Journal of Public Economics 17 (1982) 181–211. North-Holland Publishing Company

OPTIMUM TAXATION WITH ERRORS IN ADMINISTRATION

Nicholas STERN*

University of Warwick, Coventry CV4 7AL, UK

(with programming by Alan Carruth and David Deans)

Received March 1980, revised version received August 1981

The paper compares the welfare levels which can be achieved by two distinct tax regimes: lump-sum taxation, where one attempts to identify individuals and allocate transfers or subsidies on the basis of characteristics, and income taxation, where characteristics are not observed but personal incomes are measured and taxed. Where there are no errors in classifying individuals, lump-sum taxation is superior, but, where mistakes are made in the allocation of lump-sum grants or subsidies, income taxation may be more attractive. The level of errors where the regimes are equally desirable in terms of social welfare is computed in simple models following that of Feldstein (1973). Where there is strong aversion to inequality, then income taxation becomes preferable at quite small errors.

In analysing income taxation it is shown that in the Feldstein model with endogenous wages the marginal tax rate (in optimum income taxation) on the more skilled is negative and that on the less skilled is positive in contrast to the standard results [see Seade (1977)] with exogenous wages.

1. Introduction

The basic theorem of welfare economics tells us that, under standard assumptions, the first best can be achieved as a competitive equilibrium with zero taxes on commodities and the appropriate lump-sum tax for each individual. The calculation of the appropriate set of lump-sum taxes requires information on individuals which they have an incentive not to reveal — for example Mirrlees (1974) has shown that, where individuals differ in skills, it is likely that the first best will require utility to decrease with skill. It is then natural to ask how well one can do with a tax system which does not discriminate between individuals. This has led, following Mirrlees (1971), to the theory of optimum income taxation where we assume that only income

* This paper forms part of work on a programme of research on Taxation, Incentives and the Distribution of Income, financed by the Social Science Research Council. The paper has benefited from the comments of participants at a number of seminars since the first presentation of an early version in Sarlat in April 1976. I should like to thank in particular, Franklin Allen, Alan Carruth, Peter Diamond, Avinash Dixit, Peter Hammond, James Mirrlees, Kevin Roberts, Jesus Seade and an anonymous referee for very helpful comments.

0047-2727/82/000–000/$02.75

is observed and all individuals face the same income tax schedule. In this sense the system is anonymous. The schedule is then chosen to maximise welfare.

The optimum income tax formulation is not, however, without difficulties. The first problem applies to any non-linear form of income taxation. Knowledge of individual incomes is required and individuals would in general have an incentive to be misleading when reporting those incomes. This is not such a severe drawback for a linear system, with grants to individuals and a constant marginal tax rate, since incomes can be taxed at source. Of course if one abstracts from this problem an optimum non-linear system can never be worse than an optimum linear system.

Secondly, the calculation of the optimum income tax system is complex [see Mirrlees (1971)]. The calculation is inherently more difficult than that for lump-sum taxes since in the former problem individuals maximise with respect to the non-linear budget constraint associated with an income tax system and then this non-linear constraint is itself chosen in order to maximise the social welfare function.

Thirdly, we can, and do, discriminate between individuals in our tax and social security systems. Such discrimination is usually crude in its criteria and frequently only partially successful in that mistakes are made in classification, but the possibility does exist. Examples in the U.K. are discrimination in lump-sum grants between different categories of the disabled, or according to whether a woman with dependents has a permanent male cohabitor. An interesting example, but not one of overwhelming current significance, has been introduced by Hahn (1973) who drew attention to the different lump-sum taxes on dukes, squires and so on under the Poll Tax Act of 1660. Of greater practical importance is discrimination by age. And it would be possible to attempt to discriminate on the basis of some index of 'natural ability'. To ignore such possibilities may lead to a considerable sacrifice in welfare — that is something we wish to investigate.

These criticisms should not be taken, and are certainly not intended, as an attack on the optimum income taxation literature. They are intended to justify interest in a model where there is discrimination in that individuals of different types receive different lump-sum grants. However, in our model the authorities make mistakes in their classification of individuals and so do not reach the first best, with optimum lump-sum grants and taxes and zero marginal taxation. We shall assume that, using sampling techniques, the authorities know the proportion of people misclassified. Our question is how the knowledge that such mistakes exist should affect optimum policy. The size of error would depend on the discrimination being attempted: it might be small for age but large for an index of 'natural ability'. We shall assume for the most part that the proportion misclassified does not depend on the behaviour of the individuals but we shall be returning to this point below.

We shall be comparing two types of system: optimum income taxation and lump-sum taxation with errors. The lump-sum taxation scheme involves the difficulties of classifying individuals but requires no observation of personal incomes — we assume that income or output may be taxed proportionately at source but not in any other way. The income tax system involves the difficulties of observing personal incomes but can dispense with the office which classifies individuals. The optimum income tax can be calculated using knowledge only of the distribution of individuals amongst types. A sampling scheme could establish this.

The two types of system being compared each have their own information requirements, problems and administrative costs. It is hard to see how one could compare their administrative costs, and we shall not attempt to do so. To keep things simple we may assume that the administrative costs for the two schemes are the same and that the set-up costs of each is such that it would never be desirable to have both. Thus, we simply compare social welfare (based on individual levels) under the two schemes, abstracting from administration costs, in order to see which is preferable. It would be a straightforward modification of the analysis to assume different, but fixed, administrative costs for the two schemes (indeed, the results presented below could be used to examine this case).

It is clear that if we assume administration costs are zero then some combination of the two schemes would improve on either. There would still be interesting questions however as to whether we should want lump-sum taxes to be different from zero when superimposed on a non-linear tax scheme: the social costs from misclassification may be deemed to be too high. But this would be a different question from that addressed here and the required assumption about administration costs is unsatisfactory.

The model we shall be using will be very simple and there will be just two regimes which may be regarded as polar cases: but as such they are of interest and there is the further consideration of set-up costs which leads us to consider schemes which concentrate on just one feature. Notwithstanding the simplicity of the model, the issues to which the model is addressed, however, are of great importance and are central to both public finance and social administration: in summary, they concern the extent to which our tax and transfer system should be personalised (here lump-sum taxes) or anonymous (income taxation).

It is clear that if no errors are made then the lump-sum tax system is better than optimum income taxation (we assume that individuals get no utility or disutility from classification *per se*). On the other hand, if the government's classification scheme carries no information at all (it is completely random) then one would expect (see below section 2) that everyone would receive the same lump-sum grant and the system would be essentially the optimum linear income tax. It is clear that the optimum linear income tax is inferior to the optimum income tax. Thus, an important question will concern the degree of error which can be tolerated before the lump-sum system becomes inferior to that of the income tax. The question leads one naturally to the computation of solutions in particular models.

The analysis of a model where the government makes errors in the administration of grants and taxes is of interest for a reason additional to those already described. The consequence of the mistakes is that identical individuals are treated differently — for example a disabled individual whose disability is not officially recognised would receive a lower grant than other individuals with the similar disability. This dissimilar treatment violates the principle of horizontal equity, where we define the principle as stating that individuals who are *ex ante* identical should be treated *ex post* in an identical manner. Note that the utilitarian or Bergsonian calculus will take account of resultant utility levels but takes no account *per se* of dissimilar treatment. Thus, our model allows us to address questions of horizontal equity. Whilst these questions are not the central issue of this paper we shall return to them briefly in section 5 [for further discussion of these principles, particularly as they concern crime and punishment, see Carr-Hill and Stern (1976)].

The plan of the paper is as follows. We shall throughout be considering a model with just two types of individual who differ only in their labouring skill. The government attempts to classify individuals into the two groups but makes mistakes in so doing. Everyone has the same utility function of consumption of a single good, and labour in clock hours. Production is a function only of the total quantities in clock hours of the two types of labour. The objective is to maximise an increasing function, usually the sum, of utilities.

In section 2 we present the model with general functional forms and discuss the potential of different policies such as optimum income taxation, providing a diagrammatic treatment of the different tax schemes. In section 3 we discuss the calculation of the optimum income tax for our model and the relevance of standard theorems on optimum income taxation. Some interesting questions concerning those theorems emerge.

In section 4 we use constant elasticity of substitution utility functions and a Cobb–Douglas production function for the two types of labour. Analytic formulae for the optima do not appear to be possible even in the simple case and extensive numerical computations of optimum lump-sum taxation with errors and optimum income taxation are presented.

A brief discussion of considerations of equity is offered in section 5 and concluding remarks in section 6.

2. The model and the potential of different policies

The model is an elaboration of that used by Feldstein (1973). There are two types of individual, skilled and unskilled, indexed S and N. There is one consumption good which is produced by labour of the two types. Each person has the same utility function and an individual of type i maximises a utility function $U(C_i, L_i)$ subject to the constraint

$$C_i = (1-t)w_i L_i + G_j, \tag{1}$$

where w_i is the hourly wage of labour type i, t is the marginal tax rate, L_i is the amount of labour supplied, C_i his consumption and G_j is the lump-sum grant for individual type j. The indices i and j take the values S or N: if an individual is incorrectly classified $i \neq j$, if correctly $i = j$. There are β individuals of type S and $(2-\beta)$ individuals of type N. The model is static and whether individuals are skilled or unskilled, exogenous.

Mistakes are made in classifying individuals for their lump-sum grants and a proportion δ_i of each type is classified in the wrong group: thus, some skilled individuals receive a grant G_N and some unskilled a grant G_S. If, say, $G_S < G_N$, individuals of type S will have an incentive to try to be wrongly classified and individuals of type N to contest a classification as type S; thus, one would be interested in a model where the proportion misclassified is endogenous. A simple but tractable version is to put the proportion misclassified for the unskilled group to zero on the assumption, say, that they successfully contest a misclassification. We consider the three cases $\delta_S = \delta_N$, $\delta_S > \delta_N$, and $\delta_N = 0$ in our calculations in section 4, but for the present analysis the important feature is that δ_S and δ_N are constants.

Labour supply and consumption of an individual will depend on his type and whether he is correctly classified. Individuals who are correctly classified have a superscript 0 and those incorrectly classified a superscript 1. The labour supply functions derived from the maximisation of U subject to the constraint (1) are as follows:

$$L_S^0 = L((1-t)w_S, G_S), \tag{2}$$

$$L_S^1 = L((1-t)w_S, G_N), \tag{3}$$

$$L_N^0 = L((1-t)w_N, G_N), \tag{4}$$

$$L_N^1 = L((1-t)w_N, G_S). \tag{5}$$

Consumption levels then follow from (1). Note that there is no problem in identifying individuals at their place of employment so that each individual receives the correct hourly wage, which we assume is equal to the marginal product of an hour of the type of work supplied. The organisation distributing the lump-sum grant (which may, of course, be negative) is not the employer.

The average labour supply of type S and type N individuals respectively is

$$L_S = (1-\delta_S)L_S^0 + \delta_S L_S^1, \tag{6}$$

$$L_N = (1-\delta_N)L_N^0 + \delta_N L_N^1. \tag{7}$$

Output Y is a function of total labour supplies of each type βL_S and $(2-\beta)L_N$:

$$Y = F(\beta L_S, (2-\beta)L_N) \tag{8}$$

and

$$w_S = F_1, \tag{9}$$

$$w_N = F_2, \tag{10}$$

where subscripts to F denote partial derivatives. We assume F shows constant returns so that total payments to labour are equal to output. In thinking of S as denoting skilled and N unskilled we have in mind $w_S > w_N$. The wage rates are endogenous but we shall choose parameters so that w_S will usually be larger than w_N.

The government budget constraint is

$$[\beta(1-\delta_S)+(2-\beta)\delta_N]G_S+[\beta\delta_S+(2-\beta)(1-\delta_N)]G_N = tY - R, \tag{11}$$

where R is the revenue requirement. To keep things simple we exclude any possible benefits from government expenditure from the utility functions. R might then be interpreted, for example, as a fixed cost of production.

Eqs. (2)–(11) are a system of ten equations in twelve unknowns — the l.h.s. of eqs. (2)–(10) plus G_S, G_N, and t. Thus, given t and G_N we hope to solve for the other variables. The government's maximisation problem is therefore of two dimensions. We take t and G_N as the variables to be chosen, and it remains only to write down the maximand, W_ν

$$\begin{aligned}\nu W_\nu = {} & (1-\delta_S)\beta V^\nu(w'_S, G_S) + \delta_S \beta V^\nu(w'_S, G_N) \\ & + (1-\delta_N)(2-\beta)V^\nu(w'_N, G_N) + \delta_N(2-\beta)V^\nu(w'_N, G_S),\end{aligned} \tag{12}$$

where V is the indirect utility function corresponding to U, $w'_i = (1-t)w_i$, and ν is a parameter indicating the government's concern about inequality in utility levels. *If* U measures cardinal utility, then $\nu = 1$ is the utilitarian maximand ($\nu = -\infty$ corresponds to the maxi-min objective).

We have symmetry of all relevant properties about $\delta_i = \frac{1}{2}$ since classification with $\delta_i = \alpha$ and $\delta_i = 1-\alpha$ provides the same information with the labels reversed. If $\delta_i = \frac{1}{2}$, then the classification provides no information.

We have now described our model using general functional forms. Particular cases are discussed in sections 4 and 5. Before looking at these cases we examine some general statements about the potential of different kinds of policy.

Consider the 'first-best' U_S, U_N frontier describing the Pareto optima where there are no problems of misclassification and any desired lump-sum transfers can be made. (U_i is the utility level of the ith individual: i = S, N.) To keep things simple we assume for fig. 1 and our discussion of the various policies with general functional forms that $\beta = 1$ (equal population in the

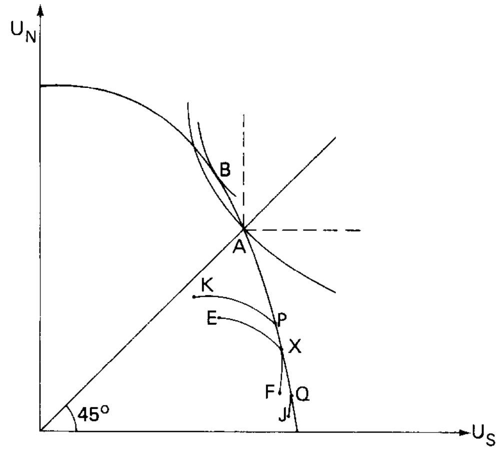

Fig. 1. The utility possibility frontiers with different tax schemes. Outer frontier is first best, lump-sum taxation; *KPXQJ* is non-linear income tax; *EXF* is linear income tax; government revenue requirement, *R*, is zero. *B* is first best optimum for quasi-concave symmetric social preferences represented by solid indifference curve. *A* is first best maxi-min optimum, *K* for the non-linear income tax and *E* for the linear income tax.

two groups) and $R=0$. It is straightforward to generalise the argument and results. With these assumptions any point on the frontier can be achieved with $G_S=-G_N$ and $t=0$. A point on the frontier is therefore identified by its G_N. The frontier is denoted in fig. 1, where we suppose that the minimum utility level is zero. Where U, W and F are concave, the first-best utility possibility frontier will also be concave.

Let us now ask what can be achieved by income taxation. No problems of misclassification arise since all invididuals face the same tax system. All individuals of type S make the same choice and have the same utility and similarly all unskilled individuals have the same (lower) utility level. Thus, with income taxation individuals allocate themselves to the different groups.

Consider the point on the first-best frontier given by $G_N=0$. The income tax schedule with zero grant to all individuals and zero marginal tax rate achieves this particular first-best optimum. Consider now some point on the first-best frontier given by $G_N>0$ and let the corresponding allocations be (C_i, L_i), $i=\mathrm{S, N}$. This is illustrated in fig. 2 where the consumption points are labelled H, I. It is clear that provided we have

$$U\left(C_S, \frac{w_S}{w_N} L_S\right) \leqq U(C_N, L_N) \tag{13a}$$

(so that type N individuals do not want to earn C_S post-tax)

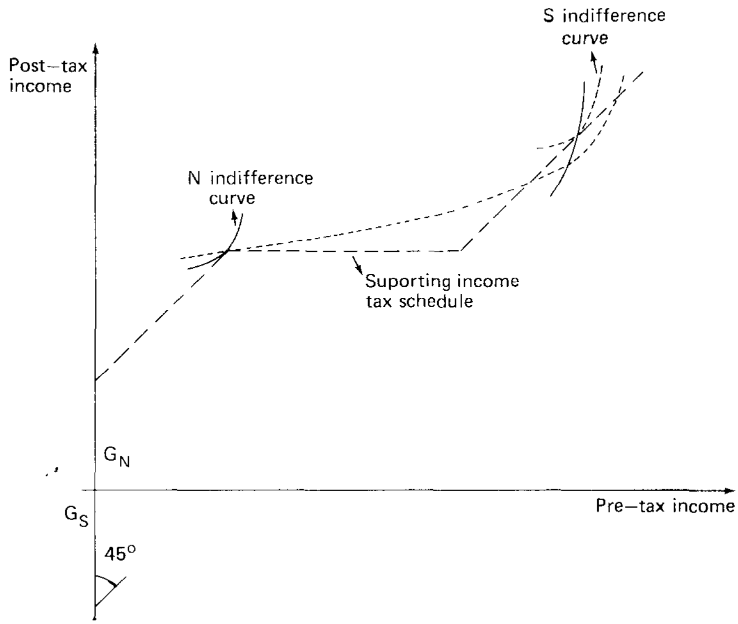

Fig. 2. A non-linear income tax schedule for a point on XP (fig. 1).

and

$$U\left(C_N, \frac{w_N}{w_S} L_N\right) \leqq U(C_S, L_S) \tag{13b}$$

(so that type S individuals do not want to earn C_N post-tax),

then this first best can be achieved by the income tax schedule given by the heavy dotted line in fig. 2. Similarly, provided the two inequalities (13) above are satisfied, we can reach points on the first-best frontier given by $G_N < 0$ using income taxation. We have therefore a portion of the first-best frontier, including the point given by $G_N = 0$ which can be achieved by income taxation. This portion lies entirely below the 45° line since any income tax schedule which is the same for all individuals must leave type S better off if $w_S > w_N$. For further discussion of the form of optimum income taxation in this model see section 3. Guesnerie and Seade (1982) examine the more general problem of optimum income taxation with n individuals. The conditions which characterise the optimum are generalisations of (13).

We suppose in fig. 1 that the income tax can achieve all points along the frontier between P and Q, given by $G_N^P(>0)$ and $G_N^Q(<0)$, respectively. For $G_N > G_N^P$, (13b) is violated, and for $G_N < G_N^Q$, (13a) is violated. It is possible that Q coincides with the point where the first-best frontier meets the axis.

Recall that we are assuming no government revenue requirement, $R = 0$, for this discussion. If $R > 0$ then the point $G_N = G_S = 0$ with $t = 0$ is no longer feasible: a point on the first-best frontier can still be characterised by its G_N (with $t = 0$) but now $G_S = R - G_N$. We know (13a) and (13b) can be satisfied on the frontier by $G_N = 0$ for $R = 0$ thus given sufficient continuity assumptions income taxation will be able to reach a portion of the first-best frontier for small R. For larger R there is no guarantee of this.

Let us suppose that the optimum income tax when the objective is maxi-min achieves utility levels represented by the point K in fig. 1. The maxi-min objective is represented by right-angle indifference curves in utility space with the kink along the 45° line. The optimum with the maxi-min objective and optimum income taxation is represented by the point J to the right and below Q. The possible frontier in utility space which can be achieved by income taxation is then $KPQJ$. The frontier is horizontal at K and vertical at J.

The best that linear income taxation (that is, a positive or negative grant G uniform across individuals, together with a wage tax or subsidy) can do is given by the frontier EF. E corresponds to maxi-min where the frontier is horizontal and F to maxi-min where the frontier is vertical. It touches the other two frontiers at the point X corresponding to $G = 0$ (although this would not be the case for $R > 0$).

We have in fig. 1 three feasibility frontiers in utility space according to whether lump-sum, non-linear income, or linear income taxation is used. In an optimum system of taxation of a given type the social indifference curve in (U_S, U_N) space touches the appropriate frontier. Where the social welfare function is symmetric in utilities then indifference curves have gradient -1 along the 45° line.

Mirrlees (1974) has shown that if and only if leisure is a normal good, then the first-best optimum has utility decreasing in skill. His model had a continuous distribution of skills and exogenous relative wage rates but an examination of the proof of this proposition shows that the argument can be applied to the case with endogenous wages and a discrete distribution of skills. The first-best optimum in our case then has $U_S < U_N$ (see point B in fig. 1). This implies (with symmetric quasi-concave indifference curves) that along the first-best frontier $-dU_N/dU_S > 1$ at the point A where the 45° line meets the first-best frontier. If social preferences are represented by W [see eq. (12) with $\delta_i = 0$] then $\nu = 1$ corresponds to straight line indifference curves gradient -1; the curvature (or 'concavity') of the indifference curves increases as ν decreases. The optimum under the maxi-min objective ($\nu = -\infty$) with lump-sum taxation occurs at A where $U_S = U_N$ (see fig. 1). Note that A gives a higher level of social welfare than optimum income taxation, corresponding to a point on PK, for a welfare function which is symmetric and quasi-concave in utilities.

The case where lump-sum taxation is possible but where mistakes are made in classification cannot be represented simply in utility space since there are four levels of utility [see eq. (12)]. We can however make some simple comparisons of welfare levels that are feasible in this case with those that arise from the three forms of taxation just considered. When $\delta_i = 0$, no mistakes in classification, we can reach the first-best and the whole outer utility possibility frontier of fig. 1 is available. If the government so chooses it can give the same grant to everyone so the utility possibility frontier given by linear taxation and represented by *EXF* will be feasible whatever δ_i. When $\delta_i = \frac{1}{2}$, each i, the classification carries no information and one might expect that *EXF* represents the best one can do. One cannot assert in general, however, that when $\delta_i = \frac{1}{2}$, each i, the optimum linear income tax represents the best solution since whilst the maximand is concave in G_S and G_N there is no guarantee that the constraints [which include the labour supply functions (2)–(5)] have convenient concavity properties. We shall be calculating in section 4 the optimum G_S and G_N for $\delta_i = \frac{1}{2}$, each i, in a particular model and it transpires that the optimum G_S and G_N are indeed equal. Note that the argument presented by Stiglitz (1976) for random taxation does not apply here since he was dealing with commodity taxes and exploiting the quasi-convexity of the indirect utility function with respect to prices. We are considering here the possibility of different lump-sum transfers for *ex ante* identical individuals but not the possibility of different prices.

Crudely speaking then, and with the qualifications given above, one can illustrate the possibilities as δ varies (we suppose $\delta_S = \delta_N = \delta$ here) by saying that the welfare levels available when $\delta = \frac{1}{2}$ are represented by points along *EXF*, when $\delta = 0$ the outer first-best frontier, and for δ between 0 and $\frac{1}{2}$ the possibilities lie somewhere between. We see that with symmetric social preferences and with low enough δ, lump-sum taxation, even though classification is occasionally mistaken, can do better than optimum non-linear income taxation, but with δ near a half the attempt to discriminate will not do as well as optimum income taxation. [There will, however, always exist non-symmetric preferences such that optimum income taxation reaches the full optimum and any $\delta > 0$ will make lump-sum taxation worse than optimum income taxation — this is where the optimum lies along *PQ* on the outer boundary.] We calculate in section 5 the value $\hat{\delta}$ of δ which gives equal welfare under the two regimes of optimum income taxation and optimum lump-sum taxation with errors. For $\delta < \hat{\delta}$ the lump-sum scheme will be preferable and for $\delta > \hat{\delta}$ the optimum taxation scheme.

3. Optimum income taxation[1]

In the previous sections we have explained that a major objective of this paper is to compare the potential of income taxation with that of lump-sum

[1]Discussions with Avinash Dixit and Jesus Seade have been of particular help with this section.

taxation. In order to do this we must calculate optimum income taxation. The calculation of the optimum income tax can be quite complex [see Mirrlees (1971)] and it is natural to ask whether one can invoke the standard theorems on optimum income taxation to simplify the analysis. There are three general theorems which one might think would be of assistance and these are as follows. First, the optimum tax rate should lie between zero and one; secondly, the marginal tax rate on the highest income should be zero; and thirdly, the marginal tax rate on the lowest income should be zero [see Mirrlees (1971, propositions 2 and 3) for the first result, and Seade (1977, theorem 2) for the second and third]. These results require a number of assumptions and we shall not go into the detail here, but it is widely thought that they provide the general statements that are available on the shape of tax schedules and that to say more one has to go to particular functional forms. It transpires that none of these three results holds for our model and it is interesting to see why.

To derive optimum income taxation for our model we proceed as follows. It is clear from fig. 2 that any allocation satisfying the production constraint and inequalities (13) can be decentralised by an income tax system. It is also clear that any feasible allocation which is the outcome of an income tax system must satisfy these conditions. The optimum income tax will, therefore, be given by the solution to:

$$\underset{C_S, L_S, C_N, L_N}{\text{maximise}}\ W(U(C_S, L_S), U(C_N, L_N))$$

$$U(C_S, L_S) - U\left(C_N, \frac{F_2}{F_1} L_N\right) \geqq 0, \tag{14}$$

$$-C_S - C_N + F(L_S, L_N) \geqq 0. \tag{15}$$

To keep things simple we have put $\beta = 1$ and $R = 0$ (the results of this section are easily extended to cover different β and R). Eq. (14) follows from (13b) after substituting from (9) and (10). With a lower social marginal utility of consumption for the skilled (see below) it will be (13b) rather than (13a) that will be relevant. $W(\cdot)$ is the social welfare function. We assume that all of C_S, L_S, C_N, and L_N are strictly positive at the optimum (note that with an elasticity of substitution less than or equal to one $L_i = 0$, $i = \mathrm{S}$ or N, would imply zero output). Taking Lagrange multipliers λ and μ for constraints (14) and (15) and differentiating with respect to C_S and L_S we have

$$W_S \frac{\partial U}{\partial C}(C_S, L_S) + \lambda \frac{\partial U}{\partial C}(C_S, L_S) - \mu = 0, \tag{16}$$

$$W_S \frac{\partial U}{\partial L}(C_S, L_S) + \lambda \frac{\partial U}{\partial L}(C_S, L_S) - \lambda L_N \frac{\partial h}{\partial L_S}(L_S, L_N)$$

$$\times \frac{\partial U}{\partial L}(C_N, hL_N) + \mu F_1 = 0, \tag{17}$$

where W_S is the partial derivative of W with respect to the utility of the skilled person, $\partial U/\partial C$ and $\partial U/\partial L$ denote partial derivatives of U with respect to its first and second arguments, and $h(L_S, L_N) = F_2(L_S, L_N)/F_1(L_S, L_N)$.

From (16) and (17) we have

$$1 + \frac{1}{F_1} \frac{(\partial U/\partial L)(C_S, L_S)}{(\partial U/\partial C)(C_S, L_S)} = \frac{\lambda}{\mu F_1} L_N \frac{\partial h}{\partial L_S} \frac{\partial U}{\partial L}(C_N, hL_N). \tag{18}$$

We can interpret the left-hand side of (18) as the marginal tax rate on the more skilled individual (it is one minus the ratio of the marginal rate of substitution between consumption and leisure to the wage). From the right-hand side of (18) we see that this marginal tax rate must be *negative*. We examine the elements of the right-hand side of (18). F_1 and μ being the marginal product of skilled labour and the shadow price on the resource constraint, must be positive, as is L_N. Since F is homogeneous and concave and there are just two factors, $\partial h/\partial L_S$ is positive (an increase in the quantity of skilled labour increases the marginal product of unskilled labour). If there is a disutility of labour $(\partial U/\partial L)(C_N, hL_N)$ is negative. There remains only to consider the sign of λ.

We shall argue that (14) must bind provided

$$W_S \frac{\partial U}{\partial C}(C_S, L_S) < W_N \frac{\partial U}{\partial C}(C_N, L_N) \tag{19}$$

at the optimum (W_N is analogous to W_S). We use fig. 2. If (14) does not bind then, holding labour supplies constant, we can make a lump-sum transfer from S to N (vertical opposite shifts of the consumption points H and I) whilst preserving the conditions for decentralisation using income taxation. If (19) holds, there is an increase in welfare. Hence, the Lagrange multiplier λ must be positive. [For any solution of (14) and (15) together with the first-order conditions one can check, *ex post*, whether (19) does in fact hold.] Thus, the marginal tax rate at the top is negative.

We can give an intuitive interpretation of this result as follows. As with most interpretations of first-order conditions for optimality we decompose the effects of a change into marginal costs and marginal benefits and at the optimum these should be equal. Consider the consequences of an increase in L_S. We have the benefit F_1, the marginal product and the cost in terms of the utility of forgone leisure which is given in terms of output by the marginal rate of substitution of consumption for leisure. Now in this case we have the *extra* benefit that the increase in L_S raises w_N/w_S the relative wage of the unskilled with the consequence that constraint (14) is relaxed (it would now take the skilled relatively longer to earn the same income as the unskilled). The gains to the relaxation of the constraint are, as we have just seen in our argument above that it should bind, that a beneficial lump-sum transfer is permitted. At the optimum this gain plus the marginal product should be equal to the marginal rate of substitution. Hence, the marginal rate of substitution *exceeds* the marginal product and at the optimum we have a marginal subsidy.

The importance of the endogeneity of relative wages in the above argument is clear. If there is an infinite elasticity of substitution between the two types of labour [as in Mirrlees (1971) and Seade (1977)], then w_N/w_S is constant and we are back to the standard result that the marginal tax rate at the top should be zero.

We can look at the marginal tax rate at the bottom by examining the first-order conditions for C_N and L_N. We have

$$W_N \frac{\partial U}{\partial C}(C_N, L_N) - \lambda \frac{\partial U}{\partial C}(C_N, hL_N) - \mu = 0, \tag{20}$$

$$W_N \frac{\partial U}{\partial L}(C_N, L_N) - \lambda \frac{\partial U}{\partial L}(C_N, hL_N)\left(h + L_N \frac{\partial h}{\partial L_N}\right) + \mu F_2 = 0. \tag{21}$$

From (20) and (21) we have

$$\frac{-1}{F_2} \frac{\dfrac{\partial U}{\partial L}(C_N, L_N)}{\dfrac{\partial U}{\partial C}(C_N, L_N)} = \frac{1 - \dfrac{\lambda}{\mu F_2} \dfrac{\partial U^{SN}}{\partial L}\left(h + L_N \dfrac{\partial h}{\partial L_N}\right)}{1 + \dfrac{\lambda}{\mu} \dfrac{\partial U^{SN}}{\partial C}}, \tag{22}$$

where $\partial U^{SN}/\partial L = (\partial U/\partial L)\ (C_N, hL_N)$, the latter partial derivative being with respect to the second argument, and similarly for $\partial U^{SN}/\partial C$. If the right-hand side of (22) is less than one we have a positive marginal tax rate at the bottom, and if it is greater than one a marginal subsidy. We are speaking here of the marginal tax rate as being derived from a comparison of the marginal rate of substitution between consumption and leisure with the pre-tax wage. A little care with this interpretation is necessary however since it is clear from the fact that (14) must bind, together with fig. 2, that the tax schedule cannot be differentiable at (C_N, w_N, L_N) — the point I. The reason is that the tax schedule to the left of point I must be steeper than the unskilled's indifference curve through I (so that he will not choose a point to the left of I) and to the right of I must be shallower than the skilled's indifference curve through I (since otherwise he will prefer a point just to the right of I to the point H). Clearly, there are many schedules which will do the decentralisation but they must all be non-differentiable at I. It is on this understanding that we speak of the 'marginal tax rate at bottom'.

Writing

$$\mu_{\mathrm{N}} = -\frac{\partial U}{\partial L}(C_{\mathrm{N}}, L_{\mathrm{N}}) \Big/ \frac{\partial U}{\partial C}(C_{\mathrm{N}}, L_{\mathrm{N}}),$$

$$\mu_{\mathrm{SN}} = -\frac{\partial U}{\partial L}(C_{\mathrm{N}}, hL_{\mathrm{N}}) \Big/ \frac{\partial U}{\partial C}(C_{\mathrm{N}}, hL_{\mathrm{N}})$$

and

$$\alpha = \frac{\lambda}{\mu}\frac{\partial U^{\mathrm{SN}}}{\partial C},$$

we have from (22)

$$\frac{\mu_{\mathrm{N}}}{F_2} = \frac{1 + \alpha \dfrac{\mu_{\mathrm{SN}}}{F_2}\dfrac{\partial(hl_{\mathrm{N}})}{\partial L_{\mathrm{N}}}}{1 + \alpha}. \tag{23}$$

From (23) we have:

$$\text{either } 1 \leqq \frac{\mu_{\mathrm{N}}}{F_2} \leqq \frac{\mu_{\mathrm{SN}}}{F_2}\frac{\partial(hL_{\mathrm{N}})}{\partial L_{\mathrm{N}}}$$

$$\text{or } 1 \geqq \frac{\mu_{\mathrm{N}}}{F_2} \geqq \frac{\mu_{\mathrm{SN}}}{F_2}\frac{\partial(hl_{\mathrm{N}})}{\partial L_{\mathrm{N}}}, \tag{24}$$

and (24) imples that

$$\frac{\mu_{\mathrm{N}}}{F_2} \gtrless 1 \quad \text{as} \quad \frac{\mu_{\mathrm{N}}}{\mu_{\mathrm{SN}}} \lessgtr \frac{\partial(hL_{\mathrm{N}})}{\partial_{\mathrm{LN}}}. \tag{25}$$

We can see from (25) that we will usually have $\mu_{\mathrm{N}}/F_2 < 1$, i.e. a positive marginal tax rate at the bottom since if consumption is normal, $\mu_{\mathrm{N}} > \mu_{\mathrm{SN}}$, whereas $\partial(hL_{\mathrm{N}})/\partial L_{\mathrm{N}} \leqq h < 1$. Note that the second inequality in (25) involves a comparison between curvature of indifference curves and curvature of isoquants. For a Cobb–Douglas production function $\partial(hL_{\mathrm{N}})/\partial L_{\mathrm{N}} = 0$ and we certainly have $\mu_{\mathrm{N}}/F_2 < 1$ [this can be seen directly from (22)]. If the relative wage is exogenous, so that $\partial(hL_{\mathrm{N}})/\partial L_{\mathrm{N}} = h$ then, as we have seen, normality of consumption is sufficient to guarantee a positive marginal tax rate at the bottom.

We shall compute optimum income taxation by using a numerical algorithm to maximise the social welfare function subject to constraints (14) and (15). With specific functional forms, (14) and (15) give L_{S} and L_{N} as function of C_{S} and C_{N} (we know the constraints will bind at the optimum) and we can then vary C_{S} and C_{N} to maximise.

Given that the result that the marginal tax rate at the top is negative is in contrast to previous results on optimum income taxation and that some might object directly to a marginal subsidy at the top, we also computed optimum income taxation subject to the constraint that the marginal tax rate at the top should equal one.

4. Cobb–Douglas production function and CES utility function

We suppose that the production function has the special form

$$Y = k_0(\beta L_S)^{\gamma}((2-\beta)L_N)^{1-\gamma} \tag{26}$$

and the utility function $U(\cdot)$ has the constant elasticity of substitution form

$$U(C, L) = [(1-\alpha)C^{-\mu} + \alpha(1-L)^{-\mu}]^{-1/\mu}. \tag{27}$$

The elasticity of substitution ε is equal to $1/(1+\mu)$. It appears that a value of ε equal to $\frac{1}{2}$, and thus $\mu = 1$, has some empirical plausibility [see Stern (1976)]. It is again straightforward to calculate the labour supply functions of eqs. (2)–(5) and they are not presented explicitly.

We explained in section 2 that we can, by solving eqs. (2)–(11), think of the maximand (12) as a function of (t, G_N). The optimum is then calculated by numerically searching over the two-dimensional space (t, G_N). Details of the procedure are given in Stern (1979).

The optimum linear income tax $G_S = G_N = G$ can be calculated by searching over a one-dimensional space — the government budget constraint gives a relation between the marginal tax rate t and the grant G. This is called the constrained case. Optimum non-linear taxation was calculated as described in section 3.

The maximand, social welfare, was calibrated using the notion of the equally-distributed, leisurely-equivalent consumption ${}^0C$, defined as follows. Given a certain pattern of utilities resulting from t and G_N, we assign to social welfare W_ν the number ${}^0C$ which is that consumption which, if equally distributed, and when hours of work were zero for everyone, would give social welfare level W_ν. Formally

$$2U^\nu({}^0C, 0) = \nu W_\nu, \tag{28}$$

where in the CES case $U(C, 0) = [(1-\alpha)C^{-\mu} + \alpha]^{-1/\mu}$. It is clear that ${}^0C$ will depend on ν in general.

A convenient and interesting standard of comparison for welfare levels achieved under lump-sum taxation with errors in classification is that point on the first-best frontier with $U_S = U_N$ — the point A in fig. 1. This is the first-best when the objective is maxi-min (corresponding to $\nu = -\infty$). The value of ${}^0C$ for the point A is independent of ν, for if U^* is the common level of utility at A we have $U({}^0C, 0) = U^*$ for all ν. The level of welfare at A, ${}^0C_A$, will be lower than for the first-base optimum $(\nu \neq -\infty)$ but higher than that for optimum income taxation (see fig. 1).

There are four types of optima to be calculated: lump-sum taxation with errors, non-linear income taxation, linear income taxation, and first-best maxi-min. Errors in classification are relevant only for the first of these. There are a number of parameters to be varied: ν, which measures attitudes to inequality (see below); R, the government revenue requirement; ε, the elasticity of substitution in preferences between consumption and leisure; γ, the (gross-of-tax) competitive share of the skilled; and β, (twice) the proportion of the skilled in the population. In addition we must examine the effects of any difference between the errors in classifying the skilled δ_S and the unskilled δ_N. For further computations see Carruth (1982). With the various different optima and parameters we have a large number of cases to consider: the presentation of the results is organised as follows.

We define a 'base run': $\nu=-1$, $R=0$, $\varepsilon=0.5$, $\gamma=0.67$, and $\beta=1$. Parameters are varied one at a time from this base holding the values of the other parameters constant. For the base run we put $\delta_S=\delta_N=\delta$. We shall provide an interpretation of the magnitudes of the parameters after the results have been presented. For each set of values of the parameters the four different optima are calculated; optimum taxation with errors is calculated for $\delta=0, 0.1, 0.2, \ldots, 0.5$. The parameters α and k_0, set at 0.5 and 1, respectively, were not varied.

The results for the base-run are presented in table 1(a), and for $\nu=-2$ and $\nu=0.97$ in table 1(b) and (c) [$\nu=1.0$ resulted in convergence problems — for a discussion of algorithms, accuracy and convergence see Stern (1979)]. A graph showing these three cases is provided in fig. 3. Variation in the parameters R, ε, λ and β is depicted in table 2 and the effect of allowing $\delta_S>\delta_N$ is shown in table 3. The case of $\delta_N=0$ is of particular interest (see section 2) and these results are presented in table 4 and illustrated with fig. 4.

For each set of values of the parameters a grid was produced for the two-dimensional searches so that optimisation could be checked by eye. One cannot guarantee concavity of the maximand in either optimum lump-sum taxation with errors or optimum income taxation but no problems of multiple local maxima were encountered. Computing times and resources were very small.

There are a number of checks on the reliability of our computations. First, for $\nu=-1$ the social direct utility function $(1/\nu)U^{\nu}$ is $-U^{-1}$ or $-((1-\alpha)C^{-1}+\alpha(1-L)^{-1})$, where $\mu=1(\varepsilon=\frac{1}{2})$. The first-best optimum requires equality of the social marginal utility of consumption and hence in this case of consumption itself. It is comforting that the optimisation routine did indeed give this result. Secondly, reassurance of the accuracy of the computations is offered by the replication of the results of Feldstein (1973). Thirdly, we have the very close proximity between the optimum linear income tax and the optimum lump-sum tax with errors when $\delta=0.5$ (see tables 1 and 2). We argued in section 2 that we should expect (but could not guarantee) these two optima to be the same.

We now examine the selection of parameter values. The interpretation of the parameter ν is most easily seen if we work with indirect rather than direct utility functions. The indirect utility function corresponding to $U(C, L)$ as in (27) is

$$V(w', G) = (w' + G)[(1-\alpha)^{\varepsilon} + \alpha^{\varepsilon} w'^{(1-\varepsilon)}]^{-1/(1-\varepsilon)}, \tag{29}$$

where w' is the net-of-tax wage and G is the lump-sum income. V is proportional to 'full' income (value of labour endowment of one unit plus lump-sum income) since U is homogeneous degree one in leisure and consumption.

The choice of ν is a matter of selection of value judgements. A value of $\nu = 1$ in (12) corresponds to constant social marginal utility of full income (for fixed w') and is in this sense not egalitarian. Lower values of ν represent diminishing social marginal utility of full income — for example with $\nu = -1$ the social marginal utility of full income decreases as the square of full income. Elsewhere [Stern (1977)] I have argued that $\nu = -1$ for optimum saving and taxation generate 'realistic' policies. Maxi-min corresponds to $\nu = -\infty$.

The revenue requirement R may be compared with total output or GNP: for most of the calculations output, which is endogenous, was between 0.5 and 0.7. Hence, a government revenue requirement of 0.1 represents something between 14 and 20% of GNP — it should be remembered that R represents government expenditure on goods and services only and not transfer payments. The elasticity of substitution between consumption and leisure $\varepsilon[1/(1+\mu)]$ is $\frac{1}{2}$ in the base run. I have argued elsewhere, see Stern (1976), that this conforms well with many empirical estimates of labour supply schedules. The parameter γ, set at 0.67 for the base-run, is the main distinguishing feature of the skilled. If the total labour supply of the skilled and unskilled were equal then the gross-of-tax wage of the skilled would be twice that of the unskilled. The other difference between the skilled and unskilled is in their numbers: $\beta = 1$ corresponds to equal numbers in each group; with $\beta = 0.5$ there are three times as many in the unskilled group as in the skilled group.

We turn now to a discussion of the results and begin with the effects of variations in the parameter ν — see table 1 and fig. 3. The first point of interest is the striking difference between the optima for ν of (or close to) 1 and ν of -1 and -2. For lump-sum taxation with errors the degree of misclassification can be large ($\delta = 0.2$ means that only 80% of the population are correctly classified whereas random classification would achieve 50%) yet still imply small tax rates and small losses in welfare (as compared with the first best) for ν close to 1. However, tax rates and losses in welfare

Table 1

(a) The base run, $v = -1$

Optimum lump-sum taxation with errors

δ	t	G_N	G_S	w_N	w_S	Y	^0C
0	0	0.1002	−0.1003	0.3752	0.6289	0.5897	0.2094
0.1	0.2048	0.1334	0.0188	0.3638	0.6386	0.5593	0.2048
0.2	0.3149	0.1410	0.0286	0.3492	0.6516	0.5384	0.2009
0.3	0.3804	0.1365	0.0627	0.3360	0.6641	0.5239	0.1979
0.4	0.4155	0.1253	0.0888	0.3273	0.6727	0.5153	0.1960
0.5	0.4264	0.1094	0.1092	0.3242	0.6758	0.5125	0.1954

Optimum non-linear income taxation

$(1-\mathrm{MTR}_N)$	$(1-\mathrm{MTR}_S)$	G_N	G_S	w_N	w_S	Y	^0C
0.4263	1.0375	0.1094	−0.0689	0.3620	0.6401	0.5709	0.2054

Optimum linear income taxation

t	G	w_N	w_S	Y	^0C
0.4263	0.1093	0.3242	0.6758	0.5126	0.1954

First-best maxi-min

G_N	G_S	w_N	w_S	Y	^0C
0.0823	−0.0823	0.3592	0.6426	0.5870	0.2086

(b) The base run, $v = -2$

Optimum lump-sum taxation with errors

δ	t	G_N	G_S	w_N	w_S	Y	^0C
0	0	0.0947	−0.0947	0.3702	0.6331	0.5889	0.2092
0.1	0.2635	0.1355	0.0092	0.3607	0.6413	0.5493	0.2032
0.2	0.3694	0.1421	0.0529	0.3482	0.6525	0.5279	0.1989
0.3	0.4272	0.1385	0.0812	0.3375	0.6627	0.5143	0.1959
0.4	0.4575	0.1299	0.1018	0.3306	0.6694	0.5065	0.1941
0.5	0.4664	0.1176	0.1175	0.3281	0.6719	0.5041	0.1935

Optimum non-linear income taxation

$(1-\mathrm{MTR}_N)$	$(1-\mathrm{MTR}_S)$	G_N	G_S	w_N	w_S	Y	^0C
0.6714	1.0447	0.1159	−0.0712	0.3665	0.6363	0.5689	0.2051

Optimum linear income taxation

t	G	w_N	w_S	Y	^0C
0.4665	0.1176	0.3282	0.6719	0.5041	0.1935

First-best maxi-min

G_N	G_S	w_N	w_S	Y	^0C
0.0823	−0.0823	0.3592	0.6426	0.5870	0.2086

Table 1 (*contd.*)

(c) The base run, $\nu = 0.97$

Optimum lump-sum taxation with errors

δ	t	G_N	G_S	w_N	w_S	Y	${}^0C$
0	0	0.2293	−0.2293	0.5088	0.5413	0.6106	0.2151
0.1	0.0068	0.2786	−0.2745	0.5011	0.5454	0.6085	0.2147
0.2	0.0231	0.3480	−0.3341	0.4860	0.5537	0.6042	0.2138
0.3	0.0763	0.4235	−0.3784	0.4481	0.5763	0.5913	0.2115
0.4	0.1987	0.3842	−0.2727	0.3679	0.6351	0.5609	0.2059
0.5	0.2521	0.0685	0.0685	0.3095	0.6915	0.5432	0.2010

Optimum non-linear income taxation

$(1-\text{MTR}_N)$	$(1-\text{MTR}_S)$	G_N	G_S	w_N	w_S	Y	${}^0C$
0.8150	1.0209	0.0905	−0.0634	0.3512	0.6497	0.5761	0.2063

Optimum linear income taxation

t	G	w_N	w_S	Y	${}^0C$
0.2525	0.0686	0.3096	0.6914	0.5431	0.2010

First-best maxi-min

G_N	G_S	w_N	w_S	Y	${}^0C_A$
0.0823	−0.0823	0.3592	0.6426	0.5870	0.2086

Notes

Notation

G_i, i = N, S = lump-sum grant intended for individuals type i.
w_i, i = N, S = wage rate for individuals type i.
t = marginal tax rate.
δ = proportion mis-classified.
Y = output.
${}^0C$ = equally-distributed leisurely-equivalent level of welfare [see eq. (28)].
ν = parameter measuring attitudes to inequality [see eq. (12)].

The different optima

Optimum lump-sum taxation with errors: where $\delta > 0$ some individuals receive incorrect grants; t and G_N are chosen to maximise the social welfare function.

Optimum non-linear income taxation: every individual faces the same income tax schedule although they differ in their wage rates; $1-\text{MTR}_i$ is one minus the marginal tax rate and G_i is the lum-sum grant as given by the tangent to the indifference curve for individuals type i.

Optimum linear income taxation: G is the grant common to all individuals; one degree of freedom in the optimisation.

First-best maxi-min: we find the point on the first best frontier where $U_S = U_N$; each individual has welfare level ${}^0C_A$.

Other parameters

R, ε, γ, and β are the government revenue requirement, the elasticity of substitution between consumption and leisure, the Cobb–Douglas parameter in the production function, and (half) the proportion of individuals of each type, respectively. For the results of table 1 we have $R = 0$, $\varepsilon = 0.5$, $\gamma = 0.67$ and $\beta = 1$. For variation in these parameters see table 2.

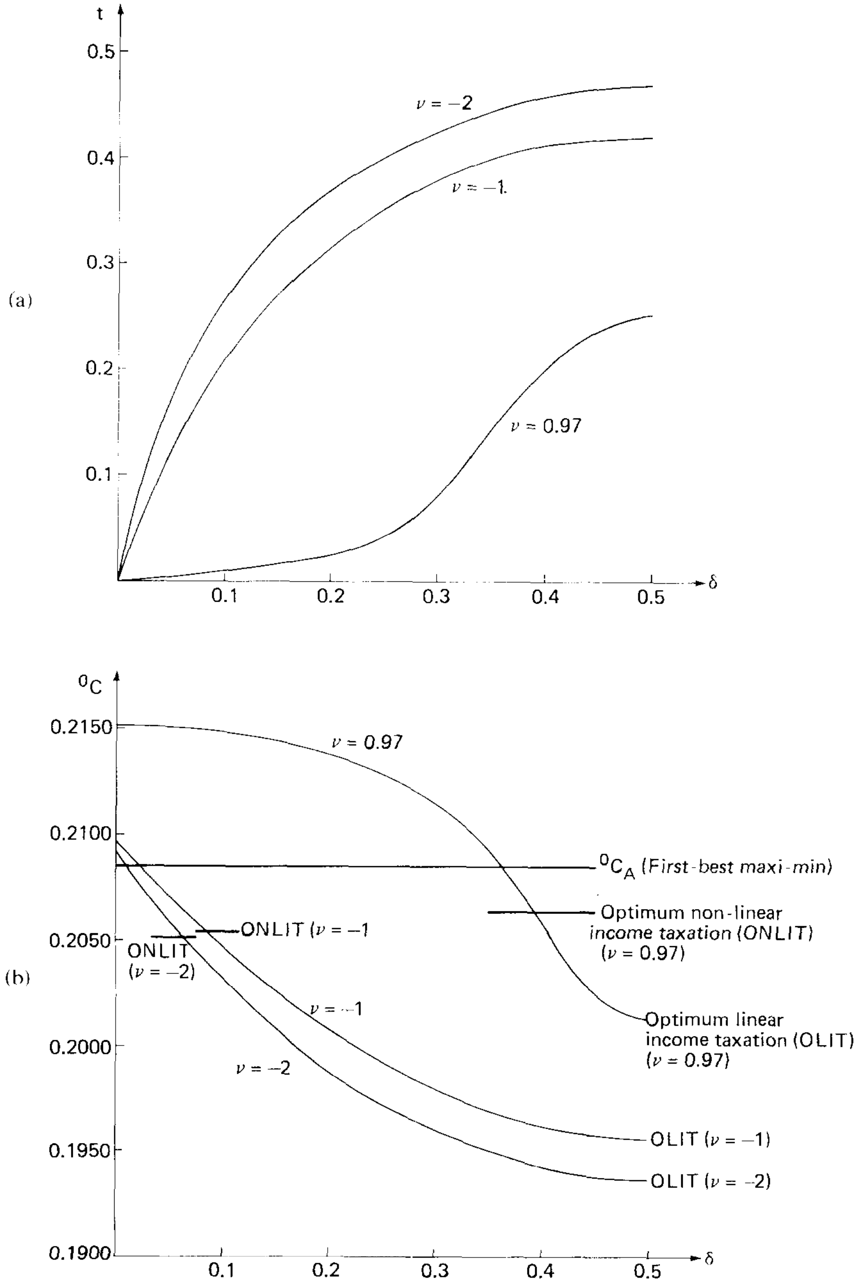

Fig. 3. (a) Optimum tax rate against proportion misclassified (see table 1(a)–(c)). (b) Welfare level at optimum against proportion misclassified (see table 1(a)–(c)).

Table 2

(a) Variation of parameters, $R = 0.1$

Optimum lump-sum taxation with errors

σ	t	G_N	G_S	Y	^{0}C
0	0	0.0572	−0.1572	0.6307	0.1847
0.1	0.2650	0.1029	−0.0454	0.5943	0.1791
0.2	0.3799	0.1118	0.0059	0.5729	0.1748
0.3	0.4434	0.1081	0.0397	0.5588	0.1717
0.4	0.4962	0.1006	0.0707	0.5466	0.1697
0.5	0.4858	0.0832	0.0832	0.5483	0.1691

Optimum non-linear income taxation

$(1-\mathrm{MTR}_N)$	$((1-\mathrm{MTR}_S)$	G_N	G_S	Y	^{0}C
0.6862	1.0342	0.0762	−0.1268	0.6120	0.1806

Optimum linear income taxation

t	G	Y	^{0}C
0.4858	0.0832	0.5483	0.1691

First-best maxi-min

G_N	G_S	Y	$^{0}C_A$
0.1406	−0.1406	0.6283	0.1840

(b) Variation of parameters, $\varepsilon = 0.7$

Optimum lump-sum taxation with errors

δ	t	G_N	G_S	Y	^{0}C
0	0	0.0888	−0.0888	0.5561	0.1709
0.1	0.1379	0.1098	−0.0370	0.5284	0.1672
0.2	0.2323	0.1153	0.0022	0.5061	0.1638
0.3	0.2933	0.1099	0.0338	0.4899	0.1609
0.4	0.3280	0.0977	0.0598	0.4801	0.1590
0.5	0.3388	0.0808	0.0808	0.4769	0.1584

Optimum non-linear income taxation

$(1-\mathrm{MTR}_N)$	$(1-\mathrm{MTR}_S)$	G_N	G_S	Y	^{0}C
0.7105	1.0591	0.0917	−0.0618	0.5344	0.1669

Optimum linear income taxation

t	G	Y	^{0}C
0.3387	0.0808	0.4770	0.1584

First-best maxi-min

G_N	G_S	Y	$^{0}C_A$
0.0712	−0.0712	0.5537	0.1701

Table 2 (*contd.*)

(c) Variation of parameters, $\gamma = 0.8$

Optimum lump-sum taxation with errors					
δ	t	G_N	G_S	Y	^0C
0	0	0.1798	0.1798	0.5995	0.2152
0.1	0.4398	0.2021	0.0277	0.5225	0.2018
0.2	0.5296	0.1894	0.0726	0.4946	0.1946
0.3	0.5727	0.1734	0.1004	0.4781	0.1899
0.4	0.5934	0.1568	0.1216	0.4691	0.1873
0.5	0.5997	0.1398	0.1398	0.4663	0.1864
Optimum non-linear income taxation					
$(1-\mathrm{MTR}_N)$	$(1-\mathrm{MTR}_S)$	G_n	G_S	Y	^0C
0.6234	1.0143	0.1587	−0.1233	0.5700	0.2080
Optimum linear income taxation					
t	G	Y	^0C		
0.5997	0.1398	0.4663	0.1864		
First-best maxi-min					
G_N	G_S	Y	^0C_A		
0.1446	0.1446	0.5902	0.2124		

(d) Variation of parameters, $\beta = 0.5$

Optimum lump-sum taxation with errors					
δ	t	G_N	G_S	Y	^0C
0	0	0.1306	−0.3918	0.4664	0.1742
0.1	0.6494	0.1498	0.0447	0.3642	0.1510
0.2	0.6767	0.1420	0.0754	0.3508	0.1458
0.3	0.6877	0.1345	0.0939	0.3440	0.1431
0.4	0.6927	0.1267	0.1073	0.3405	0.1417
0.5	0.6938	0.1178	0.1178	0.3395	0.1413
Optimum non-linear income taxation					
$(1-\mathrm{MTR}_N)$	$(1-\mathrm{MTR}_S)$	G_N	G_S	Y	^0C
0.7613	1.0151	0.1176	−0.2926	0.4397	0.1663
Optimum linear income taxation					
t	G	Y	^0C		
0.6939	0.1178	0.3395	0.1413		
First-best maxi-min					
G_N	G_S	Y	^0C_A		
0.1095	−0.3284	0.4557	0.1707		

Notes

(i) See table 1.

(ii) $\nu = -1$.

(iii) Parameters not specified at the head of a table are as in table 1.

Table 3

$\delta_S - \delta_N = 0.1$

Optimum lump-sum taxation with errors								
δ_N	δ_S	t	G_N	G_S	w_N	w_S	Y	${}^0C$
0	0.1	0.0533	0.1026	−0.0910	0.3699	0.6334	0.5821	0.2080
0.1	0.2	0.2497	0.1317	−0.0081	0.3572	0.6443	0.5511	0.2032
0.2	0.3	0.3492	0.1354	0.0405	0.3424	0.6579	0.5310	0.1994
0.3	0.4	0.4019	0.1292	0.0737	0.3312	0.6689	0.5188	0.1968
0.4	0.5	0.4242	0.1172	0.0985	0.3251	0.6750	0.5132	0.1955

Notes

(i) See table 1.
(ii) δ_i is the proportion of individuals of type i who are incorrectly classified. In previous tables we had $\delta_S = \delta_N = \delta$.
(iii) $\nu = -1$.
(iv) All other parameters as specified for table 1.

are larger for $\nu = -1$ and -2, with similar levels of δ. The value of δ, $\hat{\delta}$, above which optimum income taxation is preferred to lump-sum taxation with errors, is 0.393 for $\nu = 0.97$, 0.087 for $\nu = -1$ and 0.065 for $\nu = -2$ (interpolated from a fine grid for δ chosen to give values of the maximand close to that for optimum non-linear income taxation for the value of ν under consideration). Note that the curves in figs. 3(a) and (b) are horizontal at $\delta = 0.5$ since, as argued in section 2, we have symmetry of all relevant properties around $\delta = 0.5$.

The above interpretation of ν explains the striking difference in results between $\nu = 1$ on the one hand and $\nu = -1$ on the other. With $\nu = -1$ or -2 it is a matter of great concern if unskilled individuals are subject to a lump-sum tax, i.e. where G_S is negative. Thus, misclassification causes sharp drops in welfare. For the same reason tax rates go up quite quickly with δ since we wish to stop G_S falling too low as some unskilled will be recipients of G_S. In the extreme case of maxi-min ($\nu = -\infty$) one would want $G_S = G_N$ for $\delta > 0$ (if $G_S < G_N$ then a mis-classified unskilled person would be the worse off and one would expect to be able to raise his utility by bringing G_S towards G_N). The graph of ${}^0C$ against δ would be discontinuous at $\delta = 0$: for $\delta = 0$, the maximand takes the value ${}^0C_A = 0.209$ and for $\delta > 0$ the maximand takes the value 0.177, corresponding to optimum linear income taxation (parameter values as for the base run except that $\nu = -50$). The value of ${}^0C$ corresponding to optimum non-linear taxation with maxi-min is 0.200 ($\nu = -50$). Thus, for maxi-min $\hat{\delta} = 0$: if there is the slightest chance of making a mistake we opt for income taxation rather than lump-sum taxation.

It will be seen from fig. 3 that t and ${}^0C$ are, respectively, concave and convex functions of δ for $\nu = -1$ and -2 but that there is an inflexion for

Table 4
Unskilled correctly classified: $\delta_N = 0$

δ_N	δ_S	t	G_N	G_S	w_N	w_S	Y	${}^0C$
				$\nu = -1$				
0	0	0	0.1002	−0.1003	0.3752	0.6289	0.5897	0.2094
0	0.1	0.0533	0.1026	−0.0909	0.3699	0.6334	0.5820	0.2080
0	0.2	0.1035	0.1045	−0.0824	0.3645	0.6379	0.5744	0.2066
0	0.3	0.1539	0.1064	−0.0732	0.3593	0.6425	0.5664	0.2052
0	0.4	0.1994	0.1075	−0.0652	0.3540	0.6473	0.5587	0.2038
0	0.5	0.2455	0.1089	−0.0564	0.3489	0.6519	0.5505	0.2023
				$\omega = -2$				
0	0	0	0.0947	−0.0947	0.3702	0.6331	0.5889	0.2092
0	0.1	0.0665	0.1000	−0.0794	0.3656	0.6370	0.5796	0.2075
0	0.2	0.1289	0.1044	−0.0648	0.3609	0.6411	0.5703	0.2058
0	0.3	0.1833	0.1077	−0.0530	0.3564	0.6451	0.5615	0.2041
0	0.4	0.2348	0.1105	−0.0415	0.3520	0.6490	0.5527	0.2025
0	0.5	0.2880	0.1138	−0.0287	0.3480	0.6527	0.5431	0.2009
				$\nu = 0.97$				
0	0	0	0.2293	−0.2293	0.5088	0.5413	0.6106	0.2151
0	0.1	0.0019	0.2252	−0.2739	0.5038	0.5440	0.6096	0.2149
0	0.2	0.0046	0.2199	−0.3264	0.4975	0.5474	0.6083	0.2146
0	0.3	0.0088	0.2123	−0.3867	0.4884	0.5524	0.6064	0.2143
0	0.4	0.0170	0.2007	−0.4511	0.4744	0.5603	0.6032	0.2138
0	0.5	0.1255	0.1752	−0.3793	0.4338	0.5856	0.5821	0.2124

Notes

(i) Apart from ν, δ_N and δ_S the parameters are the same for table 1.
(ii) Optimum non-linear income taxation, linear income taxation and first-best maxi-min are identical to their respective values for table 1 since they are independent of errors in classification.

$\nu = 0.97$. Note that we have just seen that the extreme case of maxi-min gives ${}^0C$ a horizontal function of δ for $\delta > 0$ with a discontinuity at $\delta = 0$ (and similarly for t). The existence of the inflexion for ν close to 1 may be understood as follows. For very low δ the curve is almost horizontal since the maximand, being insensitive to inequality, changes little if only very small errors are made, and we have seen that symmetry around $\delta = 0.5$ implies that the curve is horizontal there.

We varied ν between 1 and -1 (holding other parameters constant as in the base run) to attempt to discover where the inflexion disappears. It would appear that the critical value of ν is around zero. I have not been able to establish such a result analytically (given that one has to resort to the computer to calculate ${}^0C$ for a specified δ, it is hard to examine analytically the second derivative of ${}^0C$ with respect to δ).

The magnitude of welfare gains from redistributive taxation in the base run (recall $R = 0$) can be seen from table 1(a) and that the level of welfare

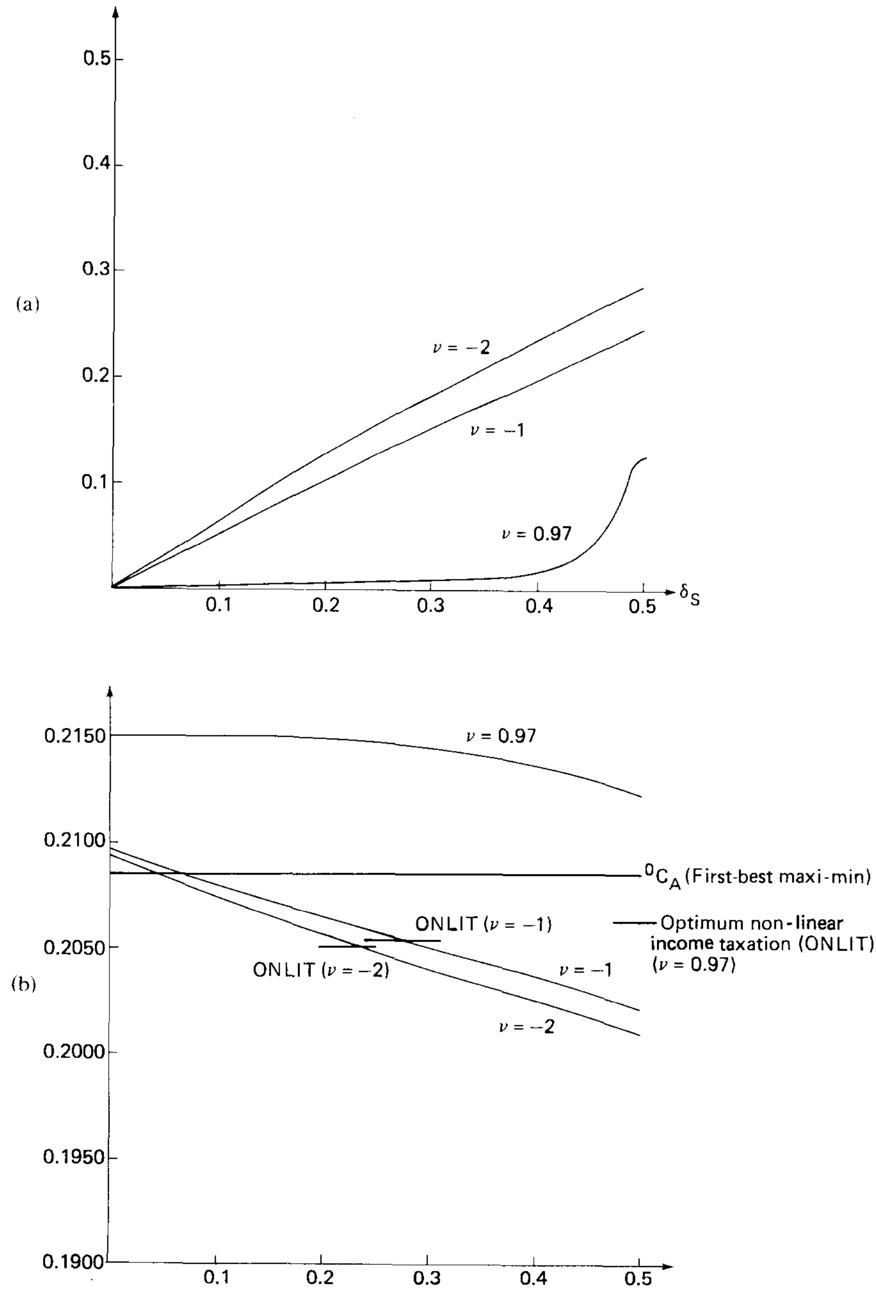

Fig. 4. (a) Optimum tax rate against proportion misclassified (see table 4(a)–(c)): $\delta_N = 0$. (b) Welfare level at optimum proportion misclassified (see table 4(a)–(c)): $\delta_N = 0$.

with zero taxation and lump-sum transfers for the base run is ${}^0C=0.184$: the first-best lump-sum has ${}^0C$ at 0.209, optimum lump-sum taxation with $\delta=0.2$ has ${}^0C$ of 0.201, non-linear income taxation ${}^0C$ of 0.205, and optimum linear taxation ${}^0C$ of 0.195. Hence, moving from no taxation to first-best lump-sum taxation provides a welfare gain of 0.025 in consumption units, or 13.6%. The gain is 0.021 or 11.4% if one is restricted to income taxation and 0.011 or 6.0% if only linear income taxation is possible. In this case there is a 4.0% drop in welfare from the first best if only 80% of the population is correctly classified and lump-sum taxation is used. The rate of change of ${}^0C$ with respect to δ is a measure of the gains to obtaining more precise classification. With an estimate of the costs of extra precision, exogenous to this model, one could make a judgement of whether the effort of finer classification was worthwhile.

For optimum lump-sum taxation with errors the changes in gross wage rates as we move from $\delta=0$ to $\delta=0.5$ are not small. As δ increases the gross relative and absolute wage of the unskilled falls. Note that G_N is not monotonic in δ although G_S is.

Optimum non-linear income taxation involves a small marginal subsidy to the skilled: the sign but not the magnitude was demonstrated in section 3. Again we argued in section 3 that there would be a positive marginal tax rate on the unskilled: this tax rate has turned out to be quite substantial.

Mirrlees (1971) found in his calculations that 'Perhaps the most striking feature of the results is the closeness to linearity of the tax schedules' (p. 206). He warned however (p. 207) 'we have not explored the welfare loss that would arise from restriction to linear schedules'. Here we have found that the welfare differences between optimum linear and non-linear income taxation are substantial (of the order of 5% of consumption). And finally, on the base-run note the very close proximity between the optimum linear income tax and the optimum lump-sum tax with errors where $\delta=0.5$. For this value of δ classification conveys no information and, as argued in section 2, there is no ground for discrimination in lump-sum transfers on the basis of the classification other than a possible local non-concavity in the problem. We suggested that such a non-concavity was not to be expected and this has been confirmed in the calculations.

The comparison of optimum non-linear income taxation with optimum non-linear income taxation under the constraint that the marginal tax rate on the skilled should be equal to zero for the cases concerned here yields results in terms of welfare which are extremely close and thus computations for the latter case are not presented [see Stern (1979)]. This is unsurprising given that the marginal subsidy was found to be small.

We turn now to a discussion of table 2, which shows, for $\nu=-1$, the effects of varying other parameters. For these new values of parameters results for $\nu = -2$ and $\nu = 0.97$ were also computed and the general features described above and illustrated in fig. 3 were not significantly altered. Results for the case of $\nu = -1$ only are reported in table 2.

An increase in the government revenue requirement imposes greater demands on the economy. As a result marginal tax rates for $\delta > 0$ increase and lump-sum grants decline. Labour supplies and output increase and ${}^0C$, the welfare level, declines (note that we have not included in our welfare measure any benefits from the government expenditure). There is similarly an increase in output for both types of income taxation.

An increase in the elasticity of substitution, ε, between consumption and leisure results in a reduction in marginal tax rates for $\delta > 0$ — the dead-weight loss from taxation is larger. A reduction in the competitive (gross-of-tax) share, γ, of the skilled lowers tax rates. The greater similarity between the two types of labour lowers the desire to redistribute through the tax system.

A reduction in the proportion of the skilled in the population raises marginal tax rates and lowers output. The reduction in the number of skilled sharply reduces incomes for the unskilled making redistribution more desirable.

In presenting our results so far we have kept δ_S and δ_N equal. We now examine the effects of allowing $\delta_S > \delta_N$ — thus the government makes more mistakes in classifying the skilled than the unskilled. The motivation for examining this case is that the skilled have an incentive to be classified as unskilled (to obtain a higher grant) whereas the unskilled have an incentive, for tax purposes, to avoid being classified as skilled. At the same time we wished to avoid the complication of the classification proportions being endogenous. The results are presented in table 3 for the case $\nu = -1$ and should be compared with those where δ_S is equal to δ_N presented in table 1(a).

Comparing the first five rows of table 3 with the first five rows of table 1(a) we see that the extra error in classifying the skilled leads to an increase in the marginal tax rate and (for $\delta_N > 0$) a reduction in the grant, G_N. The error in classifying the skilled implies that more grants G_N are distributed with the consequence that the grant is reduced and the tax rate increased. The wage rate of the unskilled falls as a result of the increase in marginal taxation.

We explained in section 2 that the case of $\delta_N = 0$ is of particular interest since it incorporates incentives connected with misclassification in a precise and rational way: the unskilled have an incentive to avoid being misclassified and $\delta_N = 0$ is the assumption that they successfully contest the misclassification. The skilled have no incentive to contest a misclassification. Results corresponding to tables 1(a), (b), and (c) for the case $\delta_S = \delta_N$ are presented in tables 4(a), (b), and (c) i.e. $\nu=-1$, $\nu=-2$, and $\nu=0.97$. Variations in proportions misclassified now refer to δ_S only. Results are illustrated in fig. 4 which is analogous to fig. 3.

Given that all the unskilled are correctly classified one would expect to be able to achieve a higher level of social welfare in table 4 for a given δ_S as compared with a δ of equal magnitude in table 1, and this is indeed the case. Further lump-sum taxation is always strictly superior to optimum linear taxation. Concern over misclassification of the unskilled and consequent very low utility levels now disappears (since it does not arise) and the lump-sum tax on the skilled is higher (again comparing equal values of δ_S and δ across the corresponding tables in table 1 and table 4). Note, however, that the lump-sum subsidy for the unskilled decreases as ν increases, i.e. aversion to inequality decreases.

The critical value of δ_S, $\hat{\delta}_S$, for $\nu=-1$ above which optimum income taxation is preferred to lump-sum taxation is 0.285, which may be compared with $\hat{\delta}$ of 0.087 when $\delta_S=\delta_N$. The corresponding $\hat{\delta}_S$ for $\nu=-2$ is 0.240 compared with $\hat{\delta}=0.065$ when $\delta_S=\delta_N$. For $\nu=0.97$ lump-sum taxation with errors is preferred to income taxation whatever the level of δ_S.

The lesson is clear: if the unskilled are perfectly classified then the lump-sum system becomes much more attractive and the level of errors required before one would switch to income taxation is much higher. For further calculations in and generalisations of this model the reader should consult Carruth (1982).

5. Equity

Our maximand hitherto has been of the Bergson–Samuelson type, and thus takes account of egalitarian values, but does not acknowledge the notion of equity as defined in section 1. Recall that this notion of equity required equal treatment *ex post* of individuals who are *ex ante* identical.

The cost of applying this notion of equity in its absolute form can be quite high. For if $\delta>0$ the above principle of equal treatment would require $G_S=G_N$ in a system of lump-sum grants. For example, with $\delta=0.1$ in the base run, the constraint $G_S=G_N$ loses roughly 5% of ${}^0C$ as compared with the lump-sum optimum with errors.

Income taxation does not violate the notion of equity described above — the income tax schedule is anonymous, in the sense that it is the same for all individuals, and all individuals of the same type make the same choice. Thus, those who regard equity as important would regard comparisons of the type suggested in section 4 as insufficiently favourable toward income taxation as against lump-sum taxation with errors.

The introduction of equity as an absolute notion may be considered too strong and one might be prepared to trade-off violations of that principle against increases in the more usual social welfare function. In this case the isssue arises of how we should formulate that trade-off. Presumably, concern would depend both on the number in a group treated unequally and on the consequences of unequal treatment. For example, if U_N^0 and U_N^1 are the utilities of the correctly and incorrectly classified unskilled, respectively, we may have a term in the maximand representing equity considerations of the form $H(|U_N^0 - U_N^1|, \delta)$. Such calculations should be relatively straightforward but the precise functional form of expressions like $H(\cdot)$ is an open question. Indeed, one can argue that the amalgamation of two different ethical prinicples, such as the utilitarian/Bergsonian and that of horizontal equity, into one grand or 'supra' social welfare function is an unappealing way of meeting a philosophical difficulty. Some, for example, would want to argue that one must make up one's mind whether the principle of horizontal equity should be taken seriously or not. If not, we forget about it, if 'yes' then we impose the constraint $G_S = G_N$. Others might want to take both principles into account and then form a judgement as to the appropriate tax rates and grants without an appeal to a grand social welfare function which combines the two principles.

For those in the latter category we can provide some information to assist their judgement. In table 5 we present the four utility levels (for the base run) for different values of δ. There are four levels since a person in a given skill category can be either correctly or incorrectly classified. Recall that 0 represents correct classification and 1 incorrect. The calibration of utility is in consumption units being the leisurely equivalent consumption (${}^0C$ corresponding to a utility level $\bar{U}$ satisfies $U({}^0C, 0) = \bar{U}$). We see from table 5 that for the lower values of δ the difference in utility levels between correct and incorrect classification are substantial. For $\delta = 0.5$ classification makes no

Table 5

Ex Post inequity in lump-sum taxation with errors

	Base run			
δ	${}^0C_N^0$	${}^0C_N^1$	${}^0C_S^0$	${}^0C_S^1$
0	0.2238	0.1182	0.1968	0.2933
0.1	0.2176	0.1291	0.2001	0.2798
0.2	0.2070	0.1374	0.2058	0.2676
0.3	0.1941	0.1465	0.2136	0.2553
0.4	0.1810	0.1568	0.2226	0.2436
0.5	0.1683	0.1683	0.2326	0.2327

Notes

(i) $\nu = -1$.

(ii) All other parameters as in table 1.

(iii) ${}^0C_i^0$ is the welfare level for skill type i correctly classified as i $(i = S, N)$. ${}^0C_i^1$ is the welfare level for skill type i incorrectly classified as $j \neq i$ $(i, j = S, N)$.

difference to utility levels since the utilitarian/Bergsonian criterion leads us to have $G_S = G_N$ when classification carries no information. Note that for low values of δ the correctly classified unskilled are better off than the correctly classified skilled. This conforms with the Mirrlees result (see section 2 above) that in the first-best ($\delta = 0$) optimum the unskilled will be better off.

6. Concluding remarks

We have, in this paper, been concerned with an issue of considerable importance: the advantages of selective or discriminatory taxation where errors are made in administration as against anonymous, here income, taxation. The modern literature on optimum income and commodity taxation has assumed directly that lump-sum taxation is impossible and Hahn (1973) has complained that we should not assume certain taxes are impossible without giving a reason, and he went on to give examples of lump-sum taxes that have been administered. We have not assumed lump-sum taxes are impossible but we have recognised that we may make errors in administering them; in particular we may not be able to determine the particular features of individuals which we regard as important for deciding lump-sum taxes and we may have to resort to less satisfactory indicators. The lump-sum system under consideration involves no taxation of personal incomes other than a proportional output or income tax at source.

In comparing income taxation and lump-sum taxation with errors we argued in section 2 that provided errors were sufficiently small, lump-sum taxation was to be preferred but that income taxation was more desirable if errors are large. In section 4 we computed the size of errors ($\hat{\delta}$ in our notation) that would make the two types of taxes equally desirable, and found that the value of $\hat{\delta}$ was very sensitive to our distributional values (for $\nu = 0.97$, $\hat{\delta} = 0.393$ and for $\nu = -1$, $\hat{\delta} = 0.087$ — recall $\delta = 0.5$ corresponds to no information from classification). Thus, our predilection for selective rather than anonymous taxation will depend strongly on our estimate of our ability to administer a selective system, and our egalitarian values. Where misclassification of the unskilled is avoided lump-sum taxation becomes more attractive: $\delta_S = 0.285$ for $\nu = -1$ and 0.240 for $\nu = -2$. For $\nu = 0.97$ lump-sum taxation is more attractive than income taxation whatever δ_S.

We have throughout ignored the differences in costs between different kinds of system and maintaining different degrees of accuracy in the selective system. The computations in this paper should be seen as a contribution to the benefit side of the analysis and provide information with which differences of cost can be compared. It should be emphasised that we are comparing two different kinds of system and not the costs and benefits of introducing a lump-sum element into a non-linear tax system.

The issue of horizontal equity arises in the model because errors are made in discriminatory lump-sum taxation. The application of the absolute principle in a system of lump-sum grants forces equal lump-sum grants for all, as linear income taxation. The loss in the social welfare function can be 5 or 10% in consumption units.

It should be emphasised too that although the issues to which the paper is addressed are substantial, the model is exceedingly simple and even though considerable parameter variation has been provided the conclusion must be viewed with some circumspection. One would like to see corresponding computations for different kinds of models. the reader should consult Carruth (1982).

Finally, in the course of our analysis and computation of optimum income taxation for the comparison with lump-sum taxation we found that the optimum income tax schedule in our model with endogenous relative gross wage rates for different kinds of labour had certain features, in striking contrast to those from models where relative gross wages are exogenous. In particular we found that there should be a marginal subsidy on the income of the more skilled individuals, in contrast to preceding models with exogenous relative wages where marginal tax rates should be between zero and one, and zero at the top.

References

Allen, H.F., 1982, Optimal linear income taxation with general equilibrium effects on wages, Journal of Public Economics 17, 135–143 (this issue).

Carr-Hill, R.A. and N.H. Stern, 1976, Theory and estimation in models of crime and its social control and their relations to concepts of social output, in: M.S. Feldstein and R.P. Inman, eds., The economics of the public services (Macmillan). (Proceedings of I.E.A. conference in Turin, April 1974.)

Carruth, A., 1982, On the role of the production and consumption assumptions for optimum taxation, Journal of Public Economics 17, 145–155 (this issue).

Feldstein, M.S., 1973, On the optimal progressivity of the income tax, Journal of Public Economics 2, 357–376.

Guesnerie, R. and J.K. Seade, 1982, Nonlinear pricing in a finite economy, Journal of Public Economics 17, 157–179 (this issue).

Hahn, F.H., 1973, On optimum taxation, Journal of Economic Theory 6, 1, 96–106.

Mirrlees, J.A., 1971, An exploration in the theory of income taxation, Review of Economic Studies 38, 175–208.

Mirrlees, J.A., 1974, Notes on welfare economics, information and uncertainty, in: M. Balch, D. McFadden and S. Wu, eds., Essays on economic behaviour under uncertainty (North-Holland).

Seade, J.K., 1977, On the shape of optimal income tax schedules, Journal of Public Economics 7, 203–236.

Stern, N.H., 1976, On the specification of models of optimum income taxation, Journal of Public Economics 6, 1 and 2, 123–162.

Stern, N.H., 1977, Welfare weights and the elasticity of the marginal valuation of income, in: M.J. Artis and A.R. Nobay, eds., Current economic problems (Blackwell). (Proceedings of the A.U.T.E. conference in Edinburgh, April 1976.)

Stern, N.H., 1979, Optimum taxation with errors in administration, Warwick Economic Research Papers, No. 157 August (University of Warwick).

Stiglitz, J.E., Utilitarianism and horizontal equity: The case for random taxation, Technical Report No. 214 August, The Economics Series, Institute for Mathematical Studies in the Social Sciences, (Stanford University).