

ON THE SPECIFICATION OF MODELS OF OPTIMUM INCOME TAXATION

N.H. STERN*

St. Catherine's College, Oxford, England

with programming by D. Deans

Revised version received September 1975

The main concerns of the paper are the problems of estimating labour supply functions for use in models of optimum income taxation, and the calculation of the effect on the optimum linear tax rate of varying the elasticity of substitution, ϵ , between leisure and goods from 0 to 1. Backward sloping supply curves are commonly observed and they imply $\epsilon < 1$. Our calculation of ϵ from estimates of supply curves by Ashenfelter and Heckman gives $\epsilon = 0.4$. Optimum marginal rates decrease with ϵ when taxation is purely redistributive but may be nonmonotonic if positive revenue is to be raised. It is proved that optimum (linear or nonlinear) taxation involves a marginal rate of 100 percent when $\epsilon = 0$.

1. Introduction

There are four main ingredients for a model of optimum income taxation: an objective function, a preference relation or supply function for individuals, a skill structure and distribution, and a production relation. They are closely intertwined. An individualistic social welfare function would take into account the preference structure of individuals. The supply of various kinds of skills will depend on individuals' wishes or ability to produce these skills. The production relation must state how skills of different kinds are combined to produce outputs.

The optimum income taxation problem as usually posed is to maximise a social welfare function, which depends on individual utilities, subject to two constraints. The first is that each individual should consume goods and supply factors in amounts which maximise his utility subject to the constraint of the tax function, which describes how much post-tax consumption can be acquired from pre-tax earnings. We are searching for the optimum function. The second is that the total labour supplied can produce the total quantity of goods

*This paper has benefitted greatly from discussions with A.B. Atkinson, D.L. Bevan, P.A. Diamond, J.S. Flemming, J.A. Mirrlees and K.W.S. Roberts. The comments of participants at a seminar in Cambridge were also helpful. Responsibility for all errors is mine.

The paper was presented to the ISPE Conference on taxation in Paris, January 18–20, 1975. The comments of the discussants at that conference, E. Malinvaud and M. Bruno, were helpful. The support of the SSRC under grant HR 3733 is gratefully acknowledged.

demanded. It is the former constraint which characterises the optimum income taxation problem and which makes it a problem of the second best. Without this constraint, that individuals are on their supply curves, we have a first-best problem.

When taxation is discussed it is often in terms of a trade-off between equality and efficiency, or the distribution of the cake and its size. The optimum income taxation problem is one way of formalising this trade-off and it is, perhaps, surprising that it was not until Mirrlees (1971) that a suitable model was developed. We are still at the stage of understanding the structure of these models and the importance of the various components. It should be clear at the outset that the purpose of this paper is not to make recommendations to the Treasury as to appropriate tax rates, but to contribute to the understanding of the discussion of equality versus efficiency through examination of a particular model.

The particular concern of this paper is the supply function, and attention is focussed on the special case of labour supply. We shall examine the problem of estimation, which preference structures obtain support from the empirical literature on labour supply, and then the influence such estimates should have on our view of the appropriate level of income taxation. It will be suggested that most previous calculations of optimum tax rates may have been biased low.

The next section presents the models of Mirrlees (1971) and Atkinson (1972) and contains a brief discussion of their numerical results.¹ The problems of specifying and estimating skill distributions are discussed in section 3, together with calculations of the elasticity of substitution (ε) between leisure and goods, based on empirical estimates of labour supply functions.

The calculations of section 3 suggest that elasticities of substitution around $\frac{1}{2}$ are of interest, and in section 4 the optimum linear income tax, for values of ε between 0 and 1, is calculated in a model similar to that of Mirrlees (1971). The extreme case of $\varepsilon = 0$ is examined, in the Mirrlees model, in section 5 and we find the optimum income taxation (linear or nonlinear) involves marginal taxation at 100 percent. It is not surprising, therefore, that the calculations of section 4 show that, for small ε , the optimum linear tax rate increases to 100 percent as ε decreases to zero. However, where taxation is imposed to raise revenue, as well as to redistribute, the optimum marginal rate may increase as ε increases over a certain range. In section 6 the numerical discussion is evaluated.

The remainder of this section is devoted to a brief examination of those elements of the model, the objective function and the production relation, which receive no further attention in the later discussion.

Most previous writers have worked with a concave transformation of individual cardinal utilities. The transformation ranges from the linear utilitarian sum to the case where the 'degree of concavity' goes to infinity – the maximin,

¹I originally intended to do a survey of theoretical and empirical work in progress but became more involved with my own investigations.

or Rawlsian, solution. Some might wish to claim that one is merely specifying the value judgements of the decision-maker by using an arbitrary numbering of individual indifference curves together with a method by which individual utilities are aggregated. The specification of a particular cardinal numbering for individuals and the form of the social preference relation over utilities may well be difficult, if not impossible, to disentangle, but I find it hard to understand a quantitative comparison between different forms of social welfare function for the same indifference structure (for individuals) if some benchmark of cardinality is not involved.

The cardinality problem is much less severe when a one argument utility function is used – see, for example, Atkinson (1973a). One can then suppose that the government defines its values over the vectors whose components are household incomes. However, when supply functions are central to the model a one-argument utility function seems out of place. It then becomes more difficult to wriggle out of the problem of numbering individual indifference curves. It is possible that part of the attraction of maximin objective functions is that the cardinality problem is less troublesome – maximising the lowest utility level will give the same policy whichever cardinalisation is used when the same monotonic increasing transformation of utilities is applied to all individuals.

The above discussion and most of the literature has supposed that the Bergson–Samuelson social welfare function (nondecreasing in each argument) is the appropriate tool for capturing social values in such analyses. Leaving aside the question of whether it *should* be used, it is possible that many people have some different underlying notion of welfare or distributional justice when they discuss income taxation. We illustrate the possible phenomenon with a few quotations and arguments which might be thought plausible and yet imply *non-Paretian* objectives. We begin with three quotations on inequality each of which clearly involves a non-Paretian position.

Tawney:²

When the press assails them with the sparkling epigram that they desire not merely to make the poor richer but to make the rich poorer, instead of replying, as they should, that, being sensible men, they desire both, since the extremes of both of riches and poverty are degrading and anti-social, they are apt to take refuge in gestures of depreciation.

Simons:³

The case for drastic progression in taxation must be rested on the case against inequality – on the ethical or aesthetic judgement that the prevailing distribu-

²See Atkinson (1973b, p. 19). I am grateful to Kevin Roberts for drawing my attention to this quote.

³See Simons (1938, p. 15). Kevin Roberts drew my attention to this quote too.

tion of wealth and income reveals a degree (and/or kind) of inequality which is distinctly evil or unlovely.

Fair (1971) quotes Plato as follows:

Plato felt that no one in a society should be more than four times richer than the poorest member of society for 'in a society which is to be immune from the most fatal disorders which might more properly be called distraction than faction, there must be no place for penury in any section of the population, nor yet for opulence, as both breed either consequence.'

Certain arguments on tax proposals and structures might seem plausible to many and also involve non-Paretian judgements. For example, Sadka (1973) has shown that with a finite number of individuals or skill levels the optimum marginal tax rate at the very top is zero. One can express his argument verbally as follows. Suppose that a given tax structure is a candidate for the optimum and it results in the most skilled person earning $\text{£}Y$. Consider the announced marginal tax on the $(Y+1)$ pound and suppose it is positive. Reduce it to zero. The most skilled person may work more and if he does he is better off. Similarly, others of lower skill may also work more. If they do, then they are better off (exploiting opportunities that were not available to them before) and they pay more tax since they move through tax brackets with nonnegative marginal rates. Thus, our change has produced more tax revenue and has made everyone at least as well off as before.⁴ A Paretian should approve. Many, however, might regard a zero marginal rate at the top as offensive. It is conceivable that they may wish to retain this view even after they have understood the above argument. We should note that one cannot deduce that, where the skill distribution has positive density, for all positive skill levels the optimum marginal tax rate tends to zero. Indeed, Mirrlees (1971) gives examples where it does not. The structure of the model is similar to an optimum growth model where we cannot infer from the result that a finite horizon model should have zero capital stock at the end, the conclusion that the capital stock tends to zero on the infinite horizon path.

Some might propose a 100 percent tax on inheritance on the grounds of equality of opportunity for children. It is non-Paretian (if one rules out envy as the basis of the argument), since the ability to confer the inheritance makes the parent better off (the desire is to give rather than consume) and, presumably, the offspring as well.⁵

Many have found⁶ the 'equal absolute sacrifice' proposal an attractive basis for optimum income taxation. This abstracts from incentive problems and states that to raise a given revenue everyone should give up that amount of his income

⁴One can throw away the extra tax revenue if it is so desired. The argument is clearly rather general. 'Better off' has been used here in the weak sense of 'at least as well off.'

⁵Mirrlees drew my attention to this argument.

⁶This principle is discussed and fitted to U.K. tax schedules in Stern (1973).

which makes the sacrifice of utility equal. It turns out that one can choose a utility function which fits the U.K. income tax structure rather well.⁷ Although it does not violate the Paretian condition, the proposal cannot be based on any symmetric strictly concave Bergson–Samuelson welfare function since, abstracting from incentive effects, such welfare functions lead to equal post-tax incomes.

The above examples indicate that the standard welfare economics procedure based on the usual welfare functions would not be regarded as the obvious starting point by many who might be prepared to comment on income tax structures.

Most of this paper will use a production structure with one basic input – labour in efficiency units – with a fixed wage. This does not mean that we are assuming constant returns to scale. We can regard the wage as the marginal product at the level of optimum total production and any profits that accrue as lump sum income for the government. Nevertheless, the assumption of one basic input is worrying. It is often asserted that a particular skill is lacking (say, management in the U.K.) and this carries with it a strong notion of complementarity with other factors rather than the complete substitutability assumed in the case of labour in efficiency units. Feldstein⁸ has made a start in this direction and incorporates two different kinds of labour into his model.

The frequent assumption of public ownership seems less serious. If, for example, there are profits in the system, one can carry out an analysis of the optimum levels (presumably subject to some constraints). The constraints on income taxation would then take account of the presence of these other taxes. Further work is necessary, however, and Atkinson and Stiglitz (1976) have begun an examination of appropriate combinations of various taxes.

The absence of further discussion of the production assumptions should not be taken as a belief that they do not matter. The specification of the way different skills interact in the production process embodies an aspect of income taxation that many would regard as crucial. It is an important area for further research.

The models discussed here will all be static and will not, therefore, involve capital and the elasticity of its supply in any essential way. These models allow the discussion of the important questions of labour supply and raise sufficient significant and difficult questions to warrant study. Some progress has been made with dynamics but the components of the models have to be kept rather simple.⁹

2. The model and numerical results of the studies of Mirrlees and Atkinson

This discussion is not intended as a comprehensive survey since Atkinson (1973a) has recently provided a thorough discussion of previous numerical

⁷See Stern (1973).

⁸Feldstein (1973). The different types of labour combine through a Cobb–Douglas production function to produce output.

⁹See, e.g., Feldstein (1973).

work. The main purpose of this section is to draw attention to the levels of calculated optimum marginal tax rates, in models similar to those of section 4 which are, on the whole, lower than one might have predicted. Indeed Mirrlees (1971, p. 207) remarked '... I must confess that I had expected the rigorous analysis of income-taxation in the utilitarian manner to provide arguments for high tax rates. It has not done so.'¹⁰ A partial response to these results has been the use of strongly egalitarian ('highly concave') social welfare functions and the limiting case the 'maxi-min welfare function.'¹¹ We shall suggest later that there is no need to use these more extreme social welfare functions to obtain tax rates that seem closer to observed rates, and that one has merely to use labour supply functions which seem closer to those which are usually estimated. For the moment, however, we give a brief sketch of these earlier results and, in the process, set out the model of income taxation to be used later.

The original work on the current models of income taxation was that of Mirrlees (1971). In his model individuals supply labour of different qualities and hence face different pre-tax wage rates. They choose how much to supply by maximising $u(c, l)$ subject to $c = g(nlw)$, where c is consumption, l the hours worked, nw the hourly wage of an n -man – he produces n efficiency hours per hour worked – w is the wage per efficiency hour and $g(\cdot)$ the tax function giving post-tax income as a function of pre-tax income.

The aggregate production constraint is $X = \int cf(n) dn = H(\int nlf(n) dn) = H(Z)$, where X (total consumption) is a function $H(Z)$ of effective labour Z , and $f(n)$ is the density of the distribution of individuals. The problem is to vary $g(\cdot)$ to maximise $\int G(u)f dn$, where $G(\cdot)$ is a concave function and the constraints are that the amounts individuals choose to supply of labour and consume of goods be compatible with the production relation. Note that the formulation involves taxation of nlw and does not require (nw) and l to be separately observable. If one can identify an n -man without affecting his behaviour, then the first-best optimum can be achieved by levying an appropriate lump sum tax for each n with a zero marginal rate of taxation.

Mirrlees provided detailed calculations for the cases where $u(c, l) = \log c + \log(1-l)$, n distributed lognormally (parameters of the associated normal distribution being $\bar{\mu}$ and σ), H linear and $G(u) = u$ or $-e^{-u}$. Using a value of $\sigma = 0.39$, derived from the work of Lydall,¹² *on the distribution of earnings*, he obtained median marginal tax rates for the case of $G(u) = u$ of 22% and 20%.¹³ The higher rate was for the case where 7% of product was required by the government and the lower where 17% could be added – the additions or subtractions corresponding respectively to cases where profits or revenues

¹⁰The utilitarian optimum ignoring incentives involves 100 percent taxation. One is initially surprised therefore when the introduction of incentives drops the rate down to 20 percent.

¹¹See Atkinson (1972).

¹²See Lydall (1968) and Mirrlees (1971).

¹³Interpolated from Mirrlees (1971, tables I-IV, p. 202).

elsewhere outweighed or were outweighed by fixed costs, or necessary expenditure. With a net government expenditure of 12% of product and $G(u) = -e^{-u}$, the median marginal rate rises to 33%.

The highest marginal rates for the three cases respectively are 26%, 21% and 39%. The marginal rates rise at first but begin falling before the median is reached. Mirrlees proves that, for the log-normal distribution and where the elasticity of substitution between consumption and leisure is less than one, the marginal rate tends to zero as n tends to ∞ . There is a higher limit in the case of the Pareto distribution where, with the same condition on the substitution elasticity, the marginal rate tends to $1/(1+\gamma)$ as $n \rightarrow \infty$ when $(nf'/f) \rightarrow -(\gamma+2)$. Examination of distributions of earnings (see section 3.2) suggests values of γ from 0.5 to 2.5 giving limiting marginal rates from 67% to 29%.

Higher rates can also be produced by widening the distribution of skills – if σ in the log normal case is increased to 1.0 (from 0.39), the median rate is 56% for the case $G(u) = -e^{-u}$ and a government requirement of 7% of product. Presumably with a wider distribution of skills, inequality considerations increase relative to those concerned with incentives. However, Mirrlees (1971, p. 207) suggests that such a σ ‘does not seem to be at all realistic . . .’ since it gives a dispersion of skills too wide to be compatible with observed distributions of earnings.¹⁴

There are two main features of the calculated tax schedules which look different from actual income tax structures.¹⁵ Marginal rates are not monotonically increasing – most of the population is in the region where they are falling – and the highest marginal rates are low. For the ‘realistic’ case of $\sigma = 0.39$, applying to 5 out of 6 of Mirrlees’ examples, the highest marginal rate is 39%.

Atkinson (1972) and (1973a) discusses the effect of increasing the concavity of $G(\cdot)$ and the limiting case of maximin. It seems clear that he was in part influenced by the low rates in the Mirrlees calculations – see Atkinson (1972, p. 2) and (1973a, pp. 390–391). The maximin criterion in the Mirrlees model yields tax rates around 50% for the median person [see Atkinson (1972, p. 28)].

We have already given the Sadka argument which explains why, for a finite population, we should expect zero marginal tax rates at the top of the distribution. This argument may also have some intuitive force for distributions with an infinite domain, provided the weight in the tail is not too big. We have noted, for example, that the log-normal gives a limiting marginal rate of zero but the Pareto does not. The zero limit of the marginal rate for certain distributions suggests that a declining rate at the upper end may be a feature of many models of optimum income taxation. We shall say no more (except for the special case of section 5) about the shape of the tax function, and concentrate on the labour supply function and its relation to optimum linear taxation.

¹⁴We discuss in section 3.2 whether the distribution of earnings gives a misleading impression of the distribution of skills.

¹⁵These are announced rates rather than effective rates.

3. The estimation of supply functions and skill distributions

3.1. Supply functions¹⁶

The work on optimum income taxation has dealt exclusively with situations where individuals have the same preference relation but differ in their earnings capacity. One can also imagine cases where individuals differ in their preference relations but face the same earnings function which is determined, as far as they are concerned, exogenously. In this subsection we shall be discussing such alternative specifications, and the different problems they pose for estimation.

We shall suppose, for the moment (but see section 3.2) that the number of hours of work is the appropriate argument of an individual's utility function and that the pre-tax wage measures the skill or efficiency of a worker per hour of work. For estimation (but not taxation) purposes we suppose that the wage and hours are separately observable.

To make some of our formulae explicit we shall consider utility functions of the constant elasticity of substitution (CES) form, although it is clear that many of the problems we shall discuss do not depend on the particular form of the utility function.

We suppose an individual maximises

$$u(c, l) = [(1 - \alpha) c^{-\mu} + \alpha(h(L - l))^{-\mu}]^{-1/\mu}, \quad (1)$$

subject to the budget constraint

$$c = A + (nw)l. \quad (2)$$

We thus have a linear tax schedule. The individual is characterised by the triple (h, n, L) and one could consider a distribution of this triple over the population. We shall be discussing some special cases. We should think of L as the number of hours available to the individual for allocation between work and leisure, given his family commitments, sleeping requirements, physical attributes and so on. The parameter h measures the ability to enjoy leisure and n the ability to produce efficiency hours of work from clock hours. Different specifications of the relations between h, n and L may lead to very different interpretations of data on wages and hours.

The first-order condition for maximisation of utility subject to the budget constraint is

$$\frac{(A + (nw)l)}{h^{1-\varepsilon}(L-l)} = \left[nw \cdot \frac{(1-\alpha)}{\alpha} \right]^\varepsilon, \quad (3)$$

where $\varepsilon = 1/(1 + \mu)$.

¹⁶The comments of A.B. Atkinson on this subsection were particularly useful.

In the Mirrlees case individuals have identical preferences so that h and L are constant over the population. Putting $h = 1$ and taking logarithms, we have

$$\log \left(\frac{A + (nw)l}{L - l} \right) = \varepsilon \log (nw) + \varepsilon \log \left(\frac{1 - \alpha}{\alpha} \right). \tag{4}$$

We see immediately that where the total quantity of hours available (L) is known or specified, we can estimate ε and α by regressing consumption per hour of leisure on the wage rate (nw).

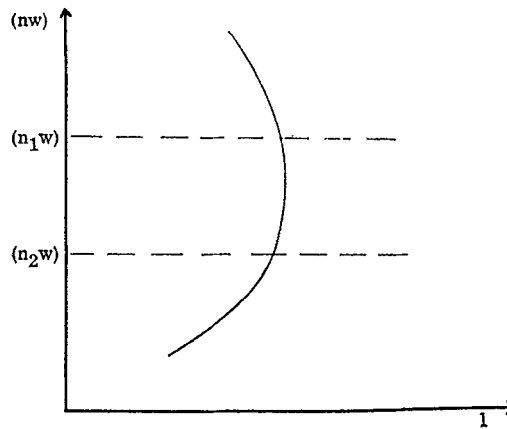


Fig. 1

Note that our assumption of identical preferences enables us to identify the supply function by merely plotting the relation between l and the post-tax wage rate per clock hour (nw) (see fig. 1). This formulation is, therefore, especially convenient for estimation purposes (see section 3.3). The skill distribution is then given by the distribution of wage rates.

The above procedure is very sensitive to the assumption of identical preferences. We give two examples to illustrate this point. First, suppose L is constant in the population but $h = n$. In other words, individuals have identical available hours but those who produce more efficiency hours of work obtain a similarly increased satisfaction per hour of leisure. And suppose, for the sake of illustration, that $A = 0$. We see from (3) that l is independent of n . In other words, everyone works the same number of hours. Thus we might infer, on seeing a distribution of wages and no variation of hours, that the supply curve was inelastic when an increase in w (the wage per efficiency hour) would change hours worked.

A second example has been used by Hall (1974). He supposes $h = n = 1$ but L varies in the population. He deals in particular with the case where $\varepsilon = 1$ (equivalent to $\mu = 0$ or $u(c, l) = c^{1-\alpha}(L-l)^\gamma$) so that we have $(L-l) = \alpha(A + wL)/w$. He assumes $L = (1-\theta)\bar{L}$, where θ has the beta density on $[0, 1]$: $f(\theta) = 6\theta(1-\theta)$. He applies the model to the Penn-New Jersey negative income tax (NIT) experiment. Families were offered a choice between (A_0, w_0) (participation) and (A, w) (nonparticipation) with $A_0 > A$, $w_0 < w$. His model predicts both participation rates and changes in hours given participation fairly well. Hall argues that the representative individual is not a sensible concept when we see a dispersion of hours worked for a given (A, w) , and that a theory of labour supply should account for this dispersion.

3.2. *Some problems of estimating the skill distribution*

In the previous subsection we suppose for our discussion of estimation that nw and l were separately observable. We had been interpreting l as clock-hours and regarding l as the relevant argument for the disutility of labour and as the basis of the productivity measure. The problem is more complicated than this, however. Both disutility and productivity of labour may be a function primarily of the effort required rather than the number of hours, although the latter is obviously of importance. In the absence of a direct measurement of effort we should discuss estimation problems when we can observe nwl , total pre-tax labour income, and not nw and l separately. Here we interpret l as effort.

There is one special formulation¹⁷ which makes the problem disappear. If individuals maximise $(1-\alpha) \log c + \alpha \log (1-l)$ subject to $c = a(nwl)^\delta$, where a and δ define the tax function, then l is constant and (pre-tax) incomes are distributed as a constant times n . We can, therefore, read off the distribution of skills from the distribution of labour income. Since l is not directly observable, the assumption that it is constant is not violated, although we cannot estimate α . It is clear, however, that the trick is rather special and will not work for more general utility and tax functions.

In general then, if l is not directly observable, we cannot pass from a distribution of labour income to a distribution of n unless we have full knowledge of the utility function and the tax function, when l can be deduced. We can, however, gain information on the utility function and skill distribution separately if the tax schedule *changes*. We can illustrate this as follows. Put $\varepsilon = 1$ in (3), and we have

$$(nwl) = (1-\alpha)nwL - \alpha A. \quad (5)$$

We can now use (5) to estimate α . Let us suppose that the current post-tax wage

¹⁷The formation was used by Vickrey (1947) and Bevan (1974).

per effective hour is one (where the model presumes a linear tax schedule). This is merely choosing a linear scale for n . A tax change occurs which increases A to A_0 and decreases w from 1 to w_0 – as in the Penn–New Jersey NIT experiment. We observe only nwl and A in both cases but we know that nw has decreased from n to nw_0 , where w_0 is known – since $1 - w_0$ is the increase in the marginal tax rate. For a given individual we then have two equations in two unknowns (α, n) which we can solve for α and n . Given a population subject to the experiment we could find the distribution of n (and of α) in the population.

The kind of experimental information we have just been discussing is rather rare.¹⁸ Usually we have a given tax structure, a distribution of labour income and we may be uneasy about measuring effort by hours worked. We should like to know if the income distribution is a good proxy for the skill distribution. We saw at the beginning of this subsection a special case where the distributions were identical. This case is unusual, however, and the income distribution may be very misleading as an estimate of the distribution of skills. For example, one can imagine a utility function where an individual has a target level of consumption or income (c_0) upon which he insists, but he is not prepared to work to raise his consumption beyond this level,

$$u(c, l) = \begin{cases} 1-l & \text{if } c \geq c_0, \\ -\infty & \text{otherwise.} \end{cases}$$

If we have a community of individuals of different skills, all of whom have this utility function, we should observe a completely equal distribution of incomes. However, some individuals would have to exert a great deal of (unobservable) effort to achieve c_0 and consequently would have low utility. Others would achieve c_0 with comparative ease.

This is an extreme example but it illustrates the point that where the unobserved supply curve of effort l with respect to nw is backward sloping, the distribution of skills is more unequal than the distribution of incomes. We shall see in section 3.3 that supply curves of hours are usually found to be backward sloping for much of their range.

On the other hand there may be many factors in actual situations which affect wage rates but should not be described as innate ability: for example, age, education, luck or power. In ideal circumstances one would examine a population which had constant values of these complicating factors. While this may be possible for age or education it is difficult in the case of luck or power. Note that if the acquisition of education is sensitive to earnings, education should be included in supply functions and not 'skill' distributions. We return to this briefly below.

¹⁸Hall (1974) makes use of the experimental Penn–New Jersey NIT data for his actual estimation (see section 3.1).

Suppose the wage rate m is equal to an (additive) combination of innate ability n and some other factor x . Then

$$\text{var}(m) = \text{var}(n) + \text{var}(x) + 2 \text{cov}(n, x).$$

If the covariance is zero or positive the distribution of wage rates is more unequal than the distribution of abilities. It seems more likely that the factors mentioned are positively correlated with ability.

We have seen that the relevant evidence for skill distribution must, therefore, be based on labour earnings and, where possible, rates, and be corrected for age and education. It is clear that a casual examination of the distribution of (earned plus unearned) income is insufficient. This is an important area for further research.

One of the main problems for this research will be the specification of the functional form of the distribution to be fitted. Pareto (1897) found that $N_x = JX^{-\alpha}$, where N_x is the number of incomes above X , gave a remarkably good fit for several countries. He estimated α and found that it was around 1.5. On the other hand, Lydall examines the upper tail of the distribution of employment incomes for different countries and finds α ranging from 2.27 for France in 1964 to 3.4 for Germany in 1964 (1968, p. 133). Lydall does examine employment incomes and, in some cases (1968, p. 33) tries to work with populations with given numbers of hours per week. He suggests further that, for precisely defined occupational characteristics (1968, p. 33), the log-normal distribution fits rather well.

An interesting approach to the problem is the recent study by Schwartz (1975). He finds, disaggregating populations by race and years of education, that the power transformation of income which gives the closest approximation to normality is the cube root of income.

We can come to no firm conclusions as to whether the current distribution of income gives an accurate picture of the distribution of skills. We saw that there were two powerful influences, backward bending supply curves and non-skill factors, pulling in opposite directions. It must be emphasised that the non-skill factors include a multitude of variables which depend on the institutions and organisation of society, and that the relative productivity of different skills depends on the capital stock. Further it is clear that the one-dimensional model of the skill distribution is a very crude representation of reality. But the problem is deeper than this. If skills are acquired, the motivation may be the potential reward, as for example in human capital models. We should then include acquired skills in the supply function rather than the skill distribution. This forces us to think of n as innate ability, a notion which is both slippery and controversial. And, what if skills (and effort) are not acquired (or supplied) for monetary reward? We have to reexamine our concepts of supply. Theoretical and empirical research on these problems is still in its infancy.

3.3. Supply curves as estimated

Empirical estimates of the response of labour supply to changes in wages and income are usually expressed in terms of a supply function. For models of optimum income taxation we usually wish to work with explicit utility functions. The purpose of this section is to describe the calculation of the parameters of a CES utility function from the estimates of income and wage responses which have been found by others.

We showed in section 3.1 that for the Mirrlees case we can estimate the labour supply function directly by assuming that everyone has the same supply function and differences in skills result in differences in wage rates (see fig. 1). We suppose that such an estimation has been performed and we have estimates of the (uncompensated) wage and income elasticities at some level of wages and for some lump-sum incomes. We want to infer a CES utility function.

We suppose the individual problem is to maximise $[(1 - \alpha)c^{-\mu} + \alpha(L - l)^{-\mu}]^{-1/\mu}$, where c is consumption of goods, l is labour supply and L the maximum possible level of work. We assume that consumption is $c = A + wl$. We are, therefore, assuming a linear tax schedule where w is the post-tax wage. We are not concerned here with the reason for the level of w so we suppress the 'n' factor.

The first-order condition for the above problem is obtained by putting $h = n = 1$ in eq. (3); we then have

$$\left[\frac{L - l}{A + wl} \right]^{\mu + 1} = \frac{\alpha}{(1 - \alpha)w} \tag{6}$$

It is obvious from (6) that

$$\frac{-\partial \log \left[\frac{L - l}{A + wl} \right]}{\partial \log w} = \frac{1}{1 + \mu} = \varepsilon,$$

the elasticity of substitution. We differentiate eq. (6) logarithmically with respect to w and A in turn, and after a little manipulation obtain

$$\frac{\partial l}{\partial w} = \frac{(A - \mu wl)(L - l)}{w(\mu + 1)(A + wL)} \tag{7}$$

$$\frac{\partial l}{\partial A} = -\frac{(L - l)}{A + wL} \tag{8}$$

Given $w, l, (w/l)(\partial l/\partial w), (A/l)(\partial l/\partial A)$, we can solve (6), (7) and (8) for L, α, μ .

Ashenfelter and Heckman (1973) estimate income and substitution effects from a cross-section of 3,203 male heads of families from the national probability sample component of the 1967 U.S. Survey of Economic Opportunity. They restricted their sample to men not receiving welfare payments and whose wives were present but not working. They write (my notation),

$$\Delta l = S\Delta w + B[l^*\Delta w + \Delta A]. \quad (9)$$

Δ represents differences from sample means, S is the substitution term, and l^* is the average of the mean labour supply of the sample and l , so that $l^*\Delta w$ represents an approximation to the income compensation and thus B an approximation to $\partial l/\partial A$. Eq. (9) is then estimated.¹⁹ Hours were calculated using annual earnings divided by hourly wage rates. Dummy variables for race, region and size of town were included as well as age and age squared. The age terms give an increase in hours to age 44 and a decline thereafter.

They find, for the mean of their sample, that $w = 3.86$ dollars per hour, $l = 2272$ hours per year, $A = 800$ dollars per year,²⁰ $(w/l)(\partial l/\partial w) = -0.15$ and $\partial l/\partial A = -0.07$. These numbers give values of L , α and μ of 3190, 0.994 and 1.45 (to 3 significant figures), respectively. Note that the value of α depends on the units of measurement of labour and income. The value of the elasticity of substitution, $\varepsilon = 1/(1 + \mu)$, is 0.408.

This is a rather striking result since the income tax models discussed in section 2 concentrated attention on the addilog case where $\mu = 0$ and $\varepsilon = 1$. We discuss the qualifications which must be attached to this estimate at the end of this subsection. For the moment, we examine its sensitivity to the values of A , $(w/l)(\partial l/\partial w)$ and $\partial l/\partial A$ – presumably the wage and hours of work at the mean of the sample can be taken as given.

It is rather hard to measure the lump sum income A available to an individual. One has to make many judgements as to how to treat social security benefits,²¹ returns on durable assets and so on. We therefore allowed A to vary across a large range, \$0–2000. The results are shown in table 1. The estimates are rather insensitive to changes in A . For $A = 0$, we obtain $\varepsilon = 0.444$; and for $A = 2000$, $\varepsilon = 0.362$.

Ashenfelter and Heckman compare the figure of -0.15 for $(w/l)(\partial l/\partial w)$ with ‘Sherwin Rosen’s (1969) estimates of -0.07 to -0.30 from inter-industrial data, T. Aldrich Finegan’s (1962) estimates of -0.25 to -0.35 from inter-occupational data, Gordon Winston’s (1966) estimates of -0.07 to -0.10 from inter-country data, and John Owen’s (1971) estimates of -0.11 to -0.24 from U.S.

¹⁹Instrumental variable techniques were used since l^* is correlated with the disturbance term [see Ashenfelter and Heckman (1973)]. The income compensation term should really allow for any differences between marginal and average tax rates.

²⁰I am grateful to Professor Ashenfelter for supplying me with this estimate of A .

²¹In fact, workers receiving social security benefits were excluded from the sample.

time-series data.' The Ashenfelter–Heckman figure of -0.15 is calculated from the sum of a substitution effect and an income effect. The former is estimated at 0.12 and the latter $[3.86 \times (-0.07)]$ at -0.27 . The standard error of the substitution coefficient is 26.0% of its estimated value and of the income coefficient 13.4% of its estimated value.²²

Given this breakdown of the -0.15 estimate, the range of the other estimates and the standard errors, we examine the sensitivity of the ϵ estimate by using 4 ways of changing $(w/l)(\partial l/\partial w)$ so that it ranges over -0.05 to -0.30 . In table 2, col. (a), we vary $(w/l)(\partial l/\partial w)$ holding the income term constant at -0.27 . All the adjustment occurs in the substitution term and ϵ decreases to 0.0685

Table 1
Central estimates of ϵ .^a

<i>A</i> (Dollars per year)	<i>L</i> (Hours per year)	$1 - \alpha$	ϵ
0	3113	0.9903	0.4449
400	3152	0.9926	0.4255
800	3190	0.9944	0.4077
1200	3228	0.9957	0.3913
1600	3267	0.9967	0.3762
2000	3305	0.9975	0.3622

^a $(w/l)(\partial l/\partial w) = -0.15$, $\partial l/\partial A = -0.07$, $w = \$3.86$, $l = 2272$ hours, as in Ashenfelter–Heckman (1973).

for $(w/l)(\partial l/\partial w) = -0.25$ (and $A = 800$). A value of $(w/l)(\partial l/\partial w)$ below -0.27 would, of course, give negative ϵ and is not, therefore, entered in the table.

In table 2, col. (b), we vary $(w/l)(\partial l/\partial w)$ holding the substitution term constant. In col. (c) we vary $(w/l)(\partial l/\partial w)$ by changing the income and substitution terms in the same proportion. For $(w/l)(\partial l/\partial w) = -0.30$, for example, the income term contributes -0.54 and the substitution term $+0.20$. Finally, in col. (d), we vary $(w/l)(\partial l/\partial w)$ so that the modulus of the substitution and income terms moves in the same direction by equal proportions. The sensitivity was analysed in terms of the income and substitution terms since these are the coefficients estimated by Ashenfelter and Heckman and are the natural parameters for an analysis based on utility. Different methods of variation are used since we have two parameters and thus must consider errors scattered on a plane. Movements along the axes of this plane are represented in table 2, cols. (a) and (b).

The row of table 2 corresponding to -0.15 replicates the central estimates since there is no change in income or substitution effects. Column (c) gives a

²²See Ashenfelter and Heckman (1973, table 7.1, line 4).

constant value of ε – it is clear from dividing eqs. (7) and (8) that given A , w and l , μ (and hence ε) depends only on

$$\left(\frac{w}{l} \frac{\partial l}{\partial w}\right) / \left(\frac{\partial l}{\partial A}\right).$$

It appears that our ε estimate of 0.408 is, if anything, a little above the ‘average’ value one would obtain using the estimates of the authors cited by Ashenfelter and Heckman.

Table 2
Sensitivity of ε estimate to assumptions on income and substitution effects.^a

$\frac{w}{l} \frac{\partial l}{\partial w}$	(a) ^b		(b) ^c		(c) ^d		(d) ^e	
	ε	L	ε	L	ε	L	ε	L
-0.05	0.7468	3190	0.6471	2780	0.4077	2517	0.6884	2895
-0.10	0.5772	3190	0.5002	2972	0.4077	2816	0.5274	3036
-0.15	0.4077	3190	0.4077	3190	0.4077	3190	0.4077	3190
-0.20	0.2381	3190	0.3441	3439	0.4077	3668	0.3152	3359
-0.25	0.0685	3190	0.2977	3730	0.4077	4303	0.2415	3546
-0.30	not applicable		0.2623	4070	0.4077	5187	0.1816	3754

^a $A = \$800$, $w = \$3.86$, $l = 2272$ hours; Ashenfelter–Heckman estimate $(w/l)(\partial l/\partial w) = +0.12 - 0.27 =$ substitution term + income term.

^bCol. (a): vary $(w/l)(\partial l/\partial w)$ holding income term constant.

^cCol. (b): vary $(w/l)(\partial l/\partial w)$ holding substitution term constant.

^dCol. (c): vary $(w/l)(\partial l/\partial w)$ by changing income and substitution terms in same proportion.

^eCol. (d): vary $(w/l)(\partial l/\partial w)$ by changing income and substitution terms so that they contribute to the changes in the same direction and in proportion to their absolute magnitudes.

Twice the standard error of the substitution term is $2 \times 0.12 \times 0.26$, i.e. 0.062; and twice the standard error of the income term is, in absolute value, $2 \times 0.27 \times 0.13$, i.e. 0.069. The sensitivity of the ε estimate to errors in the Ashenfelter–Heckman substitution term alone can be examined by looking down column (a), and to the income term alone by looking down column (b).

The estimates for ε of this subsection are, of course, qualified by the discussion of the preceding two subsections. Our use of the Mirrlees specification is important.

We should also be aware that the nature of the sample, men with nonworking wives, is likely to produce a supply function of hours that is rather inelastic [see Hall (1973)]. It is a subgroup of considerable numerical importance, however. Rosen (1976) estimates for a sample whose supply would be rather more elastic than average, women with working husbands, elasticities around 0.8.

We have used estimates from a particular source, but they do seem to be representative of findings on labour supply both in terms of parameter estimates and that the supply curve is backward sloping over much of its range. It is clear from eq. (7) that this phenomenon requires $\mu > 0$, and so $\varepsilon < 1$, in which case we have a backward sloping curve for $L > l > A/\mu w$. If one believes that supply curves are backward sloping over some range and that the CES is a good specification, then one must conclude that ε is less than one.

Finally, we should reiterate that it may not be labour supply in hours that is changed when the wage is changed but, for example, enthusiasm or effort. If the discouragement of these factors is seen as important by governments they may regard the estimates of ε appearing here as too low for a model of the decision problem they face.

4. The optimum linear tax in the Mirrlees model for constant elasticity of substitution

4.1. The model and computation of optima

The problem is to choose t and G to maximise

$$\frac{1}{v} \int_0^\infty u^v(c_n, l_n) f(n) \, dn,$$

subject to

$$\int_{n_t}^\infty n l_n f(n) \, dn = G + R. \tag{10}$$

The wage rate for an efficiency unit is one, there is a lump sum grant G , no other lump sum income, and a constant marginal tax rate t , so that the individual budget constraint is

$$c = (1 - t)nl + G. \tag{11}$$

Skills are distributed with density function $f(n)$ and we normalise so that $\int_0^\infty f(n) \, dn = 1$. The skill level n_t is that below which individuals do no work, where c_n, l_n are chosen by the individual to maximise $u(c, l)$ subject to (11).

$$u(c, l) = [\alpha(1 - l)^{-\mu} + (1 - \alpha)c^{-\mu}]^{-1/\mu}, \tag{12}$$

where $\varepsilon = 1/(\mu + 1)$ and hence, manipulating the first-order conditions,

$$l_n = \frac{1 - Gk(1 - t)^{-\varepsilon} n^{-\varepsilon}}{1 + k(1 - t)^{1 - \varepsilon} n^{1 - \varepsilon}},$$

where $k = [\alpha/(1-\alpha)]^\varepsilon$. Then

$$n_t = \frac{G^{1/\varepsilon} k^{1/\varepsilon}}{1-t}.$$

We have written the government constraint as a revenue constraint. This procedure is equivalent (see below) to using a production relation. We can interpret R as a fixed cost of production. Units are such that the marginal product of an efficiency hour is one. For a comparison of optima for different parameter values we should think of the marginal product as constant. For a single problem we can suppose that the unit is given by the marginal product at the optimum. An alternative view of R is as a public good. Note, however, that the public good does not influence labour supply or the distribution of utilities. For the case $v = 1$, for example, we could have $u_0(c_n, l_n, R) = u(c_n, l_n) + p_n(R)$, and the optimum for the u_0 problem for given R would be the same as the one we have posed. For the calculation for a particular ε and v , one can think of the given R as being optimum and the tax rate must then be chosen given that R . One can obviously extend the model to include optimisation with respect to R , although this will be more complicated in the case where R influences labour supply.

The equivalence between the production constraint specification and that of a tax revenue constraint is seen as follows:

$$\begin{aligned} \text{tax revenue} &= \text{wages} - \text{consumption} \\ &= \text{output} - \text{profit} - \text{consumption}. \end{aligned}$$

The production constraint is that

$$\text{output} = \text{consumption} + \text{government expenditure}.$$

Combining the above two equations, we have

$$\text{tax revenue} + \text{profit} = \text{government expenditure}.$$

If we count any fixed cost of production as government expenditure, measure profit before any fixed cost, and write $R = \text{government expenditure} - \text{profit}$, we have eq. (10).

Optimum taxation was calculated for the two Mirrlees log-normal cases where $\bar{\mu}$, σ (the mean and standard deviation of the associated normal distribution) are taken as $(-1, 0.39)$ and $(-1, 1)$ with the former regarded as the more realistic case,

$$f(n) = \frac{1}{n\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log_e n - \bar{\mu})^2}{2\sigma^2}\right\}.$$

The parameter α was set so that in the absence of taxation or grants the individual with mean skill would work for $\frac{2}{3}$ of the day in the case $\varepsilon = \frac{1}{2}$. Values

of ε ranged from 0.1 to 0.9 and 0.99. The tax rate (t) was varied between 0 and 1 to search for the maximum. The maximand was calibrated using 0C and ${}^{\frac{1}{2}}C$ defined as follows:

$$\frac{1}{v} u^v({}^0C, 0) = \frac{1}{v} u^v({}^{\frac{1}{2}}C, \frac{1}{2}) = \frac{1}{v} \int_0^\infty u^v(c_n, l_n) f(n) \, dn. \tag{13}$$

In other words, 0C is that consumption which, if equally distributed with zero work hours, would give the same social welfare integral as the allocation $\{(c_n, l_n)\}$ arising from a given tax rate t . A similar interpretation covers ${}^{\frac{1}{2}}C$, where we instead set work hours to half the day. Note that ${}^0C \geq G$, since the utility of each individual is at least $u(G, 0)$, and ${}^0C \leq Y$, where Y (total output) is $(G+R)/t$ (from (10)), because consumption is unequally distributed and positive work is required.

The values of v were 1, -1 , -2 . The case $v = 1$ shows no preference for equality. The utility function (12) is homogeneous of degree 1 in $(c, 1-l)$, and hence the indirect utility function can be written $(w+G)v(w)$, where w is the post-tax wage, and it is easily checked that $v(w) = [(1-\alpha)^\varepsilon + \alpha^\varepsilon w^{1-\varepsilon}]^{-1/(1-\varepsilon)}$. Thus where $v = 1$, the social marginal valuation of a unit increase of lump-sum grant G to an individual facing wage w is $v(w)$, which is independent of G . One can determine which value of v captures one's values as follows. Consider two individuals, A and B , with the same wage, w , but A has a lump sum income of $3w$ and B of w . Thus, A 's 'full' (lump sum plus w times endowment of hours (one)) income is twice that of B . If we considered one marginal unit of lump sum income to B twice as valuable as that to A , we should be opting for $v = 0$; and if we considered the marginal unit 4 times as valuable, we should be choosing $v = -1$ (since the social indirect valuation function is $(1/v)(w+G)^v v(w)$ and hence the marginal valuation is $(w+G)^{v-1} v(w)$). Note that at the optimum these marginal valuations are unequal. See Stern (1973), where the distribution is calculated explicitly.

The calculations of Mirrlees (1971) (see section 2) correspond to $\mu = v = 0$ and $\alpha = \frac{1}{2}$ since he worked with the utility function $\log c + \log(1-l)$.

The maximin, or Rawlsian welfare function, corresponds to $v = -\infty$. The objective becomes the maximisation of G since the worst-off individual, whose welfare is to be maximised, has a zero wage rate.

Where $R > 0$, there is a minimum feasible t , which we call t_b , which satisfies (10) with $G = 0$. Lower values of t would require negative G and this would prevent the worst-off individuals having positive consumption. Note that, for $0 \leq t < 1$ and $\varepsilon < 1$, $t \int_0^\infty n l_n f(n) \, dn$ is monotonic increasing in t for $G = 0$ since in this case the supply curve is backward sloping. Hence values of t larger than t_b allow a lump sum grant G , and thus there exists a $G \geq 0$ corresponding to each $t: t_b \leq t < 1$.

The computation procedure for a given ε was as follows. For a given t , the G satisfying (10) was calculated using Newton's method. Finite integrals were

Table 3
Optimum linear income taxation.^a

ε	t_b	$\nu = 1$			$\nu = -1$			$\nu = -2$			$\nu = -\infty$							
		t_{opt}	${}^0C_{opt}$	π_t	0C_b	t_{opt}	${}^0C_{opt}$	π_t	0C_b	t_{opt}	${}^0C_{opt}$	π_t	0C_b	t_{opt}	${}^0C_{opt}$	π_t	0C_b	
(a) $R = 0$																		
0.1	0	0.546	0.264	0.000	0.262	0.753	0.242	0.000	0.237	0.784	0.260	0.000	0.232	\geq	0.950	0.242	0.000	0
0.2	0	0.362	0.247	0.000	0.246	0.627	0.228	0.000	0.223	0.670	0.240	0.000	0.219		0.926	0.212	0.000	0
0.3	0	0.274	0.229	0.000	0.228	0.541	0.211	0.000	0.207	0.590	0.221	0.000	0.203		0.882	0.188	0.000	0
0.4	0	0.223	0.211	0.000	0.210	0.477	0.195	0.000	0.191	0.527	0.202	0.000	0.187		0.839	0.167	0.000	0
0.5	0	0.191	0.194	0.000	0.193	0.428	0.179	0.000	0.175	0.478	0.185	0.000	0.172		0.797	0.150	0.000	0
0.6	0	0.170	0.178	0.000	0.178	0.389	0.164	0.000	0.160	0.438	0.169	0.000	0.157		0.756	0.135	0.000	0
0.7	0	0.154	0.164	0.000	0.164	0.357	0.151	0.000	0.147	0.404	0.155	0.000	0.144		0.717	0.122	0.001	0
0.8	0	0.141	0.152	0.000	0.151	0.331	0.139	0.000	0.135	0.376	0.142	0.000	0.133		0.682	0.111	0.003	0
0.9	0	0.133	0.141	0.000	0.140	0.309	0.128	0.000	0.125	0.352	0.131	0.000	0.122		0.648	0.101	0.007	0
0.99	0	0.127	0.131	0.000	0.131	0.291	0.119	0.000	0.116	0.334	0.122	0.000	0.114		0.621	0.094	0.012	0
(b) $R = 0.05$																		
0.1	0.175	0.595	0.230	0.000	0.228	0.791	0.227	0.000	0.210	0.818	0.226	0.000	0.200	\geq	0.950	0.209	0.000	0
0.2	0.177	0.406	0.215	0.000	0.214	0.681	0.209	0.000	0.197	0.720	0.208	0.000	0.189		0.938	0.182	0.000	0
0.3	0.180	0.309	0.198	0.000	0.198	0.600	0.192	0.000	0.183	0.646	0.190	0.000	0.175		0.903	0.159	0.000	0
0.4	0.183	0.254	0.182	0.000	0.181	0.540	0.174	0.000	0.167	0.588	0.172	0.000	0.160		0.867	0.139	0.000	0
0.5	0.186	0.217	0.165	0.000	0.165	0.490	0.158	0.000	0.153	0.541	0.155	0.000	0.146		0.832	0.122	0.000	0
0.6	0.189	0.189	0.151	0.000	0.151	0.450	0.143	0.000	0.139	0.501	0.140	0.000	0.133		0.798	0.107	0.000	0
0.7	0.193	0.193	0.137	0.000	0.137	0.417	0.129	0.000	0.126	0.467	0.126	0.000	0.120		0.766	0.094	0.000	0
0.8	0.197	0.197	0.125	0.000	0.125	0.389	0.117	0.000	0.114	0.438	0.114	0.000	0.109		0.736	0.082	0.001	0
0.9	0.201	0.201	0.114	0.000	0.113	0.366	0.106	0.000	0.104	0.414	0.103	0.000	0.099		0.709	0.072	0.002	0
0.99	0.206	0.206	0.104	0.000	0.104	0.347	0.096	0.000	0.095	0.395	0.094	0.000	0.091		0.685	0.064	0.003	0

(c) $R = 0.10$

0.1	0.333	0.646	0.196	0.000	0.195	0.828	0.193	0.000	0.177	0.850	0.193	0.000	0.169	≥ 0.950	0.175	0.000	0
0.2	0.338	0.456	0.183	0.000	0.182	0.733	0.177	0.000	0.167	0.767	0.176	0.000	0.159	$\cong 0.950$	0.153	0.000	0
0.3	0.344	0.344	0.168	0.000	0.168	0.661	0.161	0.000	0.154	0.703	0.159	0.000	0.147		0.131	0.000	0
0.4	0.351	0.351	0.153	0.000	0.153	0.605	0.145	0.000	0.140	0.651	0.143	0.000	0.134		0.112	0.000	0
0.5	0.358	0.358	0.138	0.000	0.138	0.559	0.130	0.000	0.127	0.608	0.127	0.000	0.121		0.095	0.000	0
0.6	0.366	0.366	0.124	0.000	0.124	0.520	0.115	0.000	0.114	0.571	0.113	0.000	0.109		0.081	0.000	0
0.7	0.375	0.375	0.111	0.000	0.111	0.488	0.102	0.000	0.101	0.540	0.100	0.000	0.097		0.068	0.000	0
0.8	0.386	0.386	0.098	0.000	0.098	0.460	0.090	0.000	0.090	0.513	0.088	0.000	0.086		0.057	0.000	0
0.9	0.398	0.398	0.087	0.000	0.087	0.436	0.080	0.000	0.079	0.488	0.077	0.000	0.076		0.047	0.000	0
0.99	0.409	0.409	0.078	0.000	0.078	0.417	0.071	0.000	0.071	0.470	0.068	0.000	0.068		0.039	0.000	0

(d) $R = 0.15$

0.1	0.476	0.701	0.162	0.000	0.161	0.863	0.160	0.000	0.146	0.881	0.159	0.000	0.138	≥ 0.950	0.142	0.000	0
0.2	0.484	0.513	0.151	0.000	0.151	0.784	0.145	0.000	0.137	0.812	0.144	0.000	0.130	$\cong 0.950$	0.123	0.000	0
0.3	0.494	0.494	0.138	0.000	0.138	0.723	0.131	0.000	0.126	0.759	0.129	0.000	0.120		0.104	0.000	0
0.4	0.504	0.504	0.125	0.000	0.125	0.675	0.117	0.000	0.114	0.716	0.114	0.000	0.108		0.086	0.000	0
0.5	0.516	0.516	0.111	0.000	0.111	0.634	0.103	0.000	0.101	0.680	0.100	0.000	0.096		0.071	0.000	0
0.6	0.530	0.530	0.097	0.000	0.097	0.600	0.089	0.000	0.089	0.648	0.087	0.000	0.085		0.058	0.000	0
0.7	0.546	0.546	0.084	0.000	0.084	0.571	0.077	0.000	0.077	0.621	0.074	0.000	0.074		0.046	0.000	0
0.8	0.565	0.565	0.072	0.000	0.072	0.565	0.066	0.000	0.066	0.598	0.063	0.000	0.063		0.035	0.000	0
0.9	0.588	0.588	0.061	0.000	0.061	0.588	0.056	0.000	0.055	0.588	0.053	0.000	0.053		0.026	0.000	0
0.99	0.613	0.613	0.051	0.000	0.051	0.613	0.046	0.000	0.046	0.613	0.044	0.000	0.044		0.019	0.000	0

*Results are given to 3 decimal places. Where 0 is entered the figure is identically zero. An upper bound of 95 percent was placed on the optimum tax rate. Where the optimum is higher than this ≥ 0.950 is printed in the table. For $\nu = -\infty$, ${}^{\circ}C \equiv G$ and ${}^{\circ}C_0 \equiv 0$. Symbols are defined in section 4 of the text.

Table 4a

$\mu = -1.0000$ $\sigma = 0.3900$, $R = 0.0500$, $\bar{n} = 0.3969$, $\alpha = 0.3864$.

ϵ	t									
	0.0000	0.1000	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000
0.1000										
G		INFEASIBLE	0.0073	0.0359	0.0643	0.0926	0.1207	0.1482	0.1749	0.1992
n_t		INFEASIBLE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.2000										
G		INFEASIBLE	0.0065	0.0343	0.0617	0.0886	0.1147	0.1396	0.1622	0.1794
n_t		INFEASIBLE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.3000										
G		INFEASIBLE	0.0056	0.0327	0.0590	0.0844	0.1085	0.1305	0.1488	0.1585
n_t		INFEASIBLE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.4000										
G		INFEASIBLE	0.0047	0.0310	0.0562	0.0800	0.1020	0.1210	0.1349	0.1372
n_t		INFEASIBLE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.5000										
G		INFEASIBLE	0.0038	0.0292	0.0533	0.0755	0.0953	0.1112	0.1207	0.1163
n_t		INFEASIBLE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
0.6000										
G		INFEASIBLE	0.0029	0.0274	0.0503	0.0709	0.0884	0.1013	0.1067	0.0965
n_t		INFEASIBLE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0034
0.7000										
G		INFEASIBLE	0.0019	0.0255	0.0472	0.0661	0.0814	0.0914	0.0930	0.0786
n_t		INFEASIBLE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0007	0.0210

0.8000											
G	INFEASIBLE	0.0009	0.0237	0.0441	0.0614	0.0745	0.0817	0.0800	0.0629		
π_t	INFEASIBLE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0048	0.0567		
0.9000											
G	INFEASIBLE	0.0217	0.0409	0.0565	0.0676	0.0723	0.0680	0.0496			
π_t	INFEASIBLE	0.0000	0.0000	0.0000	0.0000	0.0013	0.0156	0.1010			
0.9900											
G	INFEASIBLE	0.0200	0.0380	0.0522	0.0615	0.0642	0.0581	0.0395			
π_t	INFEASIBLE	0.0000	0.0000	0.0000	0.0002	0.0041	0.0311	0.1389			
ε										t_b	
	$t_{opt} v = -\infty$	G	Y	π_t							
0.1000	≥ 0.9500	0.2085	0.27	0.0000E 00					0.1750		
0.2000	0.9383	0.1820	0.25	0.1491E-40					0.1773		
0.3000	0.9031	0.1585	0.23	0.1352E-16					0.1797		
0.4000	0.8672	0.1386	0.22	0.3634E-09					0.1828		
0.5000	0.8320	0.1215	0.21	0.5216E-06					0.1859		
0.6000	0.7984	0.1067	0.20	0.2265E-04					0.1891		
0.7000	0.7664	0.0937	0.19	0.1825E-03					0.1930		
0.8000	0.7359	0.0824	0.18	0.6673E-03					0.1969		
0.9000	0.7086	0.0724	0.17	0.1634E-02					0.2008		
0.9900	0.6852	0.0643	0.17	0.2854E-02					0.2055		

*See note from table 3. All columns to the left of an 'INFEASIBLE' statement are also infeasible. E-03, for example, means 'multiplied by 10^{-3} '.

Table 4b

 $\mu = -1.0000, \sigma = 0.3900, R = 0.0500, \bar{n} = 0.3969, \alpha = 0.3864, \nu = -1.0000.^a$

ε	t									
	0.0000	0.1000	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000
0.1000										
°C		INFEASIBLE	0.2108	0.2150	0.2186	0.2218	0.2245	0.2264	0.2271	0.2252
1/2C		INFEASIBLE	0.2108	0.2150	0.2186	0.2218	0.2245	0.2264	0.2271	0.2252
0.2000										
°C		INFEASIBLE	0.1980	0.2017	0.2048	0.2073	0.2089	0.2094	0.2078	0.2012
1/2C		INFEASIBLE	0.1988	0.2025	0.2057	0.2082	0.2099	0.2104	0.2087	0.2020
0.3000										
°C		INFEASIBLE	0.1834	0.1866	0.1892	0.1909	0.1916	0.1907	0.1868	0.1755
1/2C		INFEASIBLE	0.1873	0.1908	0.1936	0.1955	0.1962	0.1952	0.1910	0.1789
0.4000										
°C		INFEASIBLE	0.1680	0.1708	0.1729	0.1740	0.1738	0.1716	0.1654	0.1499
1/2C		INFEASIBLE	0.1775	0.1808	0.1832	0.1845	0.1842	0.1816	0.1746	0.1569
0.5000										
°C		INFEASIBLE	0.1530	0.1555	0.1572	0.1577	0.1567	0.1531	0.1448	0.1255
1/2C		INFEASIBLE	0.1693	0.1724	0.1744	0.1751	0.1738	0.1694	0.1594	0.1362
0.6000										
°C		INFEASIBLE	0.1389	0.1412	0.1425	0.1424	0.1406	0.1357	0.1256	0.1032
1/2C		INFEASIBLE	0.1625	0.1654	0.1671	0.1671	0.1646	0.1584	0.1453	0.1172
0.7000										
°C		INFEASIBLE	0.1260	0.1280	0.1289	0.1284	0.1257	0.1197	0.1079	0.0836
1/2C		INFEASIBLE	0.1569	0.1597	0.1609	0.1601	0.1565	0.1483	0.1323	0.1003

0.8000																			
^o C	INFEASIBLE	0.1143	0.1161	0.1166	0.1155	0.1121	0.1050	0.0918	0.0667										
^{1/2} C	INFEASIBLE	0.1523	0.1549	0.1537	0.1541	0.1492	0.1391	0.1205	0.0856										
0.9000																			
^o C	INFEASIBLE	0.1052	0.1052	0.1054	0.1038	0.0997	0.0917	0.0775	0.0526										
^{1/2} C	INFEASIBLE	0.1508	0.1508	0.1512	0.1488	0.1426	0.1307	0.1098	0.0734										
0.9900																			
^o C	INFEASIBLE	0.0963	0.0963	0.0962	0.0942	0.0894	0.0807	0.0660	0.0420										
^{1/2} C	INFEASIBLE	0.1478	0.1478	0.1477	0.1446	0.1373	0.1239	0.1012	0.0643										
ϵ	t_{opt}	G	Y	r_t	${}^oC_{opt}$	${}^{1/2}C_{opt}$	t_b	oC_b	${}^{1/2}C_b$										
0.1000	0.7914	0.1727	0.28	0.0000E 00	0.2271	0.2271	0.1750	0.2096	0.2096										
0.2000	0.6805	0.1349	0.27	0.0000E 00	0.2094	0.2104	0.1773	0.1970	0.1977										
0.3000	0.6000	0.1085	0.26	0.8098E-53	0.1916	0.1962	0.1797	0.1826	0.1865										
0.4000	0.5391	0.0889	0.26	0.6348E-34	0.1741	0.1846	0.1828	0.1673	0.1767										
0.5000	0.4898	0.0734	0.25	0.1287E-24	0.1577	0.1751	0.1859	0.1525	0.1687										
0.6000	0.4500	0.0609	0.25	0.3161E-19	0.1426	0.1673	0.1891	0.1386	0.1621										
0.7000	0.4172	0.0507	0.24	0.8481E-16	0.1290	0.1609	0.1930	0.1258	0.1565										
0.8000	0.3891	0.0420	0.24	0.1578E-13	0.1166	0.1557	0.1969	0.1141	0.1521										
0.9000	0.3656	0.0347	0.23	0.6383E-12	0.1055	0.1513	0.2008	0.1036	0.1485										
0.9900	0.3469	0.0289	0.23	0.6460E-11	0.0964	0.1481	0.2055	0.0949	0.1457										

*See notes from tables 3 and 4a.

calculated using Gauss–Legendre quadrature and infinite integrals using Gauss–Laguerre quadrature. With t and G known we can calculate the level of the maximand. In this manner tables (such as table 4) are constructed giving values of G and 0C over a (one-dimensional) grid of tax rates t , for each ε .

We search, for each ε , over the grid of t to find the optimum. The procedure depends on the existence of at most one local maximum. It is clear from an inspection of tables 4a and 4b that this is satisfied for our problem. The existing grid is examined to find a triple of values of t such that the maximand is higher at the middle value t_2 than the two outer values t_1 and t_3 . The optimum must then lie in the range (t_1, t_3) . (There is an obvious modification if the largest value of the maximand lies at one of the boundaries of the grid.) A fourth point, t_4 , is defined as the midpoint of the larger of the intervals between t_3 and t_1 , t_2 . One of the outer values (t_1 or t_3) is then replaced by t_4 so that we again have a triple with the maximand highest at the interior point. It is clear that in two moves we must have an interval at most half the length of that of the original. The process is continued until the length of the interval is less than a specified value (here 0.001) and then the interior point of the triple is taken as the optimum. Thus our optimum tax rates are accurate to $\pm 0.1\%$.

The calculations presented in tables 3, 4a and 4b are for the case $\bar{\mu} = -1$, $\sigma = 0.39$. We have $\nu = 1, -1, -2, -\infty$; and $R = 0, 0.05, 0.10, 0.15$. A method for understanding how ν measures our value judgements has already been given. The magnitude of R can be judged by comparison with the mean of the $f(n)$ distribution, \bar{n} ($= 0.3969$ in the case $\bar{\mu} = -1$, $\sigma = 0.39$). Output \bar{n} is the maximum conceivable output – it would be produced if everyone worked all day ($I_n = 1$, all n). Alternatively, one can compare R with output $Y (= (G + R)/t)$ which is given in table 4. ${}^0C_{\text{opt}}$ is the value of 0C associated with t_{opt} , the optimum t , and 0C_b is the value associated with t_b .

A pictorial expression of the results is given in the graphs of figs. 2 and 3.

4.2. Discussion of the results

The purpose of the calculations is to examine the sensitivity of the optimum tax rate to the parameters ε , ν and R . We also report, but do not give details on, the results of changing σ , a measure of the dispersion of the distribution of skills.

Table 3(a) gives the optimum tax rates for different values of ε and ν for the case $R = 0$ where taxation is *purely* redistributed. The results are illustrated in fig. 2a. As ε increases from 0 to 1, the optimum tax rate decreases (from 100 percent at $\varepsilon = 0$ for all ν – see section 5) along a curve which is convex to the origin. Higher values of ν (less egalitarian) give lower tax rates as one would expect. The calculations of Mirrlees correspond to the case ($\nu = 0$, $\varepsilon = 1$) and of Atkinson to the case ($\nu = -\infty$, $\varepsilon = 1$). The Mirrlees median rates of around 20 percent compare reasonably with the middle of the interval between $\nu = 1$ and $\nu = -1$. Atkinson found an optimum t of 64 percent for the linear case with

$R = 0$, and a value of α close to the one we are using [see Atkinson (1972, table 2, $a = 0.5$)]. Hence our results correspond with previous studies.

Results with higher values of R are given in tables 3(b), 3(c) and 3(d), and portrayed graphically in figs. 2b, 2c and 2d. It remains true that for a given R and ε , optimum tax rates increase with ν and, understandably, for a given ε and ν they increase with the expenditure requirement R . However, for a given R and ν it is not necessarily true that the optimum tax rate decreases with ε .

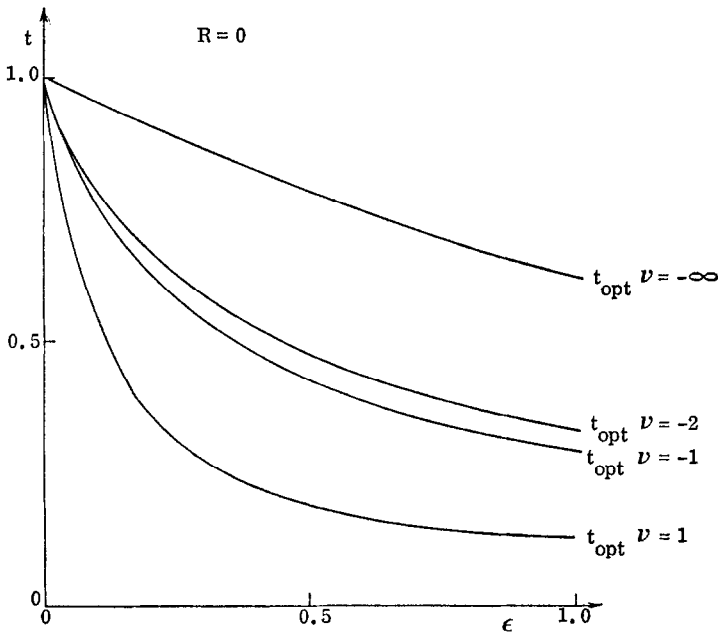


Fig. 2a

For example, for $\nu = 1$ and $R = 0.10$, the optimum tax rate, as a function of ε , has a minimum of approximately 35 percent around $\varepsilon = 0.35$. This failure of monotonicity is more likely the higher is ν and the higher is R .

The reason for the nonmonotonicity is fairly straightforward and can best be understood by examining the behaviour of t_b , the minimum feasible tax rate, as a function of ε . From the calculations it can be seen that t_b increases with ε . The region below the t_b curve (see figs. 2b, 2c and 2d) is infeasible. For higher (less egalitarian) values of ν the optimum has lower tax rates and lower G . When the optimum involves G very close to zero, the optimum tax rate is close to t_b , which *increases* with ε , and thus the optimum lying near to the frontier of the infeasible region also increases with ε .

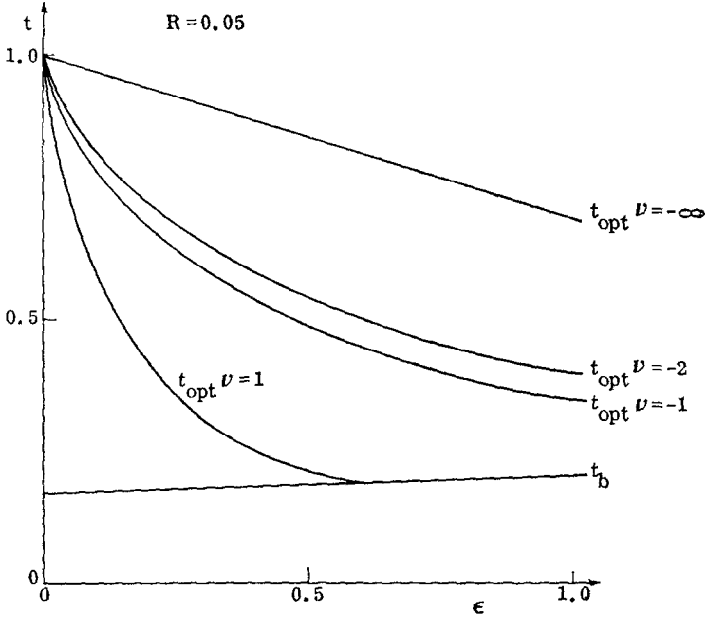


Fig. 2b

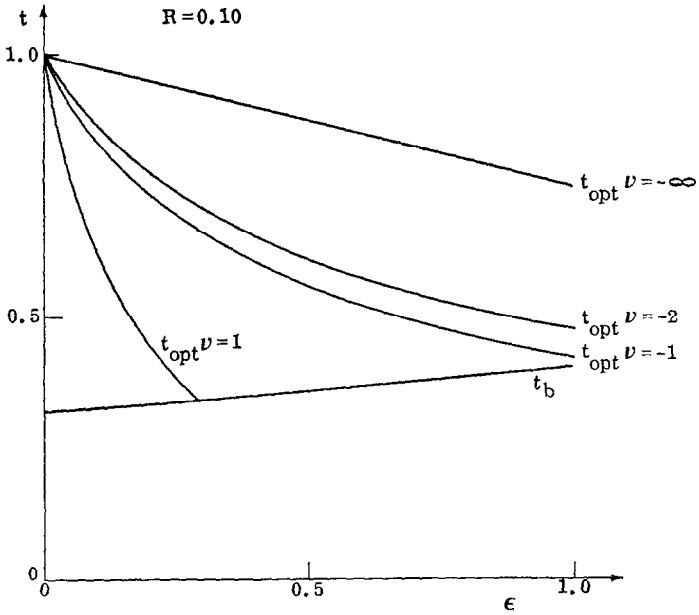


Fig. 2c

Mirrlees (1971, pp. 206–7) found that the use of a much wider dispersion of skills ($\sigma = 1$) gave tax rates which were very much higher. The same conclusion emerges here. For $\nu = 1, R = 0, \sigma = 1$, the optimum tax for $\epsilon = 0.4$ was 62.3%, and for $\epsilon = 0.99$ was 42.9%, compared with $\sigma = 0.39$, which gives a rate of 22.5% for $\epsilon = 0.4$ and 12.5% for $\epsilon = 0.99$. Further, the higher values of ϵ together with $\sigma = 1$ provided the only cases where the proportion of the population not working was high. For $R = 0, \nu = 1, \epsilon = 0.99$, the optimum

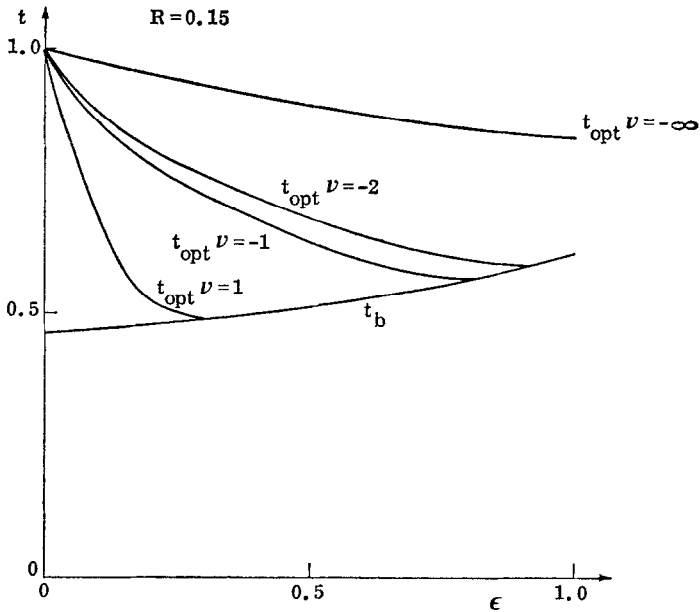


Fig. 2d

proportion of the population not working was 11% (and compare with the n_t figures in table 3).

The total product in our examples usually lies in the range 0.25 to 0.30 and thus a government revenue requirement of 0.15 is very large. Government expenditure on goods and services in the U.K. is approximately 20% of GNP.²³ Thus 0.05 might be viewed as an appropriate figure for R . We have argued that $\nu = 1$ represents no egalitarian preference and have given a method for judging which values of ν capture one's values. I prefer $\nu = -1$, corresponding to an assertion that the social marginal valuation of income should decrease as the

²³The major components of this 20 percent are health services, education and defence. It is arguable whether these should be classified as public goods – fixed costs since at least the first two in the list may also be productive and none are equally available to all.

square of income.²⁴ Our estimate of ε is 0.4 – see section 3. Our *central case*, therefore, is $\varepsilon = 0.4$, $R = 0.05$ and $\nu = -1$, which gives an optimum tax rate of 54%, a grant G of 0.089 and income Y of 0.257. Thus the minimum (0.089) below which no income is allowed to fall is 34% of the average income (total population is one, so Y is also the average income). The utilitarian approach therefore gives taxation rates which are rather high without any appeal to extreme social welfare functions, and need only invoke labour supply functions of the type which are commonly observed. Total net direct and indirect tax rates were around 50% for a broad range of British incomes in 1974 (see *Economic Trends*, Feb. 1976).

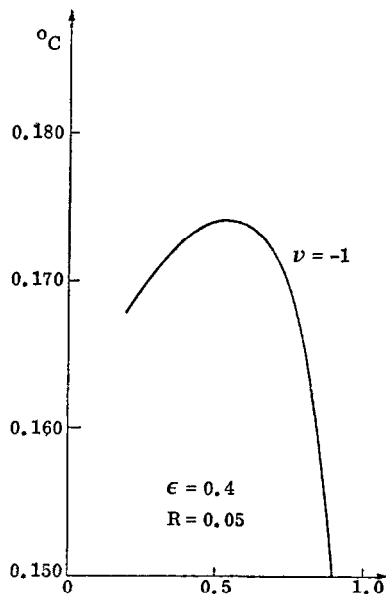


Fig. 3a. (Horizontal axis is tax rate (t)).

The redistributive benefits of taxation can be judged by the comparison of 0C_b and ${}^0C_{opt}$ in table 3. The former gives the welfare level when taxes cover only R (and $G = 0$) and the latter gives the welfare level with the optimum linear income tax. In the central case we have ${}^0C_b = 0.167$ and ${}^0C_{opt} = 0.174$. Thus the redistributive benefits of taxation are worth approximately 5% of income.

It might be argued that such a figure is small when compared with the possible incentive and administrative costs of taxation. This argument would be mistaken. If R is positive, a tax system is necessary, with its administrative costs. The difference in administrative costs between a tax rate t_b and the optimum is that associated with a positive grant G . The incentive aspects of taxation are already

²⁴See Stern (1973).

incorporated in the model. It would be reasonable to argue that our estimate of ε is too small because, for example, in a fully articulated model risk-taking, savings and effort should all be responsive to rewards, but that is a different position. Whichever value of ε one selects there will still be benefits to be obtained from redistribution and, if a tax system is necessary in any case, the benefits, if they are agreed to be such, should be taken. That being said it seems that the benefits are not large for our central case. It is clear that this conclusion is very sensitive to our specification of v . Where $v = -\infty$, the maximin welfare function, the

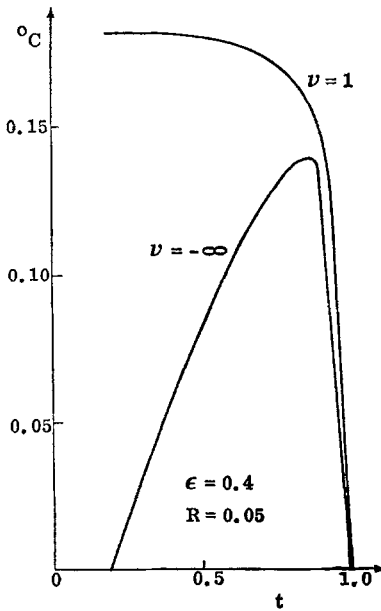


Fig. 3b

measure $^{\circ}C$ is equal to G . Hence, for this case, $^{\circ}C_b = 0$. The redistributive benefits of taxation, here raising the income of the least-skilled from zero, are very large. Where $v = 1$ and we have no special preference for equality the redistributive benefits are very small (see table 3 and fig. 3).

A further suggestion that emerges from the examination of the welfare measure $^{\circ}C$ is that uncertainty about the level of ε might lead us to choose a tax rate lower than the optimum (under certainty) associated with our mean estimate of ε , if our central estimates of ε are in a region where the optimum t falls with ε . The reason (see figs. 3a and 3b) is that, for a given ε , increases in taxation above the optimum seem to give larger welfare losses than deviations below the optimum. Further underestimates of ε lead to bigger changes in the choice of t than overestimates.

An example of the full set of results available from the author for each of the cases discussed here is given in tables 4a and 4b, for the case $\nu = -1$, $R = 0.05$. Table 4a shows the G available from given t , for each ε , together with the t giving maximum G (t_{opt} for $\nu = -\infty$). Table 4a is independent of ν . The values of the maximand for ε , t (calibrated using $^{\circ}C$ and $^{\frac{1}{2}}C$) are shown in table 4b, followed by the optimum t and t_b with corresponding values of other figures associated with t and t_b .

The objective of this section has been to display the sensitivity of optimum tax schedules to the elasticity of supply (given by ε), social values (given by ν) and the expenditure requirement (R). It would, perhaps, be misleading to say that rates are more sensitive to ε than to ν and R since we have no precise measure of sensitivity. But we can claim that values of ε based on observation do give values of the optimum tax rate substantially higher than those which Mirrlees (1971) was surprised to find so low (see section 2) and that an argument for very high rates (say above 70%) must be based (if it is rooted in our model) on a claim that ε is very low (say less than 0.1 or 0.2) rather than an extreme view of values (ν) or the government revenue requirement (R).

5. The optimum tax schedule in the Mirrlees model for the case $\varepsilon = 0$

We saw in section 2 that hitherto attention has been focussed on the addi-log case where the elasticity of substitution ε is one. We saw also in section 3.3 that current estimates of labour supply elasticities give ε around or less than one-half. We should, therefore, see $\varepsilon = 1$ as a polar case with $\varepsilon = 0$ as another polar case. It transpires, not surprisingly, that the case $\varepsilon = 0$ is much easier to work out than the case $\varepsilon = 1$, and the answer is independent of the production function and distribution of skills.

The result is that the optimum tax rate, amongst both linear and general schedules, is 100%. This is a fairly obvious result since consumption and leisure are consumed in fixed proportions and there are no dead-weight losses from income taxation. This does not mean that the first-best optimum can be achieved, however. We should emphasise however that a zero elasticity of substitution does *not* imply inelastic labour supply. The substitution effect of a wage change is zero, but we still have the income effect which gives a backward bending supply curve. An inelastic labour supply is the prerogative of the case $\varepsilon = 1$ where (in the absence of a lump sum income) substitution and income effects exactly cancel.

The CES utility function for the case $\varepsilon = 0$ is

$$u(c, l) = \text{Min}(c, 1-l). \quad (14)$$

We consider first the optimum income taxation problem for the utilitarian maximand where individual utilities are given by (14). We pose this formally as:

Problem P. Find c_n, l_n and a function $g(\cdot)$ such that $\int_0^\infty u_n f(n) dn$ is maximised where $u_n = u(c_n, l_n)$ (u as (14)) and $g(\cdot)$ is such that (i) the individual problem maximise $u(c, l)$, subject to $c \leq g(nl)$ $c \geq 0, 1 \geq l \geq 0$, has a solution (c_n, l_n) , (ii) $\int_0^\infty c_n f(n) dn \leq \int_0^\infty nl_n f(n) dn$. The density function $f(n)$ satisfies $f(n) \geq 0, \int_0^\infty f(n) dn = 1$ and $\infty > \bar{n} = \int_0^\infty nf(n) dn$.

We call an allocation $\{(c_n, l_n)\}$ which satisfies (i) and (ii) feasible. Note that as the problem is posed we are assuming that where there is more than one pair (c_n, l_n) which maximises u the government can select whichever such pair it wishes. Mirrlees (1971, proposition 2) shows that we can restrict attention to $g(\cdot)$ that are nondecreasing and right-continuous. Note that, as the problem is posed, units are such that the n -man produces n if he has zero leisure and the kinks in his indifference curves in (c, l) space lie on the line joining $(0, 1)$ and $(1, 0)$.

Theorem 1. An optimum for problem P is $c_n = \bar{n}/(\bar{n} + 1), l_n = 1/(\bar{n} + 1)$, all n , and $g(x) = \bar{n}/(\bar{n} + 1)$ all x .

Proof. Consider a feasible allocation $\{(c_n^0, l_n^0)\}$. If $1 - l_n^0 > c_n^0$, then we can replace l_n^0 by $l'_n = 1 - c_n^0$ and both (i) and (ii) remain satisfied. If $c_n^0 > 1 - l_n^0$, then we can replace c_n^0 by $c'_n = 1 - l_n^0$ and again (i) and (ii) remain satisfied. We can, therefore, confine attention to feasible allocations $\{(c_n, l_n)\}$ where $c_n = 1 - l_n$. The utility maximising (c_n, l_n) (where we now take $c_n = 1 - l_n$) satisfies

$$g(nl_n) \geq 1 - l_n \geq \lim_{a \rightarrow nl_n} g(a).$$

It is clear (see fig 4) that c_n is nondecreasing with n [see also Mirrlees (1971, theorem 1)].

Now

$$\int_0^\infty u_n f(n) dn = \int_0^\infty c_n f(n) dn \equiv A. \tag{15}$$

From (ii), we have, putting $l_n = 1 - c_n$,

$$A \leq \bar{n} - \int_0^\infty c_n n f(n) dn. \tag{16}$$

But c_n is a nondecreasing function of n , hence ²⁵

$$\int_0^\infty c_n n f(n) dn \geq (\int_0^\infty c_n f(n) dn)(\int_0^\infty n f(n) dn). \tag{17}$$

²⁵Where c_n is continuous, write $c(n^*) = \bar{c}$, the average of c . Then since c is nondecreasing, $n(c - \bar{c}) \geq n^*(c - \bar{c})$ and

$$\int ncf - \bar{n}\bar{c} = \int n(c - \bar{c})f \geq n^* \int (c - \bar{c})f = 0.$$

c_n is continuous as defined here, but if not, a similar argument using $\lim \inf$ in defining n^* will work.

(16) and (17) imply

$$A + A\bar{n} \leq \bar{n},$$

or

$$A \leq \frac{\bar{n}}{\bar{n} + 1}.$$

But $\bar{n}/(\bar{n} + 1)$ is exactly the utility integral obtained from putting c_n, l_n and $g(\cdot)$ as in the statement of the theorem. We, therefore, have the desired result.

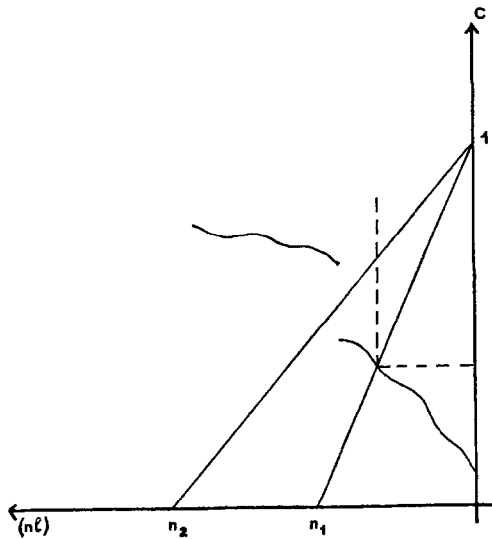


Fig. 4

Suppose the constraint (ii) in problem P is replaced by $\int_0^\infty c_n f(n) dn \leq H(\int_0^\infty n l_n f(n) dn)$, where $H(\cdot)$ is a monotonic increasing function satisfying $H(\bar{n}) > 0$ (where everyone works full-time positive output is possible) and $H(0) < 1$ (satiation is not possible with zero work input). Call the modified problem P'.

Theorem 2. An optimum for the problem P' is $c_n = c^*, l_n = 1 - c^*, g(x) = c^*$, all x , where c^* satisfies $c^* = H(\bar{n} - c^*\bar{n})$. Note that under the conditions imposed on $H(\cdot)$ the equation for c^* defines a unique c^* and $0 < c^* < 1$.

Proof. We follow the proof of theorem 1 but (16) becomes

$$A \leq H(\bar{n} - \int_0^\infty c_n n f(n) dn). \tag{16'}$$

(16') and (17) imply

$$A \leq H(\bar{n} - A\bar{n}).$$

Hence $A \leq c^*$ (check from the definition of c^* and the conditions on H – see fig. 5). But c^* can be achieved by putting c_n, l_n and $g(\cdot)$ as in the theorem, and we have the desired result.

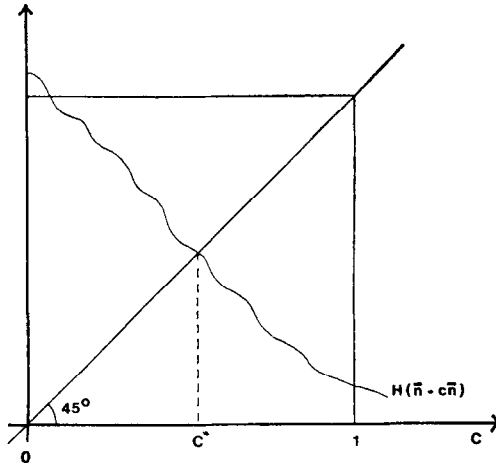


Fig. 5

One can easily see that c^* lies between 0 and 1 by plotting c and $H(\bar{n} - c\bar{n})$ against c and looking at the point of intersection (this also provides a demonstration that $A \leq c^*$) (see fig. 5).

Replace the maximand in problem P by $\int_0^\infty G(u)f(n) dn$, where $u(c, l)$ is as in eq. (14) and G is concave and increasing. Call this problem Q.

Theorem 3. An optimum for problem Q is given by the solution described in theorem 1.

Proof. We can again restrict attention to allocations where $c_n = 1 - l_n$ and where the maximand becomes

$$\int_0^\infty G(c_n)f(n) dn. \tag{18}$$

The problem Q^0 : maximise (18) subject to $\int_0^\infty c_n f(n) dn \leq \bar{n}/(\bar{n} + 1)$, has a solution with a higher utility integral than problem Q since $\bar{n}/(\bar{n} + 1)$ is the maximum

output obtainable under an income taxation scheme and Q^0 is, therefore, a less constrained problem than Q . But a solution to problem Q^0 ,

$$c_n = \frac{\bar{n}}{\bar{n}+1}, \quad (\text{and } l_n = 1/(\bar{n}+1)),$$

can be achieved with the tools of problem Q and we, therefore, have an optimum for problem Q .

Remark. A similar statement to theorem 3 applies if we modify problem P' to Q' by introducing $G(\cdot)$.

We should note two points about these theorems. First, if G is convex we should not expect a solution with 100% marginal tax rates. Presumably, one might be prepared to reduce total output for the 'benefits' of unequal consumption. Secondly, we are forced to rely rather heavily on the notion that the government can choose from the set of bundles that maximise utility for the individual.

The full optimum involves maximising $\int_0^\infty u_n f(n) dn$, subject to $\int_0^\infty c_n f(n) dn \leq \int_0^\infty n l_n f(n) dn$. As before, we can restrict attention to configurations where $c_n = 1 - l_n$ and hence $\int_0^\infty u_n f(n) dn = \int_0^\infty c_n f(n) dn$. Taking a Lagrange multiplier v for the constraint, we have

$$\left. \begin{array}{l} -1 + (n+1)v \leq 0 \\ l \geq 0 \end{array} \right\} \text{comp,}$$

$$\left. \begin{array}{l} -1 + (n+1)v \geq 0 \\ l \leq 1 \end{array} \right\} \text{comp,}$$

$$\left. \begin{array}{l} \int_0^\infty c_n f(n) dn \leq \int_0^\infty n l_n f(n) dn \\ v \geq 0 \end{array} \right\} \text{comp,}$$

as necessary and sufficient conditions for optimality. Thus, the optimum requires²⁶ that for

$$n \leq \frac{1}{v^*} - 1, \quad c_n = 1, \quad l_n = 0,$$

$$n > \frac{1}{v^*} - 1, \quad c_n = 0, \quad l_n = 1,$$

²⁶We need not be specific about c_n at $n = (1/v^*) - 1$ if there is no probability atom concentrated on this level of n .

with v^* determined by

$$\begin{aligned} \text{total consumption} &= \int_0^{(1-v^*)/v^*} f(n) \, dn = \int_{(1-v^*)/v^*}^{\infty} nf(n) \, dn \\ &= \text{total output.} \end{aligned}$$

The l.h.s. decreases from 1 to 0 as v goes from 0 to 1, and the r.h.s. increases from 0 to \bar{n} as v goes from 0 to 1, and we therefore have a unique solution for some $v^* > 0$.

The solution was calculated for n distributed log-normally with the parameters used by Mirrlees (1971). The cases are specified by $\bar{\mu} = -1, \sigma = 0.39$ and $\bar{\mu} = -1, \sigma = 1$, where $\bar{\mu}$ and σ are the mean and standard deviation of the underlying normal distribution. Mirrlees suggested $\sigma = 0.39$ was the more realistic. In the former case $v^* = 0.766$ and the proportion of the population not working is 32.0%. The ratio β of output in the optimum income taxation solution to output in this, the full optimum, is 0.889. Output is the obvious measure of welfare here since the utility function is linear in consumption and the Atkinson (1970) equally-distributed-equivalent (EDE) measure is equal to output itself.

The corresponding figures for $\bar{\mu} = -1, \sigma = 1$ are $v^* = 0.727$, 50.8% of the population not working, and a ratio β of output in the optimum income tax case to output in the full optimum of 0.744. With the wider spread of skills, more of the population is idle, yet the increased availability and use of skills at the upper end gives a greater proportional output increase in the movement from optimum income taxation to the full optimum.

The full optimum here provides an illustration of the Mirrlees result that the full optimum has utility decreasing with skill level, provided leisure is a normal good. The form of decrease is rather bizarre, however, with consumption dropping from its maximum, one, to its minimum, zero, at $n = (1/v^*) - 1$.

The full optimum was also calculated for maximand $\int_0^{\infty} G(u)f(n) \, dn$, where $G(x) = \log_e x$ and $-1/x$. The analysis is similar and the solution, where $h(\cdot) \equiv G'^{-1}(\cdot)$, is

$$\begin{aligned} c_n &= h((n+1)v^*) \quad \text{if } h((n+1)v^*) \leq 1, \\ c_n &= 1 \quad \quad \quad \text{if } h((n+1)v^*) \geq 1, \end{aligned}$$

with $l_n = 1 - c_n$ and v^* determined from

$$\int_0^{\infty} c_n f(n) \, dn = \int_{((x_0/v^*)-1)}^{\infty} nl_n f(n) \, dn,$$

if $(x_0/v^*) - 1 > 0$, where x_0 is defined by $h(x_0) = 1$, and

$$\int_0^{\infty} c_n f(n) \, dn = \int_0^{\infty} nl_n f(n) \, dn$$

if $(x_0/v^*) - 1 \leq 0$.

The solutions for the two log-normal cases and $G(x) = \log_e x$, and $G(x) = -1/x$, involved the whole population working. In all four cases $x_0 = 1$. The value of the maximand at the full optimum was calibrated by c^* where $G(c^*) = \int_0^\infty G(c_n)f(n) dn$ (again analogous to Atkinson's EDE). We then calculated the ratio β of output (equals consumption per head) in the optimum income tax case to c^* . For $\bar{\mu} = -1, \sigma = 0.39$, we have

$$\beta = 0.994, \quad G(x) = \log_e x,$$

$$\beta = 0.997, \quad G(x) = -1/x.$$

For $\bar{\mu} = -1, \sigma = 1$, we have

$$\beta = 0.935, \quad G(x) = \log_e x,$$

$$\beta = 0.963, \quad G(x) = -1/x.$$

In the case where $\varepsilon = 0$, therefore, a small amount of inequality aversion (concavity of $G(\cdot)$), brings the full optimum rather close to the (completely equal) optimum income tax solution. Presumably the increased availability and use of skills at the upper end is again giving bigger output increases, for the case $\sigma = 1$, when we move from optimum income taxation to the full optimum.

The small welfare difference between the full optimum and the optimum income taxation solutions contrasts with the impression one has from Mirrlees' (1971, p. 206) calculations that the full optimum gives substantial welfare differences from the optimum income taxation solution.

6. Concluding remarks

We have discussed most of the ingredients of a model of optimum income taxation both in terms of how the different components should be specified and the effects of varying the specifications of optimum tax schedules. It was suggested that the Bergson-Samuleson social welfare function, nondecreasing in each argument and which is almost universally adopted in the literature on welfare economics, may not be a good representation of the values of many who would wish to comment on appropriate income taxation. This did not involve the rejection of the criterion, and the usual form of welfare function was used in most of the paper. Our particular concern, however, was the elasticity of labour supply and the related problem of the skill distribution.

It was argued in section 3 that the assumption that individuals differ only in skills, and not in preferences between work and leisure, is very convenient for estimation purposes. Since backward bending supply curves of hours of work are generally observed in practice, such a specification must lead to estimates

of the elasticity (ε) of substitution between work and leisure which are considerably less than one. We used the figure $\varepsilon = 0.4$. On the other hand, a specification where individuals differ in their preferences as well as their skills can produce very different results, and we saw how a completely inelastic labour supply might be inferred when supply was in fact sensitive to the wage rate (positively or negatively).

We suggested that it is hard to guess whether the distribution of earnings overestimates or underestimates the dispersion of the distribution of skills since backward bending supply curves make the earnings dispersion narrower, and nonskill factors wider, than the dispersion of skills.

There are deeper questions, however, concerning the supply of effort and skills. If a skill is acquired, then the motive for acquisition should be in the model. Thus any distribution of fixed skills in the model should be of unacquired skills. We do not know how to measure skill or the difference between acquired and unacquired skills, if indeed the distinction is sound. Similarly, the measurement of effort is extremely complex and very difficult to disentangle from skill. The motive for supply of effort may not always be monetary reward – for example, many work who would receive more on public welfare. It is precisely because of the incentives for the individual to conceal his levels of skill and effort, that we build models of income taxation rather than ability taxation.

Our central estimate (see section 4), using $\varepsilon = 0.4$, of the optimum linear income taxation rate was 54%, compared with levels of 20 or 30% which emerge from models where $\varepsilon = 1$. We found that the optimum tax rate was rather sensitive to ε and proved in general, (see section 5) that for $\varepsilon = 0$, the optimum tax rate (linear or nonlinear) is 100%.

Very high tax rates can only be justified by appeal to low ε and not to high revenue requirements or extreme preference for equality. The optimum tax rates are, however, rather sensitive to: ε , social values (v) and revenue requirements (R).

An interesting feature to emerge was, where there is a large revenue to be raised and values are not particularly egalitarian, that the optimum tax rate may not be monotonic in the elasticity of substitution. The reason is that the minimum tax required to raise the revenue increases with ε when the desired grant G is small.

We found that, in our central case, the gains from optimum linear taxation, as compared with minimum taxation to meet revenue requirements, were not large but that this conclusion was, not surprisingly, very sensitive to distributional values. On the other hand, in the case where $\varepsilon = 0$, there is no loss from a restriction to optimum linear income tax as opposed to nonlinear, and the optimum income tax solution gives welfare close to the full optimum.

Finally we should emphasise that the study of optimum income taxation is in its infancy, there is much work, empirical and conceptual as well as theoretical, to do, and therefore all our estimates and calculations must be viewed with

circumspection and as attempts to understand the best model currently available rather than prescriptions for policy.

References

- Ashenfelter, O. and J. Heckman, 1973, Estimating labour-supply functions, in: Cain and Watts (1973).
- Atkinson, A.B., 1970, On the measurement of inequality, *Journal of Economic Theory*, and in Atkinson (1973b).
- Atkinson, A.B., 1972, 'Maxi-min' and optimal income taxation, (mimeo) presented to the 1972 Budapest meeting of the Econometric Society.
- Atkinson, A.B., 1973a, How progressive should income tax be?, in: Phelps (1973).
- Atkinson, A.B., ed., 1973b, *Wealth, income and inequality* (Penguin, Harmondsworth).
- Atkinson, A.B. and J.E. Stiglitz, 1975, The design of tax structure: Direct versus indirect taxation, *Journal of Public Economics* 6, 000-000 (this issue).
- Bevan, D., 1974, Savings, inheritance and economic growth in the presence of earnings inequality, mimeo, Oxford.
- Cain, G.C. and N.W. Watts, 1973, *Income maintenance and labour supply* (Rand McNally, Chicago).
- Fair, R.C., 1971, The optimal distribution of income, *Quarterly Journal of Economics*.
- Feldstein, M.S., 1973, On the optimal progressivity of the income tax, *Journal of Public Economics* 2, 357-376.
- Finegan, T.A., 1962, Hours of work in the United States: A cross-sectional analysis, *Journal of Political Economy*.
- Hall, R.E., 1973, Wages, income and hours of work in the U.S. labor force, in: Cain and Watts (1973, ch. 3).
- Hall, R.E., 1974, Labour supply and the negative income tax experiment, mimeo, M.I.T.
- Lydall, H.F., 1968, *The structure of earnings* (Oxford).
- Mirrlees, J.A., 1971, An exploration in the theory of optimum income taxation, *Review of Economic Studies*.
- Owen, J.D., 1971, The demand for leisure, *Journal of Political Economy*.
- Phelps, E.S., 1973, Wage taxation for economic justice, *Quarterly Journal of Economics*.
- Plato, 1960, *The laws*, translated by A.E. Taylor (E.P. Dutton, New York).
- Roberts, K.W.S., 1974, On optimal linear income tax schedules, mimeo, Oxford.
- Roberts, K.W.S., 1974, Calculations on the optimal linear income tax, mimeo, Oxford.
- Rosen, H., 1975, A methodology for evaluating tax reform proposals, *Journal of Public Economics* 6, 000-000 (this issue).
- Rosen, S., 1969, On the interindustry wage and hours structure, *Journal of Political Economy*.
- Sadka, E., 1973, *Income distribution, incentive effects and optimal income taxation*, unpublished Ph.D. dissertation (M.I.T., Cambridge, MA).
- Schwartz, J., 1975, An examination of some theoretical and methodological issues relating to the study of income, income distributions and inequality, mimeo.
- Simons, H.C., 1938, *Personal income taxation* (University of Chicago, Chicago).
- Stern, N.H., 1973, Welfare weights and the elasticity of the marginal valuation of income, (mimeo) presented to I.I.M. Berlin Symposium on Planning, 1973. Revised version to appear in the proceedings of the 1976 AUTE conference in Edinburgh.
- Tawney, R.M., 1964, *Equality* (Allen and Unwin, London).
- Vickrey, W., 1974, *Agenda for progressive taxation* (Ronald Press, New York).
- Winston, G.C., 1966, An international comparison of income and hours of work, *Review of Economics and Statistics*.