligned}\mathscr{H}^1(m) = {} & u^*(m)\mathrm{d}\Phi(m) \\ & + \gamma^1 \Bigg[\Delta_1^* - (1 + r_1^*)^{-1} x_1^m + F\left(k_0, \int_0^M l_1^m \mathrm{d}\Phi\right) - c_1^m \\ & - x_0 \Bigg] \mathrm{d}\Phi(m) \\ & + \mu^1(m) \left[\frac{-e_1^m u_e^*}{m} \right]. \qquad (15)\end{aligned}$$

Thus, the social welfare of generation 1 is to be maximized subject to the deadweight debt constraint *and* shadow rate of interest, both parameters being associated with a preselected Pareto-efficient vector, as well as the differential-equation condition expressed in (7). It cannot be stressed enough that the former constraint specifies *two* parameters, Δ_1^* and r_1^*, and thus $d_1^* = \Delta_1^* \cdot (1+r^*)$, not just the budgetary deficit, and both parameters are derived from a prior choice of some target on the Pareto locus.

We remark that, as the formulation of the optimal tax problem in (15) makes clear, any point on the Pareto locus can be supported by a proper combination of public debt and the gross rate of return on saving, $F_k(\cdot)$. To attain the point (k_1^*, x_1^*) in fig. 1, for example, the government offers the gross rate of return of $F_k(k_1^*, \mathscr{L}(k_1^*, x_1^*)) \equiv F_k^*$ at which rate of return households save $\sigma^1(\cdot, F_k^*)$. Firms will demand that amount of capital which has a marginal product, when calculated at the optimal $l_2^* = \mathscr{L}(k_1^*, x_1^*)$, equal to the same rate of return to saving which was fortuitously chosen to be F_k^*. To accommodate private saving the government will have to issue the amount d_1^* of bonds where $d_1^*/F_k^* = \sigma_1(\cdot, F_k^*) - k_1^*$. The market value of those bonds, Δ^*, can be read-off from fig. 1 by extending the tangent to $\mathscr{W}^{*2}(k_1^*, x_1^*)$ to its intersection with the vertical axis.[5]

In principle, given certain transversality conditions on $\mu(\cdot)$ and the differential equation $\mathrm{d}u^*(m)/\mathrm{d}m$, the whole optimal tax schedule $T_s(\cdot)$ should be deducible from the first-order condition for a maximum of $\mathscr{H}^1(m)$ in (15):

$$\frac{\partial \mathscr{H}^1}{\partial x_1^m} = \gamma^1 \left[-(1+r_1^*)^{-1} - \left(\frac{\partial c_1^m}{\partial x_1}\right)_{u^*(m)} \right] \mathrm{d}\Phi(m) + \mu^1(m) \left\{ \frac{-e^m}{m} \left[u_{ec}^m \left(\frac{\partial c_1^m}{\partial x_1^m}\right)_{u^*(m)} \right] + u_{ex} \right\} = 0. \tag{16}$$

In practice, however, little can be said about the shape of this schedule other than at its endpoints or when the term in the curly brace is zero. We take up those implications in section 5.

4. Optimal taxation: 'maximin'

Since the 'maximin' social welfare function can usually be represented by the limiting Bergson–Samuelson social welfare function, as the degree of concavity of the functional G is increased without bound – for example,

[5]See Phelps and Riley (1978, fig. 4). Actually the argument does not require that the point being supported lie on the Pareto locus.

Atkinson (1973) – we should expect the above analytics to carry over to 'maximin' without essential differences. We show here that this is indeed the case, and we point out some special features of maximin-optimal taxation.

The minimum life-time utility scored by persons from generation 1, which is to be maximized, must equal $u^*(0)$ since no person of generation 1 can have a smaller utility than that; of course, $u^*(m)$ may equal $u^*(0)$ for positive m so small that such persons choose not to work. Thus, our 'maximin' problem is to maximize $u^*(0)$ subject to the x_1 and k_1 constraints in (10) and (11), respectively, and to the differential equation constraint in (7). This problem is equivalent to the problem of maximizing k_1 subject to (10), (7), and the constraint that $u^*(0) \geqq \bar{u}(0)$, where $\bar{u}(0)$ is the maximum minimum utility in the former problem. The Hamiltonian corresponding to the latter formulation is

$$\mathscr{H}^1(m) = \left[-x_0 + F\left(k_0, \int_0^M l_1^m \mathrm{d}\Phi\right) - c_1^m \right] \mathrm{d}\Phi(m) + \pi^1[\bar{x}_1 - x_1^m]\mathrm{d}\Phi(m) + \mu^1(m)\left[\frac{-e_1^m u_e^*}{m}\right] + \alpha^1[u^*(0) - \bar{u}(0)], \tag{17}$$

where π^1 is the Lagrange multiplier associated with the constraint in (10), $\mu^1(m)$ is the co-state variable associated with (7), and α^1 is the Lagrange multiplier associated with the constraint that $u^*(0) \geqq \bar{u}(0)$; both π^1 and α^1 are independent of m.

It is clear at once that π^1 measures the marginal rate of substitution for generation 1 between k_1 and x_1 since this multiplier measures the increase of maximum k_1 per unit of increase in the allowable x_1 at fixed $u^*(0)$, say $\partial k_1^*/\partial x_1$ by way of notation. Furthermore, the marginal rate of substitution for generation 2 between k_1 and x_1 is $F_k(k_1, l_2)^{-1}$ if that generation maximizes its Bergson–Samuelson welfare, as was shown in section 3, or if it maximizes its own minimum utility. To show the latter one only has to note that the maximand of generation 2, on the present formulation, is

$$-x_1 + F(k_1, l_2) - c_2 \equiv k_2^*,$$

subject to constraints in none of which do x_1 and k_1 appear; by the familiar envelope theorem, therefore,

$$\left(\frac{\mathrm{d}k_1}{\mathrm{d}x_1}\right)_{k_2^*} = \frac{1}{F_k(k_1, l_2)}.$$

Now consider any interior vector (k_1^*, x_1) that is Pareto efficient between generations 1 and 2, and denote such a Pareto optimum by (k_1^*, x_1^*). At any such point the above two marginal rates of substitution are equal, and consequently $\pi^1 = F_k(k_1^*, \mathscr{L}(k_1^*, x_1^*))^{-1}$. Hence, if x_1 has been set at some value x_1^* that will secure Pareto efficiency when k_1 is maximized subject to the three constraints, the corresponding Hamiltonian reduces to

$$\mathscr{H}^1(m) = \left[-x_0 + F\left(k_0, \int_0^M l_1^m \, \mathrm{d}\Phi\right) - c_1^m - F_k(k_1^*, \mathscr{L}(k_1^*, x_1^*))^{-1}(x_1^m - x_1^*) \right] \mathrm{d}\Phi(m) + \mu^1(m)[-e_1^m u_e^* m^{-1}] + \alpha^1[u^*(0) - \bar{u}(0)]. \tag{18}$$

As before, we are to choose l_1^m and x_1^m to maximize this Hamiltonian, subject to the relation in (8) determining c_1^m as a function of the control variables, l_1^m and x_1^m, and the state variable $u^*(m)$. Thus, the optimal (l_1^m, x_1^m) for each m yield a constrained maximum of the quantity

$$F\left(k_0, \int l_1^m \mathrm{d}\Phi\right) - c_1^m - F_k(k_1^*, \mathscr{L}(\cdot))^{-1} x_1^m.$$

This Pareto-efficient maximization has an interesting interpretation. At each m, l_1^m and x_1^m are doing their bit to maximize $-x_0 + F(k_0, l_1) - c_1 - F_k(k_1^*, \mathscr{L}(k_1^*, x_1^*))^{-1} x_1$, which is just $k_1 - F_k(k_1^*, \mathscr{L}(\cdot))^{-1} x_1$, subject to the constraints. But x_1 is $F_k(k_1^*, \mathscr{L}(\cdot))k_1 + d_1$; so the maximand is just $-d_1/F_k(k_1^*, \mathscr{L}(\cdot))$, the negative of the market value of the government debt sold to generation 1 savers. Thus, the paths $(l_1(m), x_1(m))$ must maximize the algebraic budgetary surplus of the government.

More formally, we may examine the integral of the maximand, with the constants omitted, and see from the household budget equations, the assumption of competitive factor rewards, and Euler's theorem, that this integral satisfies

$$\begin{aligned}
&\int_0^M \left[F\left(k_0, \int_0^M l_1^m \mathrm{d}\Phi\right) - c_1^m - F_k(k_1^*, \mathscr{L}(\cdot))^{-1} x_1^m \right] \mathrm{d}\Phi \\
&= F_k(k_0, l_1)k_0 + F_l(k_0, l_1)l_1 \\
&\quad - \int_0^M \left[w l_1^m + g_1 + \frac{h_1}{F_k(\cdot)} - T_l(\cdot) - T_s(\cdot) - \frac{x_1^m}{F_k(\cdot)} \right] \mathrm{d}\Phi - \frac{x_1}{F_k(\cdot)} \\
&= F_k(k_0, l_1)k_0 + \int_0^M [T_l(\cdot) + T_s(\cdot)] \mathrm{d}\Phi - \left[g_1 + \frac{h_1}{F_k(\cdot)} \right].
\end{aligned} \tag{19}$$

Now let $\bar{r}_0$ denote the net rate of return on k_0 that the old must receive if their total receipts, $d_0+(1+\bar{r}_0)k_0$, are to add up to their old-age consumption claim, x_0. Then the algebraic excess earnings, $(Y_k(k_0,l_1)-\bar{r}_0)k_0$, constitute surtax revenue additional to the revenues from $T_l(\cdot)$ and $T_s(\cdot)$. Hence the maximization of our integral in (19) implies the maximization of *total* tax revenues *net* of the present value of g_1 and h_1 required to bring $u^*(0)$ up to the mandated $\bar{u}(0)$.

It is now a short, but subtle, step to the conclusion that total tax revenues are maximized – gross of g_1 and h_1. Clearly, if g and h were arbitrarily set, tax revenues could not generally be maximized without causing $u^*(0)$ to be excessive or else deficient in relation to $\bar{u}(0)$. However, the availability of the demogrant, $g+h/F_k(\cdot)$, provides a degree of freedom with which to compensate those persons having the least utility for any taxation of their savings, i.e. their second-period consumption, and thus to maintain their utility at $\bar{u}(0)$.[6] By contrast it would be empty to say that taxation to maximize Bergson–Samuelson welfare causes the revenue collected to be maximized; for while it is true that the deadweight burden on the next generation is minimized, and thus also the algebraic budgetary deficit, when the Bergson–Samuelson social welfare level is taken as given, total taxes collected could be simultaneously at a maximum only if g and h were *graduated* according to each person's m – a graduation which, if the required information were available, would permit lump-sum taxation.

There is a technical point about the revenue maximum worth noting. While the aforementioned surtax revenues figure in the total revenue collected, it is nevertheless true that at the optimum the quantity

$$\int_0^M [T_l(w_1^* l_1^m)+T_s(x_1^m)]\,\mathrm{d}\Phi$$

is being maximized where w_1^* is a *parameter* equal to $F_l(k_0,l_1)$ when evaluated at the optimal l_1. The fact that larger l_1^m over any interval of m between 0 and M would reduce $F_l(k_0,l_1)$, and thus lower both the before-tax wage bill and wage-tax revenue on that account, is just offset by an equal rise of the surtax revenue since, by what amounts to the factor-price-frontier relation,

$$\frac{\mathrm{d}}{\mathrm{d}l_1}[k_0F_k(k_0,l_1)-(1+\bar{r}_0)k_0+l_1F_l(k_0,l_1)]=F_l(k_0,l_1).$$

[6] At least in the present formulation, therefore, it is unnecessary to adopt the device employed by Ordover and Phelps (1975) of supposing that g and h are so distributed as to obviate private saving by the poorest persons to achieve their optimal consumption plan. See also Ogura (1977).

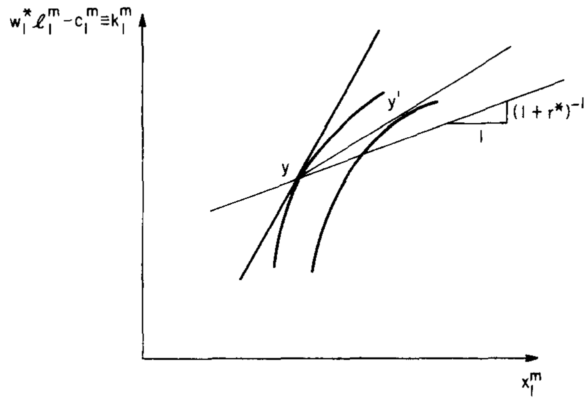

Fig. 2

in fig. 2. In this diagram each indifference curve shows the amount of x_1^m needed to compensate them for a given amount of private plus public saving extracted from them for a given wage tax function and demogrant. At y the corresponding indifference curve is steeper than the straight line with slope $(1+r_1^*)^{-1}$ because their 'budget line' has locally a slope $(1+r_1^*)^{-1}+T_s'(x_1^m)$. An infinitesimal reduction of the marginal tax rate on this x_1^m and all higher x_1^m would bend down the budget line beyond y and cause some point y' to be chosen instead of y. At y' the top savers would necessarily be saving more, through lesser c_1^m or greater l_1^m, and consuming larger x_1^m. They and possibly some near-top savers would be better off, their opportunities having been expanded; no one from generation 1 need be left worse off, since the government could introduce *ad valorem* surtaxes and subsidies to maintain the market wage rate at w^*, the rate of return at r^*, and the retired persons' receipts at x_0 despite the rise of k_1^m and perhaps l_1^m; and generation 2 would enjoy a rise of social welfare because that generation, whether of the 'maximin' or Bergson–Samuelson type, would be willing to accept the burden of additional x_1 at the rate $(1+r_1^*)$ per unit of increase in k_1 in the neighbourhood of (k_1^*, x_1^*) while the move from y to y' offers better terms than that. Therefore y could not have corresponded to a Pareto optimum – a contradiction. Hence $T_s'(\max_m x_1^m)$ cannot optimally be positive. By a similar argument it must not be negative either.

Our proposition regarding the marginal tax rate at the top is simply proved by appeal to the mathematics of the Hamiltonian formulation. If the target vector is a Pareto-efficient one, (k_1^*, x_1^*), the Hamiltonian becomes (15) in the Bergson–Samuelson case and (18) in the maximin case. In either case, it is a known property of the solution to the maximization problem that the co-state variable, $\mu^1(m)$, reaches zero at $m=M<\infty$.[7] At lower m, $\mu^1(m)>0$ to

[7]Gelfand and Fomin (1963, sec. 1.6).

The foregoing analysis has no immediately apparent implications for the maximin-optimal taxation of wages and second-period consumption by persons with utilities at or near the minimum level, $u^*(0)$. However, it seems to us improbable that the marginal tax rate on the first few dollars of wage earnings would optimally be negative, i.e. $T_l'(0)<0$. Such a property of the tax schedule would raise the eventual height of the marginal tax rate needed to extract any given revenue from middle-wage and higher-wage workers, and the heightened disincentives for lower-middle-wage workers would presumably cost more revenue than what might be gained from low-wage workers. It seems more probable that $T_l'(0)>0$. Then, with marginal tax rates unchanged at some wl_1^m and beyond, all persons earning that much more will have to pay a greater tax on their inframarginal earnings than if marginal tax rates started at zero.

We ought to add in closing that our tax-maximization proposition does not extend to a rather different conception of social welfare – a notion of economic justice more in the spirit of Rawls. One might restrict the demogrant, to be set at some humanitarian level, to those persons who do not work, whether unable or unwilling, and seek to maximize the after-tax wage rate – more naturally, the after-subsidy wage rate – of those working persons with the lowest such wage rate. Then revenue maximization would be the objective only at the high-wage-income brackets.

5. Two implications for taxation of interest and wealth

Here we discuss two consequences of our concept of optimal taxation for the appropriate tax or subsidy on second-period consumption. The first implication pertains to the optimal marginal tax rate at the highest attained level of second-period consumption, in effect the marginal tax rate that applies to the top interest-income bracket. Phelps (1973) showed the maximin optimality, and Sadka (1976) the Bergson–Samuelson optimality, of driving the marginal rate of taxation of *wage* income to *zero* at the top level of wage earnings attained – provided, as is natural, that m has some upper bound $M<\infty$ so that there exists a highest earnings level for each wage-tax schedule. Other derivations can be found in Cooter (1978), Ogura (1977), and Seade (1977) which studies also the marginal tax rate at the bottom wage-income bracket. The question we pose is whether the analogous result extends to the optimal taxation of second-period consumption.

There is an easy heuristic argument that if the vector (k_1, x_1) is required to be *Pareto efficient*, then for both Bergson–Samuelson and maximin optimality the marginal tax rate $T_s'(x_1^m)$ must indeed be zero at the highest x_1^m. For suppose, contrariwise, that $T_s'(\max_m x_1^m)$ were positive, say, at the fiscal optimum. Then persons with the highest x_1^m would be driven to a point like y

signal that, by reason of the differential equation in (7), the faster utility is made to rise around m the greater will be the utility in store for persons with higher m and hence the smaller will be the remaining room to meet the target vector, (k_1^*, x_1^*); but at $m=M$ there are no implications for utility at still higher m to worry about, so $\mu^1(M)=0$.

Hence, at $m=M$ the first-order condition $\partial\mathscr{H}^1/\partial x_1^m=0$ for a maximum of the Bergson–Samuelson Hamiltonian in (16) reduces to

$$\frac{\partial\mathscr{H}^1(M)}{\partial x_1^M}=\gamma^1\left[-(1+r_1^*)^{-1}-\left(\frac{\partial c_1^M}{\partial x_1^M}\right)_{u^*(M)}\right]\mathrm{d}\Phi(M)=0.$$

Since $\gamma^1<0$, we have $(\partial c_1^M/\partial x_1^M)_{u^*(M)}=-(1+r_1^*)^{-1}$. And since $1+r_1^*=F_k(k_1^*,\mathscr{L}(k_1^*,x_1^*))$ when the target vector is Pareto efficient, it follows that $T_s'(x_1^M)=0$. Of course, a similar argument involving $\partial\mathscr{H}^1/\partial l_1^M$ shows that since $\mu^1(M)=0$, $T_l'(wl_1^M)=0$ also. Readers may verify that maximization of the maximin Hamiltonian in (18) at $m=M$ has the same implications.

It ought to be acknowledged, as has been conceded before, that the deterministic framework here makes no provision for random influences upon interest-type income some insurance against which might be Bergson–Samuelson or even maximin optimal for the government to provide. Some extreme examples, based on an *ex ante* homogeneous population, where optimal 'regressivity' of marginal tax rates occurs only beyond the experienced range have been constructed by Varian (1978). However, it must be wondered whether the insurance function of everywhere-positive marginal tax rates is in fact very important to the optimal graduation of marginal tax rates. It may also be noted that, surprisingly, the above proof does not assume that it is the workers with the largest m who have the greatest x_1^m; but perhaps this relation is implied.

Now consider the optimal taxation of second-period consumption, hence of interest income, at *any* $m>0$. If again we impose the condition that the target be some *Pareto efficient* vector, (k_1^*, x_1^*), a first-order condition for the maximum of the corresponding Bergson–Samuelson Hamiltonian in (15) is

$$\frac{\partial\mathscr{H}^1}{\partial x_1^m}=\gamma^1\left[-(1+r_1^*)^{-1}-\left(\frac{\partial c_1^m}{\partial x_1^m}\right)_{u^*(m)}\right]\mathrm{d}\Phi(m)$$
$$+\mu^1(m)\left\{\frac{-e^m}{m}\left[u_{ec}^m\left(\frac{\partial c_1^m}{\partial x_1^m}\right)_{u^*(m)}\right]+u_{ex}\right\}=0.$$

The terms in the curly braces also turn up in the conceptually distinct maximization problem of Atkinson and Stiglitz. As is shown there, these terms sum to zero for every m if the utility function is of the weakly separable form $u[\psi(c,x),e]$ so that each worker's marginal rate of substitution between the two consumer goods is independent of the amount of first-period effort or leisure. On this further *separability condition*, then, our first-order condition implies $(1+r_1^*)^{-1} = -(\partial c_1^m/\partial x_1^m)_{u^*(m)}$, whence $T_s'(x_1^m)=0$ for every m, not just for $m=M$. Interest income must then be tax-exempt.

Does this result extend, under the conditions of Pareto-efficiency and separability, to the maximin case? One might guess not on the thought that a tax structure designed to maximize total tax revenue would shun any form of tax exemption as nature abhors a vacuum. Yet, it is a straightforward calculation to show that the first-order condition, $\partial\mathscr{H}^1/\partial x_1^m=0$, for a maximum of the maximin Hamiltonian in (18) yields essentially the same equation as above – with γ^1 replaced by 1.

An intuitive explanation of this result is that if the taxation of labor income has been preset at its first-best optimum, there is no more to be gained in revenue from a surtax on x_1^m than from a surtax on c_1^m since, under the separability condition, leisure is no more complementary to (or less substitutable for) c_1^m than to x_1^m; apparently either surtax or both together would further weaken the incentive to work and thus cause wage-tax revenues to fall by at least the amount of the surtax revenue collected. Moreover, either surtax would cause the target (k_1^*, x_1^*) to be achievable, if at all, only through a lower $u^*(0)$ since households would not be trading x_1^m for c_1^m at the r^* associated with $F_k(k_1^*, \mathscr{L}(k_1^*, x_1^*))$ and consequently second-best tax and transfer measures would have to be taken.

It must be clear that the above tax-exemption proposition fails to hold once the Pareto-efficiency condition is dropped. Indeed, the determination of the optimal tax or subsidy to second-period consumption when the target vector is Pareto inefficient is quite transparent under the separability condition. We note that an inefficient target, $(\bar{k}_1, \bar{x}_1)$, makes the marginal rates of substitution between k_1 and x_1 unequal between generations. In the Bergson–Samuelson Hamiltonian of eq. (12), therefore, the ratio of the Lagrange multipliers must satisfy $\lambda^1=\gamma^1[1/F_k(\bar{k}_1,\ \mathscr{L}(\bar{k}_1,\bar{x}_1))+\delta]$, where δ, which is a function of $(\bar{k}_1,\bar{x}_1)$, is either positive or negative. Hence the Hamiltonian in (12) can now be written as

$$\begin{aligned}\mathscr{H}^1(m) &= u^*(m)\mathrm{d}\Phi(m)\\ &\quad+\gamma^1\left[\left(\frac{1}{F_k}+\delta\right)\bar{x}_1-\bar{k}_1+F\left(k_0,\ \int_0^M l^m\mathrm{d}\Phi(m)\right)-x_0\right]\mathrm{d}\Phi(m)\\ &\quad-\gamma^1\left[\left(\frac{1}{F_k}+\delta\right)x_1^m+c_1^m\right]\mathrm{d}\Phi(m)+\mu^1(m)\left[\frac{-e_1^m u_e^*}{m}\right].\end{aligned}$$

On the separability assumption, then, the necessary first-order condition with respect to x_1^m is simply

$$\frac{\partial \mathscr{H}^1}{\partial x_1^m} = -\gamma^1\left[\left(\frac{1}{F_k}+\delta\right)+\left(\frac{\partial c_1^m}{\partial x_1^m}\right)_{u^*(m)}\right]=0 \tag{20}$$

whence, by the consumer equations (5b) and (5c),

$$-\gamma_1\left[\left(\frac{1}{F_k}+\delta\right)-\left(\frac{1}{F_k}+T_s'(x_1^m)\right)\right]=0. \tag{21}$$

From this we can deduce that $T_s'(\cdot)$ is constant for all values of x_1^m. The optimal indirect tax schedule must be linear in x – *for all inefficient* $(\bar{k}_1, \bar{x}_1)$ *vectors*. In fact we can be more exact than that: if the target vector $(\bar{k}_1, \bar{x}_1)$ is to the left of the Pareto locus, a positive tax on saving is called for; if the target vector is to the right of that locus, a subsidy is appropriate to the target.[8] The reader can confirm that these results carry over to the maximin case.

6. Piecemeal reform and 'second best'

The first-best optimality of exempting interest from taxation, under the separability condition, is an arresting result. The same proposition reached before us by Atkinson and Stiglitz, whose argument we have corrected with the needed Pareto-efficiency condition, has already aroused wide interest. In effect it restores the doctrine of Corlett and Hague (1953) which states that a tax on present consumption, equivalently a subsidy to saving, would be optimal if leisure were more complementary to present consumption than to future consumption; and a tax on saving or interest, equivalently a subsidy to present consumption, would be optimal if the differential complementarity is the reverse. Although the theorems here do not strictly apply to any of our own earlier work on maximin taxation, because our tax structures were linear rather than graduated, we somehow missed (or repressed) an important observation: there is *no utility-theoretic presumption* that the taxation of interest income would gain more in interest-income levies collected than it would lose in revenue from wage income – just as there is no presumption that a surtax on present consumption would raise revenue on balance – in those cases studied where wage-income taxation, as well as interest-income taxation, is being set to maximize total revenue collected.

That said, it must still be emphasized that the sort of piecemeal taxation reform which the above conclusions may encourage is not without risks.

[8]See also Atkinson and Sandmo (1977) for the corresponding point in relation to steady-state paths.

Even though we take separability for granted, it is not a corollary of our findings that the economy would be led (as if by an invisible hand) to the Pareto locus were interest-income taxation to be abolished; that efficient outcome would be assured only if wage-income taxation were already, or was going simultaneously to be, optimized as well. And even if exempting interest from taxation would move the economy nearer to that locus, other steps might have to be taken simultaneously to prevent a fall of social welfare for one generation or the other.

The risk that lighter taxation of interest-type income carries for the *next* generation is straightforward – assuming, as heretofore, that the next generation will tax optimally and redeem in full the public debt. Each increase of the after-tax rate of return to saving will burden the next generation with larger x_1 – necessarily if second-period consumption is a noninferior good – while the consequent k_1 may fail to increase (if it increases at all) by a compensating amount. It is immediately evident from the geometry of fig. 1 that k_1 must rise per unit of increase in x_1 at a rate at least as great as $F_k(k_1, l_2)^{-1}$ in order that the maximized welfare of generation 2 not fall as x_1 is increased. In algebraic terms, the next generation will be injured if and only if the constraint under which it operates,

$$F(k_1, l_2) - x_1 - c_2 = k_2, \tag{22}$$

is contracted on balance by the change in (k_1, x_1), i.e. if and only if $\mathrm{d}k_1/\mathrm{d}x_1 \leqq F_k(\cdot)^{-1}$.

This result has an interesting (though deceptive) interpretation. In the linear tax case, with the after-tax rate of return to saving denoted by ρ and the after-tax wage of a standard ($m=1$) unit of labor denoted by ω.

$$k_1 = F(k_0, l_1) - \omega l_1 - g + \sigma - x_0, \quad .$$
$$x_1 = h + (1+\rho)\sigma;$$

hence the derivative with respect to ρ of the left-hand side of the constraint equation in (22) is

$$F_k(k_1, l_2) \cdot \left\{ [F_l(k_0, l_1) - \omega] \frac{\partial l_1}{\partial \rho} + \frac{\partial \sigma_1}{\partial \rho} \right\} - \left[\sigma_1 + \frac{\partial \sigma_1}{\partial \rho}(1+\rho) \right].$$

Evidently that derivative is non-negative, so that generation 2 is not being harmed by rising ρ, if and only if

$$(F_l - \omega)\frac{\partial l_1}{\partial \rho} + \frac{1}{F_k}\left\{ [F_k - (1+\rho)] \frac{\partial \sigma_1}{\partial \rho} - \sigma_1 \right\} \geqq 0. \tag{23}$$

This 'marginal revenue' condition means that total tax revenue, discounting future revenues by $F_k(k_1, l_2)^{-1}$, is locally nondecreasing in ρ.

The proper deduction to be drawn, apropos discrete changes in ρ, is that the next generation will surely be harmed by a given increase of ρ if there results a rise of the par value of the *notional* debt as calculated at the *original* or no-reform rate of return, $F_k(k_1^0, \mathscr{L}(k_1^0, x_1^0))$, i.e. a rise of $x_1 - k_1 F_k^0(\cdot)$; as fig. 3 clearly demonstrates, every (k_1, x_1) outcome that is as

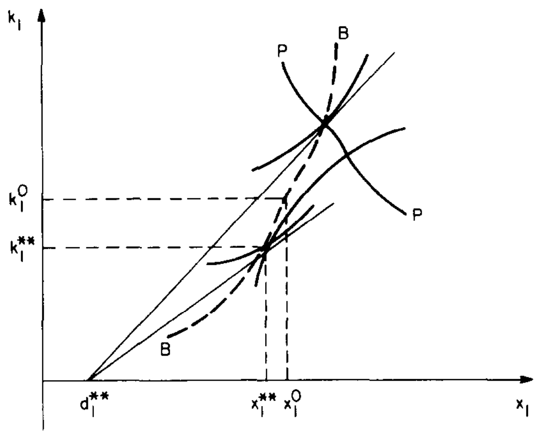

Fig. 3

good or better for the next generation than the no-reform outcome, (k_1^0, x_1^0), makes $x_1 - k_1 F_k^0(\cdot)$ smaller than the original $x_1^0 - k_1^0 F_k^0(\cdot)$. Yet it is also clear from the geometry of fig. 3 that the *actual* debt associated with the *new* and lower rate of return might rise without harm to the next generation; for example, a rise of x_1 for which the increase of k_1 just compensates, causes $x_1 - k_1 F_k(\cdot) > x_1^0 - k_1^0 F_k^0(\cdot)$ owing to the curvature of generation 2's welfare contours. The producer's surplus of the next generation is

$$F(k_1, l_2) - d_1 - k_1 \cdot F_k(k_1, l_2),$$

the derivative of which

$$-\frac{\partial d_1}{\partial \rho} - F_{kk}(k_1, l_2) \cdot k_1 \frac{\partial k_1}{\partial \rho},$$

can be positive whether or not $\partial d_1 / \partial \rho > 0$; for even if the increase of k_1 is due only to a ρ-induced rise of *private* saving, tax revenue and thus public saving being down by assumption, the next generation gains on that account by paying out a diminished marginal product on the original quantity of k_1.

We have shown that a steadily rising ρ, ultimately meeting $F_k(\cdot)-1$, will continue not to harm the next generation as long as the 'marginal revenue' condition in (23) is satisfied. The expression in the curly braces there may be written as

$$\left[\left(\frac{1+r}{1+\rho}-1\right)\frac{\partial\sigma_1}{\partial(1+\rho)}\cdot\frac{1+\rho}{\sigma_1}-1\right]\sigma_1\,;$$

its algebraic sign gives the direction of interest-income tax revenue as ρ is rising. To calculate that sign one wants something like the 25-year rates of return to saving and the corresponding elasticity. Boskin (1978) has estimated the *annual* after-tax interest-rate elasticity to be 0.4. Using his average *annual* after-tax rate of return of 0.03, we then calculate[9] the 25-year $(1+\rho)$-elasticity to be 0.76. Furthermore, taking his average *annual* before-tax rate of return of 0.07, we find the 25-year interest-factor ratio, *our* $(1+r)/(1+\rho)$, to be 2.67. Then the term in the brackets above is positive, equal to 0.27. But high confidence in the 0.4 estimate is not widespread; a mere one-quarter reduction of Boskin's elasticity and thus ours as well just about reduces the bracketed expression to zero. And we have no estimates of the wage-tax term.

It is especially uncertain whether the wholesale elimination of interest taxation would raise tax revenue figured at the original rate of return. If ω is so large that σ is substantial then, as ρ also becomes large, it may well be that $\partial l/\partial\rho$ and $\partial\sigma/\partial\rho$ will become small or even negative because the income-effects of further increases of ρ become large enough to outweigh the substitution effects. And if ω is close to $F_l(k_0,l_1)$ there can be little help from $\partial l_1/\partial\rho>0$ anyway. As ρ is increased all the way to $F_k(k_1,l_2)-1$, therefore, x_1 may eventually increase so steeply against k_1 as to jeopardize all the gain (if any) to the next generation from the first increments of ρ.

The risks to the *present* generation posed by lighter interest-income taxation raise more delicate issues. At first it might seem that the present generation will lose nothing from increasing its ρ since, as assumed above, it can protect its after-tax wage rates through lighter wage taxation if needed. It is true that the present generation need lose nothing as long as its private saving is large enough to accommodate both the desired public borrowing and the minimum capital that the next generation will need to meet old-age claims. However, the next generation may similarly avoid injury up to a point by denying the retired population some of their anticipated consumption claims – if it does not thus prompt the succeeding generation to do

[9]The 25-year $(1+\rho)$-elasticity is

$$\left(\frac{\partial\sigma_1}{\partial\rho}\frac{\rho}{\sigma_1}\right)\left(\frac{1+\rho}{\rho}\right)=(0.4)\left[\frac{(1.03)^{25}}{(1.03)^{25}-1}\right].$$

the same. Any risks to the present generation from raising ρ – even by the *fait accompli* of tax credits for saving made deductible from wage tax liabilities – therefore hinge on the presence of one or more self-imposed constraints it may place on its actions *vis-à-vis* the next generation in recognition of the latter's deterrent powers of retaliation. Then lighter interest-income taxation may induce heavier wage-income taxation and/or lower transfer payments than would otherwise have been possible in order still to meet the constraints, and the net result may be a reduction of the present generation's welfare.

One constraint that the present generation might adopt is the one-parameter restraint of budget balance:

$$x_1 - k_1 F_k(k_1, \mathscr{L}(k_1, x_1)) \leqq d_1^{**}, \; d_1^{**} = d_0.$$

While we shall see that this restraint is inappropriate for the attainment of any Pareto-efficient or 'first-best' optimum – for that one wants the appropriate two-parameter restraint, $x_1 - k_1(1+r_1^*) \leqq d_1^*$ – the balanced budget constraint has been of long-standing interest and it may be realistic.[10] If that is the only self-restraint exercised by the present generation, and if we put aside the next generation's willingness and ability to redeem the public debt, the government can then select any point of the opportunity locus, labelled *BB* in fig. 3, that is traced out by clockwise rotation of a straight line around the point ($x_1 = d_1^{**}, k_1 = 0$). We shall assume that *BB* makes x_1 a single-valued function of k_1 – at least there is a largest x_1 for every $k_1 > 0$; its slope relative to the k-axis,

$$\left(\frac{\mathrm{d}x_1}{\mathrm{d}k_1}\right)_{d_1^{**}} = \frac{F_k(\cdot) + k_1\left[F_{kk}(\cdot) + F_{kl}\dfrac{\partial l_2}{\partial k_1}\right]}{1 - k_1 F_{kl}\dfrac{\partial l_2}{\partial x_1}},$$

necessarily starts positive and approaches 1 minus the rate of depreciation as $k \to \infty$. Travelling onward along *BB* from its origin corresponds to improving potential welfare for generation 2. Hence *BB* slices the next generation's isowelfare contours from below. At each intersection *BB* is steeper than the local isowelfare contour.

If the present generation maximizes its own social welfare subject only to the balanced-budget constraint, besides the omnipresent Mirrleesian differen-

[10]In Phelps (1965) it is speculated that a generation might embrace the 'fiscal neutrality' of lifetime budget balance out of fear that unbalancing the budget would provoke the next generation to repudiate the increment of the debt. Some results on the second-best optimality of taxing interest income in this setting were recently derived by Krohn (1978). As noted earlier, Atkinson and Stiglitz also postulate budget balance.

tial equation in (7), it will choose a 'second-best' target (k_1^{**}, x_1^{**}) where – assuming an interior solution – *BB* is tangent to the best of its intersecting isowelfare contours, as illustrated in fig. 3. It follows immediately that at this second-best target the isowelfare contour of generation 1 is steeper than that of generation 2. Under the separability condition, as we showed, the slope of the former is precisely the optimal $F_k(\cdot)^{-1} + T_s'(x_m)^{**} \equiv (1+\rho^{**})^{-1}$, for all m, and the slope of the latter is $F_k(\cdot)^{-1}$. At the second-best optimum, therefore, $T_s'(x_m)^{**} > 0$, i.e. $1+\rho^{**} < F_k(\cdot)$. To prove this proposition more directly we would remark that, on the separability assumption, the first-order condition for an interior maximum of the second-best Hamiltonian, $\mathscr{H}^{**}(m)$, corresponding to the balanced-budget constraint is

$$0 = -1 + k_1 F_{kl}(k_1, \mathscr{L}(k_1, x_1)) \frac{\partial \mathscr{L}}{\partial x_1}$$
$$+ \left(-\frac{\partial c_1^m}{\partial x_1^m} \right)_{u^*(m)} \left\{ F_k(\cdot) + k_1 \left[F_{kk}(\cdot) + F_{kl}(\cdot) \frac{\partial \mathscr{L}}{\partial k_1} \right] \right\}$$
$$\equiv \frac{\partial \mathscr{H}^{**}(m)}{\partial x_1^m},$$

after dividing by the associated Lagrange multiplier. It follows that $(-\partial c_1^m / \partial x_1^m)_{u^*(m)}$ is independent of m and equal to the slope $(\mathrm{d}x_1/\mathrm{d}k_1)_{d_1^{**}}$ of *BB* at the second-best maximum.

It has just been shown that a generation bound only by budget would, for its self-interest, drive ρ below the net market rate of return to investment, moving inward along *BB* until the optimum degree of 'monopolistic' exploitation is reached. If we hypothesize that its adherence to traditional tax practice would be second-best optimal for the present generation, therefore, and if the present generation is in fact bent on preserving budget balance, then any tax reform on its part that abolished or merely lightened the taxation of interest income would necessarily reduce that generation's social welfare – given that the next generation will maximize its welfare independently of any precedents or innovations made by the present generation. However, the hypothesis of a second-best *status quo ante* is obviously overstrong.

A weaker hypothesis is that, owing to the traditional practice of taxing interest income, the no-reform outcome (k_1^0, x_1^0) would fall short of the intersection of *BB* with the Pareto locus – it would be a point on *BB* southwest of *PP* – though it would not lie farther down *BB* than the second-best (k_1^{**}, x_1^{**}). Even if the economy had been overtaxing interest income from the second-best standpoint, to take one contrary example, it might have been undertaxing wage income and thus not tending to fall short of its second-best optimum. Such a no-reform outcome, (k_1^0, x_1^0), is illustrated in fig. 3. At that (k_1^0, x_1^0), generation 1 cannot move outward along *BB* without reducing its own potential welfare; and it can never move inward without injuring generation 2. If we made the additional hypothesis that the present generation will constrain itself not to reform its taxation in ways that injure the next generation, whether out of a sense of fairness or (again) out of fear of retaliation, then the present generation will not wish to deviate from (k_1^0, x_1^0).

But if a point like (k_1^0, x_1^0) becomes the present generation's target vector, the corresponding taxation that is third-best optimal makes $\rho^{***} < F_k(k_1^0, \mathscr{L}(k_1^0, x_1^0)) - 1$. This follows from the fact that at (k_1^0, x_1^0) the isowelfare contour of generation 1 slices that of generation 2 in the same way as at (k_1^{**}, x_1^{**}) – because both target vectors lie on *BB* at points southeast of the Pareto locus. Therefore the outright abolition of interest-income taxes by the present generation would make its potential welfare smaller than it could be – and perhaps smaller than it would be if instead the generation maintained ρ at its initial level.

Of course it is not denied that there exists a subset of target vectors different from (k_1^0, x_1^0) each of which would permit a mutual welfare gain for the two generations over that available with (k_1^0, x_1^0). On the foregoing hypothesis about the no-reform (k_1^0, x_1^0), however, the set of mutually superior target vectors – including an interval of the Pareto locus – lies entirely beyond the *BB* locus. Hence, each of these targets would require an increase of the public debt for its support. Until this implication is understood and accepted in fiscal policy-making, then there is a danger that exempting interest income from tax would ultimately reduce the welfare of the present generation.

References

Atkinson, A.B., 1973, How progressive should income-tax be?, in: M. Parkin, ed., Essays on modern economics (Longman, London); reprinted in: E.S. Phelps, ed., Economic justice (Penguin, Harmondsworth).

Atkinson, A.B. and A. Sandmo, 1977, The welfare implications of personal income and consumption taxes, Discussion paper (University College, London).

Atkinson, A.B. and J.E. Stiglitz, 1976, The design of tax structure: Direct versus indirect taxation, Journal of Public Economics 6, 55–75.

Boskin, M.J., 1978, Taxation, saving, and the rate of interest, Journal of Political Economy 86, S3–S28.

Bryson, A.E. and Y.-C. Ho, 1969, Applied optimal control (Wiley, New York).

Cooter, R., 1978, Optimal tax schedules and rates: Mirrlees and Ramsey, American Economic Review 68, 756–768.

Corlett, W.J. and D.C. Hague, 1953, Complementarity and the excess burden of taxation, Review of Economic Studies 21, 21–30.

Diamond, P.A., 1965, National debt in a neoclassical growth model, American Economic Review 55, 1125–1150.

Feldstein, M.S., 1978, The welfare cost of capital income taxation, Journal of Political Economy 86, S29–S53.
Gelfand, J.M. and S.V. Fomin, 1963, Calculus of variations (Prentice-Hall, Englewood Cliffs).
Jha, R., 1978, Essays in economic theory, Unpublished doctoral dissertation, Columbia University.
Krohn, L.D., 1978, Intertemporal monopoly power: Model of aggregate growth with government intervention and ethical constraints, Unpublished doctoral dissertation, Columbia University.
Mirrlees, J.A., 1971, An exploration in the theory of optimal income taxation, Review of Economic Studies 38, 175–208.
Ogura, S., 1977, More on Rawlsian optimal income taxation, Quarterly Journal of Economics 91, 337–344.
Ordover, J.A., 1976, Distributive justice and optimal taxation of wages and interest in a growing economy, Journal of Public Economics 5, 139–160.
Ordover, J.A. and E.S. Phelps, 1975, Linear taxation of wealth and wages for intergenerational lifetime justice: Some steady-state cases, American Economic Review 65, 660–673.
Phelps, E.S., 1965, Fiscal neutrality toward economic growth (McGraw-Hill, New York).
Phelps, E.S., 1973, Taxation of wage income for economic justice, Quarterly Journal of Economics 87, 331–354; reprinted in: E.S. Phelps, ed., Economic justice (Penguin, Harmondsworth).
Phelps, E.S., 1976, Linear 'maximin' taxation of wage and property income on a 'maximin' growth path, in: B.A. Balassa and R.R. Nelson, eds., Economic progress, private values and public policy: Essays in honor of William Fellner (North-Holland, Amsterdam).
Phelps, E.S. and J.G. Riley, 1978, Rawlsian growth: Dynamic programming of capital and wealth for intergeneration 'maximin' justice, Review of Economic Studies 45, 103–120.
Sadka, E., 1976, On income distribution, incentive effects, and optimal income taxation, Review of Economic Studies 43, 261–267.
Seade, J., 1977, On the shape of optimal tax schedules, Journal of Public Economics 7, 203–235.
Varian, H.R., 1978, Redistributive taxation as social insurance, Center for Research on Economic and Social Theory Research Paper (University of Michigan).