VATT-KESKUSTELUALOITTEITA

VATT-DISCUSSION PAPERS

128

TAGGING AND TAXING: THE OPTIMAL USE OF CATEGORICAL AND INCOME INFORMATION IN DESIGNING TAX / TRANSFER SCHEMES

Ritva Immonen

Ravi Kanbur

Michael Keen

Matti Tuomala

Valtion taloudellinen tutkimuskeskus

Government Institute for Economic Research

Helsinki 1996

ISBN 951-561-181-4

ISSN 0788-5016

Valtion taloudellinen tutkimuskeskus

Government Institute for Economic Research

Hämeentie 3, 00530 Helsinki, Finland

J-Paino Ky

Helsinki, August 1996

TAGGING AND TAXING: THE OPTIMAL USE OF CATEGORICAL AND INCOME INFORMATION IN DESIGNING TAX/TRANSFER SCHEMES

Ritva Immonen*, Ravi Kanbur†, Michael Keen‡ and Matti Tuomala§

September 1993
This version: June 21st 1996

Abstract: It is widely recognised that optimal tax/transfer schemes will generally involve elements of both 'tagging' (the use of categorical benefits) and 'means-testing' (income-relation of benefits). This paper explores the optimal design of such mixed schemes. Simulations suggest a striking qualitative dissimilarity between the group-specific schedules optimally imposed on the poorer and richer groups: broadly speaking, the optimal marginal tax rate decreases with income amongst the latter, but increases with income amongst the former. This latter observation, potentially important for policy, runs counter to the conventional wisdom from previous simulations; the reconciliation, we argue, lies in the role played in optimal tax design by the revenue constraint. The simulations also suggest that gains from the appropriate use of categorical information can plausibly be substantial.

JEL Nos.: H21, I38

* University of Jyväskylä, Seminaarinkatu 15, 40100 Jyväskylä, Finland
† The World Bank, 1818 H Street NW, Washington D.C. 20433, USA
‡ University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK and the Institute of Fiscal Studies
§ University of Tampere, P.O. Box 607, SF-33101 Tampere, Finland.

Acknowledgements. We are grateful to two referees for very helpful comments, and to the Yrjö Jahnsson Foundation for financial support enjoyed by the first- and last-named authors. Errors and views are ours alone.

Contents

1 Introduction

It is now widely agreed that optimal social security schemes involve elements of both 'tagging' (the use of categorical information) and 'means-testing' (income-relation of benefit payments). On one hand, purely categorical (or 'contingent') benefits – benefits, that is, which are paid at a flat rate conditional on some event associated with neediness (old age, for instance, or the presence of children) – have the attraction of targeting those groups most at risk of poverty whilst maintaining low marginal tax rates, but the disadvantages that benefits will leak to non-poor members of poor groups and, conversely, prove inadequate for poor members of non-poor groups. Means-testing, on the other hand, has the attraction of reducing or denying benefit to those not in need whilst providing a decent minimum to all, but the disadvantage that – as is well-known from the discussion of negative income tax and social dividend schemes – the financing of a reasonable guaranteed minimum income is likely to require high marginal tax rates on much of the population.[1] Hence Atkinson (1992), for example, concludes and emphasises that: "The issue of policy design is not therefore a confrontation between fully universal benefits and pure income testing; rather the question is that of the appropriate balance of categorical and income tests."

The purpose of this paper is to explore the nature of the optimal combination of categorical and income-related benefits. For while the desirability of such a combination is now recognised, there has been, as far as we aware, no formal and general analysis of its appropriate form. In the literature on optimal non-linear income taxation initiated by Mirrlees (1971), households are assumed to differ only in the single dimension of 'ability', leaving no scope for the use of categorical information. The

[1] There are of course many other considerations that also arise in comparing categorical and means-tested benefits, not the least of which is the possibility of incomplete take-up of the latter as a consequence of stigma and/or hassle: see, for example, Atkinson (1992), Besley (1990) and Cowell (1986).

potential value of such information was emphasised by Akerlof (1978) and its optimal use analyzed further by Stern (1982), Kanbur (1987), Kanbur and Keen (1989) and Keen (1992). These analyses of categorical benefits proceed, however, on the basis of arbitrary restrictions on the form of income-relation permissible within groups: Kanbur and Keen (1989), for instance, presume within-group schedules to be linear, so precluding the ability to recoup benefit from the better-off that is a central feature of the means-testing strategy. An intriguing and thoughtful account of the essential issue – the design of distinct non-linear income tax/transfer schemes for sub-groups of the population linked by inter-group transfers – is provided by Dilnot, Kay and Morris (1984; henceforth DKM). Their analysis, however, is entirely informal. More to the point, it will be argued here that one of their most central (and influential) conclusions, whilst intuitively plausible, is inconsistent with the conventional wisdom that has emerged from the literature on optimal non-linear income taxation. One of our main objectives will then be to find a reconciliation of that tension. (To anticipate the conclusion, it is the conventional wisdom of the optimal tax literature, not the intuition of DKM, that will be found wanting).

The paper focusses, more specifically, on two key issues. The first is qualitative: What are the shapes of optimal tax/transfer schedules when categorical information can be used to apply different schedules to different groups? It is here that the DKM analysis is pivotal, suggesting as it does that marginal tax rates should rise with income in the 'poor' group and fall with income in the 'rich'. The second question is quantitative: How substantial are the potential welfare gains from applying distinct schedules to distinct groups?

These issues are explored here in the context of a simple model of optimal taxation in which categorical status is both exogenous and costlessly observed by the authorities. In practice, of course, the authorities are rarely able to distinguish between groups in such a clear-cut way. While policy may distinguish between the currently old and young, for example, sensible design of that policy requires recognition that these are, in the long run, the same people; requires, that is, a life-cycle perspective. Rather closer to the present analytical structure are the distinctions by family composition and health status commonly observed in tax-benefit systems. These distinctions commonly fall far short of the full diversity of tax structure allowed here: often a common tax schedule is applied to all, but with simple exemptions, deductions and/or credits differentiated by type. Moreover, mis-classifications of family or health position may occur either by simple error or by deliberate mis-reporting, and these characteristics may in some cases be objects of choice. The authorities may thus be more constrained in differentiating along these dimensions than is assumed in the formal model used here. Note too that there may be cases in which it is relatively straightforward to distinguish between groups – men and women, for instance, or residents of different states in a federation – but regarded as horizontally inequitable to use that distinction as a basis for differential tax treatment. There are thus many considerations bearing on the optimal use of categorical information that are not addressed here. Our reason for precluding them is familiar: understanding optimal policy in the simplest case is a prerequisite for the useful analysis of more realistic circumstances. For instance, since differentiation between groups is costly, both administratively and in terms of potential adverse selection, it seems useful to have a sense of the benefits against which such costs must ultimately be weighed. More fundamentally still, the tension between DKM and the conventional wisdom mentioned above indicates that the fundamental qualitative nature of the simple case is as yet only imperfectly understood.

Section 2 of the paper focusses on theory, developing both the informal DMK analysis and a formal treatment of the problem as one of optimal taxation. This raises more questions than can be solved analytically. In particular, and not surprisingly, the problem proves too complex for an unambiguous analytical answer to be given to the first of the questions above, concerning the intra-group patterns of optimal marginal tax rates. To pursue this, and to address the second of the issues raised above, we turn in Section 3 to numerical simulation. Section 4 concludes.

2 The optimal use of categorical and income information

The issue with which we are concerned can be formalised as a conceptually straightforward extension of the familiar one-dimensional problem of non-linear tax design. Suppose that the population (the size of which is normalized to unity) can be divided into two mutually exclusive and exhaustive groups, labelled 1 and 2. Individuals are unable to alter or disguise the group to which they belong, which is observed costlessly by the government. Members of each group i have preferences $u^i(x, y)$ defined over consumption x and labour supply y, but differ in their hourly gross wage ('ability') n. The within-group structure of the model is thus exactly as in Mirrlees (1971). The two groups differ in the form of their preferences and/or the distribution of their abilities, the latter being described for each group by a continuous density function f_i (with corresponding distribution $F_i(n)$) on support $[\underline{n}_i, \overline{n}_i]$.

The government's problem is to design possibly non-linear tax/benefit schedules $T_i(z)$ for the two groups, $z \equiv ny$ denoting gross income. Its objective, we assume, is to maximise a utilitarian social welfare function of the general form

$$W = \int_0^\infty \{\theta G[u^1(\cdot)]f_1(n) + (1-\theta)G[u^2(\cdot)]f_2(n)\}dn \,, \tag{1}$$

where θ denotes the proportion of the population in group 1 and $G(\cdot)$ intermediates between the representation of an individual's preferences and the quantification of their social worth. The government faces two constraints on this optimisation. The first is that of raising (or spending) some exogenous amount of revenue R:

$$\int_0^\infty \{\theta T_1(\cdot)f_1(n) + (1-\theta)T_2(\cdot)f_2(n)\}dn = R\,. \tag{2}$$

Note that this constraint applies collectively to both groups, opening the way to the inter-group transfers one would expect from the optimal use of categorical information. The second constraint is that of self-selection: in each group i, a person with wage n chooses y to maximise $u^i(x,y)$, with $x = ny - T^i(ny)$.

It is helpful to think of the government's problem as comprising two steps. At the first, it takes as given for each group i some arbitrary group-specific revenue requirement R_i and derives the group-specific tax schedule which maximises, conditional on that revenue requirement, group i's contribution $\theta_i \int G(u^i)f_i dn$ to social welfare (where $\theta_1 = \theta$ and $\theta_2 = 1-\theta$). This means, in effect, solving a problem of exactly the same form as the original Mirrlees (1971) one-dimensional problem. Hence, from previous analyses of that problem,[2] the marginal tax rates of these conditionally optimal schedules are characterised by the condition[3]

$$T_i'(z) = -\left(\frac{\mu_i(n)u_x^i s_n^i}{\lambda^i f_i(n)}\right) \tag{3}$$

where λ^i is the Lagrange multiplier on the group-specific revenue

constraint,

$$s_n^i(x,z,n) = -\frac{\partial}{\partial n}\left(\frac{u_y(x,z/n)}{nu_x(x,z/n)}\right) \tag{4}$$

is assumed negative,[4]

[2]See for example Kanbur, Keen and Tuomala (1994).

[3]Derivatives are indicated by subscripts for functions of several variables and a prime for functions of just one.

[4]This condition – for which normality of consumption is sufficient – is Assumption B of Mirrlees (1971): it ensures that post-tax income increases with the wage rate.

$$\mu_i(n) = \int_0^n (G'u_x - \lambda^i)(1/u_x)\exp(-\int_p^n (u_{nx}/u_x)dm)f(p)dp \qquad (5)$$

is the multiplier associated with the self-selection constraints and satisfies the transversality condition $\mu^i(\underline{n}_i) = \mu^i(\overline{n}_i) = 0$, and $u_n(x,y,n) \equiv -yu_y(x,y)/n$. Proceeding in this way for all conceivable R_i then gives the maximised contribution of group i to collective welfare as a function $W_i(R_i)$ of the revenue constraint imposed on it, with associated shadow price of revenue $\lambda^i = -W_i' > 0$.

At the second step, the government simply chooses the optimal allocation of the aggregate revenue requirement R over the two groups: selects R_1, that is, to maximise $W_1(R_1) + W_2(R - R_1)$. This is just a matter of choosing the optimal inter-group transfer, and requires equating across groups the marginal social cost of raising an additional unit of revenue, so that

$$\lambda^1 = \lambda^2 \equiv \lambda \,, \qquad (6)$$

where λ is the marginal social cost of public funds at the optimum. Using (6) in (3), the optimal pattern of within-group marginal tax rates is thus of the form

$$T_i'(z) = -\left(\frac{\mu^i(n)u_x^i s_n^i}{\lambda f_i(n)}\right) \qquad (7)$$

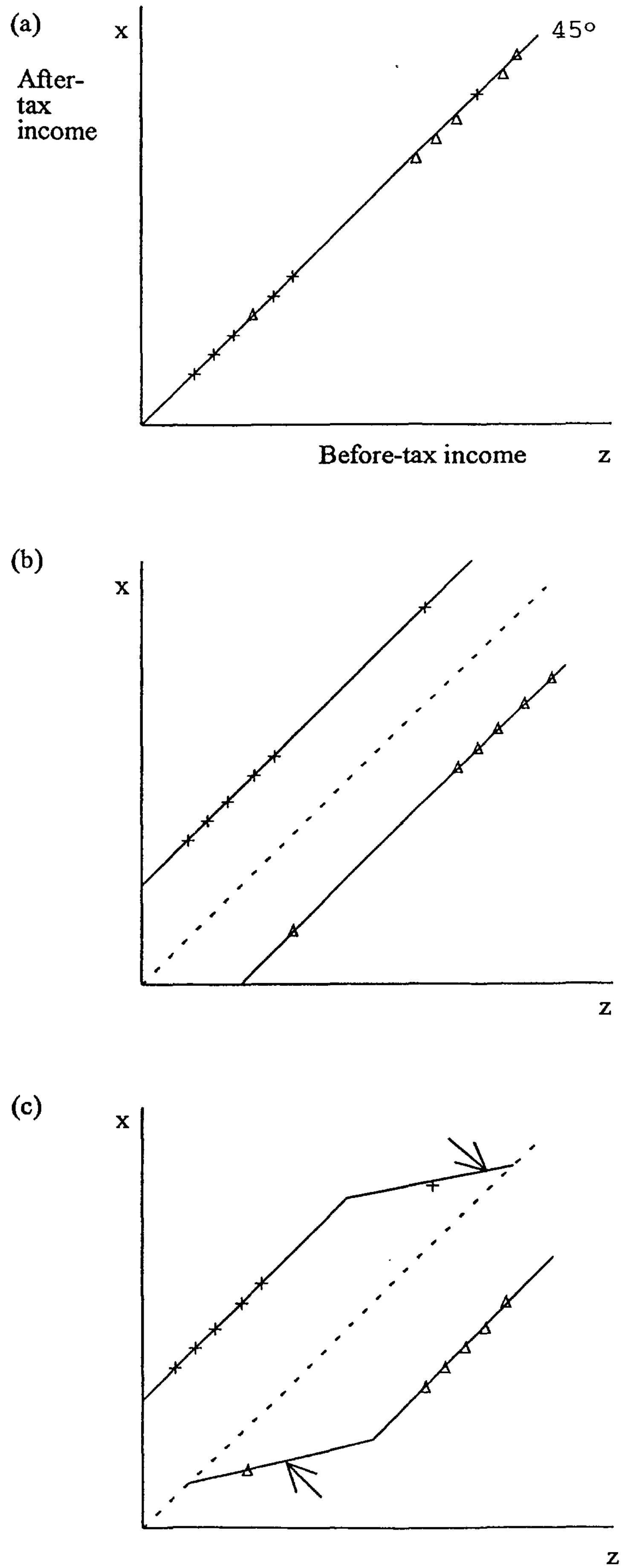
with (6) also being used in (5).

There are few insights into the present concerns to be found in (7). As one would expect, the standard qualitative conclusions from the Mirrlees problem (concerning, in particular, the end-points) apply in group-specific form: the marginal tax rate is optimally non-negative, for example, and is zero on the highest earner in each group. Beyond that there is little – at this stage, for we shall return to (7) later – to be said.

The problem just described is essentially that underlying the informal discussion of DKM, who argue diagrammatically for a particular pattern of group-specific tax schedules. Their argument is illustrated in Figure 1. Suppose, as we henceforth shall, that group 2 is the poorer of the two in the sense of having a less favourable distribution of abilities; for brevity, we shall simply call group 1 'rich' and group 2 'poor'. Representing members of group 2 by "+" and members of group 1 by "$\triangle$", the no-tax outcome is then as in Panel (a), with members of group 2 clustered at relatively low levels of income. Panel (b) shows the effect of a purely categorical benefit: a poll subsidy to members of group 2 financed by a poll tax on members of group 1. This clearly brings about a considerable equalisation of average post-tax incomes in the two groups. But clearly too the scheme is very generous to the outlying 'rich' member of the poor group 2 and, at the same time, very harsh on the outlying poor member of the rich group 1. This then points to a gain from introducing non-linearities in the group-specific schedules of the kind shown in Panel (c), increasing the tax paid by the rich outlier in the poor group and using the proceeds to reduce that paid by the poor outlier in the rich group. Such a reform has the merit of increasing inter-group transfers to the poorest of the poor but the disadvantage of raising the marginal tax rates on these outliers; but since there are, by definition, few outliers, the first-order distributional gain will (up to some point) outweigh the additional deadweight loss. The implication of this line of argument is that the optimal combination of categorical and income information has the feature that the marginal tax rate decreases with income in the rich group but increases with income in the poor group.

The argument is intuitively appealing. On closer inspection, however, there is a puzzle. For recall from the discussion above that the optimal group-specific schedules are the solution to a standard one-dimensional Mirrlees problem (in which the revenue to be raised reflects the optimal inter-group transfer determined in the second step of two-dimensional problem). And while there are no general analytical results on the pattern of marginal tax rates required to solve the Mirrlees problem, there

Figure 1: The Dilnot-Kay-Morris argument



is a conventional wisdom that has emerged from its numerical simulation: optimal marginal tax rates have generally been found to be monotonically *decreasing* with income[5] (see, for instance, Tuomala (1990)). These observations would lead one to expect decreasing marginal tax rates within both groups. But the DKM argument points, on the contrary, to *increasing* marginal rates within the poor group.

Something, it would seem, has to give: either the DKM argument or the conventional wisdom of the optimal tax literature[6] that marginal tax rates should optimally fall over much of the income range. Certainly the DKM argument cannot apply at the top of the income distribution: since the marginal tax rate on the richest of the poor is non-zero and the marginal tax rate is everywhere non-negative, at the top of the distribution the marginal tax rate can only fall. But it is well-known that such conclusions may apply only within the very upper reaches of the income range, and so do not preclude the possibility that the marginal tax rate may increase over most of the income distribution. The central question thus remains: which is it that must be jettisoned, the DKM prescription or the conventional wisdom?

Given the difficulty of deriving analytical results on this from the model above, there are two main approaches that might be adopted. One is to pose the question instead in a setting in which the variation of ability within categorical groups is discrete rather than continuous. With only two ability types in each group, the result of Stiglitz (1982) clearly applies: the optimal marginal tax rate on the richest in each categorical group is zero, that on the poorest is strictly positive. Marginal rates will thus optimally fall within each group. At least three groups are therefore required to admit the possibility that the DKM prescription of increasing marginal tax rates within the poor group[7] may be correct. But analytical conclusions seem no easier to obtain with three ability types than with a continuum. In what follows we therefore

[5] Except, perhaps, in a small range at the bottom of the income distribution.

[6] The conventional wisdom of the non-economist, of course, seems to be quite the reverse: that marginal tax rates should increase with income.

[7] Except, as noted above, at the very top of the income distribution.

pursue the second approach: numerical simulation of the continuum model. Simulations are clearly needed to address the quantitative issues raised at the outset. On the analytical side, of course, imposing structure on the form of preferences and the distributions of ability[8] clearly limit the generality of the conclusions that can be drawn. It will be seen, however, that the central finding has a clear intuition that suggests a wide applicability.

3 Numerical simulations

The functional forms and parameter values used in simulating the optimal two dimensional scheme characterised above are chosen to be comparable, so far as is possible, with those conventionally used in simulating the one-dimensional Mirrlees problem. The distribution of the pre-tax wage within each group is assumed lognormal with mean α_i and variance σ_i. For Group 1, these parameters are set at values typical of those which have been commonplace, following Mirrlees (1971) – who in turn cites Lydall (1968) – in the simulation of optimal income taxes: α_1 = -0.8 and σ_1 = 0.4. Group 2 is taken to be poorer, having a lower average wage, with the particular parameter values being chosen so that groups 1 and 2 stand in broadly the same relation as do non-pensioners and pensioners (respectively) in Finland: specifically, we take α_2 = -1.4 and σ_2 = 0.55. Our purpose, it should be emphasised, is not to investigate that particular application, but merely to ensure reasonable orders of magnitude.

To begin with, we assume Cobb-Douglas preferences

$$u^i = (1-\rho_i)lnx_i + \rho_i ln(1-y_i) \qquad \rho_i \in (0,1) \, , \tag{8}$$

the time endowment being normalised at unity.

[8]An alternative approach would be to leave these unrestricted but instead constrain the shape of the tax schedule: requiring the marginal tax rate to be linear in income, for example.

In the benchmark Case A, the two categoric groups differ only in the distribution of ability, as just described. In Case B, members of the poorer group also have a lower disutility of effort, and so work more hours at any given wage (in the absence of tax): this does not sit well with thinking of this group as pensioners of course (one reason why we do not claim to be capturing that case), but it seems of interest to consider a simple case in which the poor are also the hardest working.

Social welfare is taken to be the simple sum of utilities cardinalised as in (8), so that $G(\cdot)$ in (1) is the identity function. This of course implies a particular view of the relative neediness and social deservingness of the two groups. For the benchmark case in which preferences are identical across the two groups, the judgement is that individuals with the same (x, y)-bundle contribute the same amount to social welfare, irrespective of the group to which they belong. When preferences differ, the judgement is simply that (8) correctly allows for differences in needs and deservingness. Though perfectly coherent, this is certainly *ad hoc*; but so too – in the absence of a more complete treatment of the profound and to some degree unresolved issues raised by differences in neediness or worth (see, for example, Atkinson and Bourguignon (1982), Keen (1990) and Lambert (1993)) – would be any other social cardinalisation of preferences.

The aggregate revenue requirement R is set at zero, so that the sole purpose of taxation is redistributive.[9] We take $\theta = 0.7$, so that 30 % of the population belongs to the poor group. The computational methods used, which are somewhat unfamiliar, are described in an Appendix.

Table 1 reports key aspects of the comparison in these two cases between the optimal tax scheme when the two groups cannot be distinguished (so that a common schedule must be applied to both) and the optimal scheme when categorical information is

[9]Similar qualitative results to those reported below are obtained if R is instead set at low but not implausible levels of around 10 per cent of aggregate income. Higher revenue requirements are liable to give a rather different pattern of optimal schedules, for reasons that will become apparent – and indeed a central concern – shortly.

Table 1. Effects of the optimal use of categorical information

	Case A	Case B
Welfare gain	4%	15%
Transfer to group 2	130%	60%

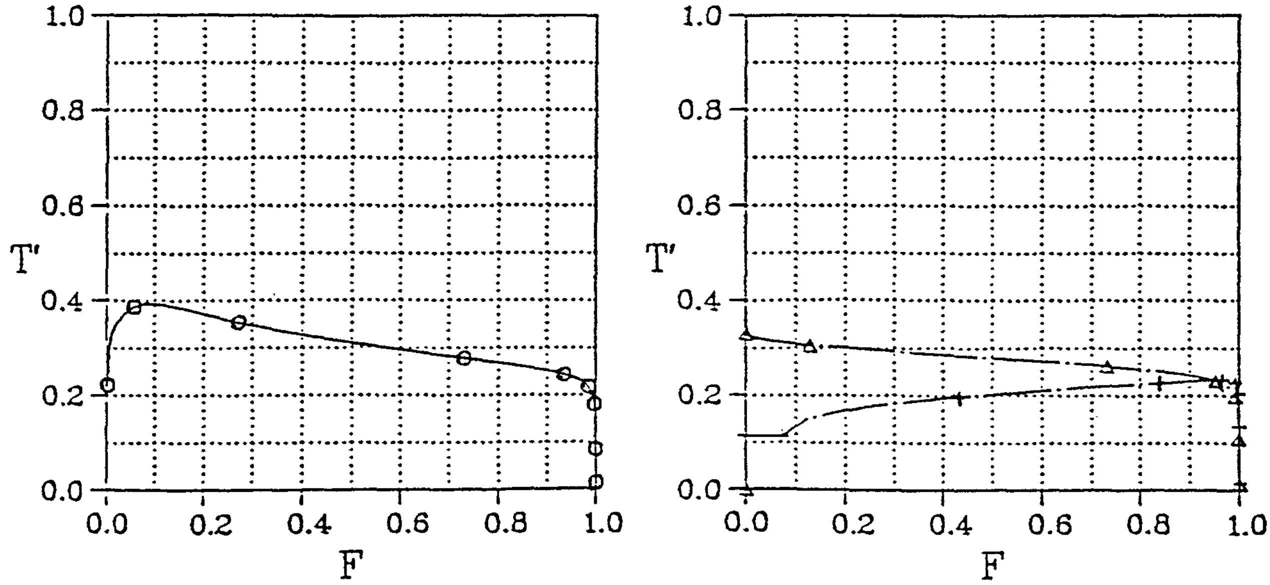
Notes: Cases A and B have the common structure described in the text. The difference is that in Case A $\rho_1 = \rho_2 = 0.5$ while in Case B $\rho_1 = 0.5$ and $\rho_2 = 0.15$.

used to apply separate schedules to them. Maximised social welfare, it is clear, can be no lower in the latter situation than in the former: for it remains an option, when categorical information can be used, nevertheless to apply a common schedule to both groups. What is much less obvious – the central quantitative question raised in the Introduction – is how sizeable the gains from introducing a categorical element in tax design are liable to be. To assess this, we follow Stern (1976) in taking as a measure of the social welfare attained under some tax scheme the level of consumption which, if equally distributed at zero hours, would generate that same level of social welfare; with distinct schedules applied to the groups, this equivalent consumption is calculated separately for each group, weighted by their populations and then aggregated over the two. The welfare gain reported in the first row of Table 1 is the proportional increase in equivalent consumption in moving from the optimal single schedule to the optimal group-specific schedules. The gain in the benchmark case is modest, but far from negligible. In Case B the gain is clearly very substantial:[10] aggregate equivalent consumption rises by 15%. The impression that emerges is thus one of potentially sizeable gains from the optimal use of categorical information.

Turning to the qualitative issue (and puzzle) raised in Section 2, Figures 2 and 3 show the patterns of optimal marginal tax rates in Cases A and B respectively. Both the optimal common schedule and the optimal group-specific schedule for the richer group 1 display the familiar feature that, through most of the income range, the marginal

[10] The common schedule for this case is calculated with $\rho_1 = \rho_2 = 0.4$, reflecting the difficulty of solving the Mirrlees problem with heterogeneous individuals.

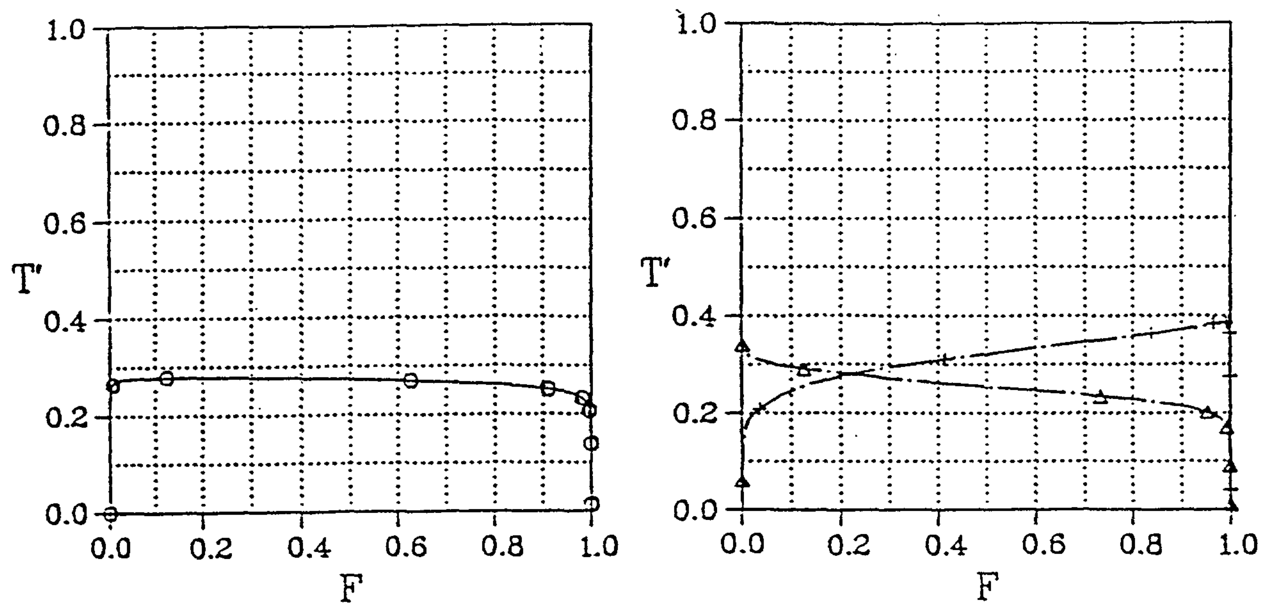
Figure 2: Optimal marginal tax rates in Case A



Key:

o One tax schedule
△ Group 1
+ Group 2

Figure 3: Optimal marginal tax rates in Case B



Key: As Figure 2.

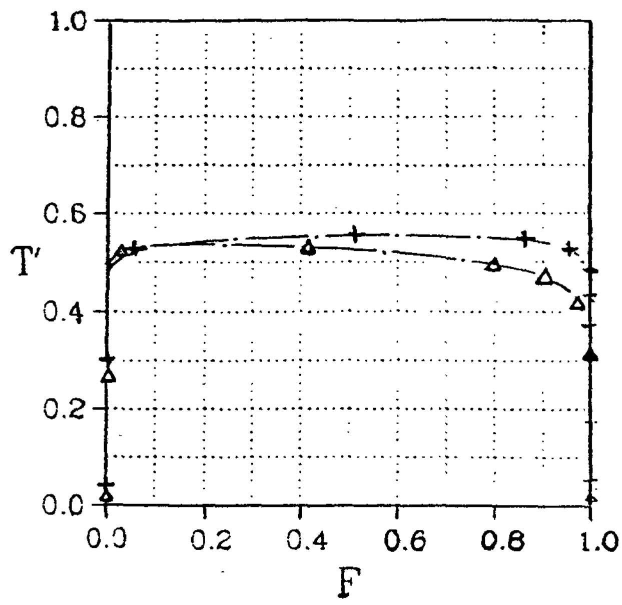
tax rate decreases with income. For the poorer group 2, in contrast, the striking finding emerges that the marginal tax rate optimally increases with income. Figure 4 shows that the same qualitative difference between the group-specific schedules also emerges if preferences are of the form

$$u^i(x,y) = -(1-\rho_i)(1/x) - \rho_i(1/(1-y)) \tag{9}$$

(with $\rho_1 = 0.4$ and $\rho_2 = 0.15$) implying an elasticity of substitution between consumption and leisure for both groups of 0.5 (rather than unity, as in cases A and B). The contrast between the two group-specific schedules is less sharp in this Case C, reflecting the general tendency for optimal marginal rate schedules to become flatter as labour supply becomes less responsive. The important point, however, is that the same general pattern emerges from all three figures, as indeed it did from all simulations performed. Thus it is the DKM intuition described in Section 2 above that is vindicated by the simulations, not the conventional wisdom from the previous optimal tax literature.

It is possible, however, to reconcile the two. The key lies in the revenue requirement. In the simulations of the one-dimensional Mirrlees problem from which the conventional wisdom of optimally decreasing marginal rates has emerged, the revenue requirement is usually set at levels taken to correspond broadly to observed levels of national expenditure on public goods: between, say 10, and 30 per cent of aggregate gross income. The optimal group-specific revenue requirements for the two-dimensional problem, however, reflect optimal inter-group redistribution and so can be of an entirely different order of magnitude and indeed sign: much higher for richer groups; much lower, and quite possibly negative, for poorer. The simulations do indeed point to very substantial inter-group transfers. This can be seen from the second row of Table 1, which reports the amount optimally transferred to the poorer group 2 (expressed as a proportion of that group's collective gross income): in Case A, for instance, the revenue requirement optimally imposed on group 2 is a subsidy

Figure 4: Optimal marginal tax rates in Case C



Key: As Figure 2.

Table 2. Optimal marginal tax rates and the revenue requirement

F(n)	R			
	-100	-50	0	+50
0.10	12.0	15.6	23	44
0.50	13.4	15.9	21	34
0.90	13.4	15.1	19	25
0.99	11.7	12.6	13	15
F(n*)	0.85	0.65	0.0	0.0

Notes: The revenue requirement is specified as a proportion of aggregate gross income. Preferences are as in Case A; the pre-tax wage is distributed lognormally with α = -1 and σ = 0.4. In the final row, n^* is the ability level at which the highest marginal tax rate occurs.

equal to 130 per cent of their gross income. This is financed by extracting from the richer group about 18 per cent of their gross income (the overall revenue requirement, recall, being zero).

The suspicion these observations create that the qualitative pattern of marginal tax rates is altered by substantial – but, in the context of categorical taxation, perfectly plausible – variation in the revenue requirement is confirmed by Table 2, which shows how the pattern of optimal marginal tax rates for the one-dimensional Mirrlees problem (simulated under standard assumptions) changes as the revenue requirement is decreased, from one in which, over most of the income range, marginal rates decrease with income to one in which they increase. With a revenue requirement of zero, for example, the marginal tax rate is everywhere decreasing; with a revenue requirement of -100% of gross income, on the other hand, it increases with income up to the 85th percentile.

Some sense of the analytical basis of these numerical results can be derived from the discussion in Section 2. It is clear from (3) that the variation of the optimal marginal tax rate T' with level of income is a complex matter. One consideration, however, is the variation of $\mu(\cdot)$ with n. It is straightforward to show that $\mu(n)$ starts and finishes with a value of zero (the transversality condition) and has an inverse-

U shape in-between. Intuitively, $\mu(n)$ measures the social welfare gain from slightly increasing the marginal tax rate at n and distributing as a poll subsidy to those below n the revenue raised from the consequent increase in average tax rates above n.[11] At low levels of n, this gain tends to increase with n: for the hypothetical reform then benefits more of those whom the government would like to help, in the sense that their marginal social utility of income ($G'u_x$) exceeds the marginal social cost of public funds (λ). Over this range (and putting aside the efficiency losses that also enter the calculation in (3)), the optimal marginal tax rate thus increases with income. But beyond the ability level n^* at which

$$G'(u(n^*))u_x[x(n^*), y(n^*)] = \lambda \,, \tag{10}$$

the hypothetical reform starts to benefit those who are undeserving, in the sense that $G'u_x \leq \lambda$; the case for a high marginal tax rate now becomes progressively weaker as n continues to increase. Some clue to the role found for the revenue requirement can then be seen in (9). When the revenue requirement is low, so too is λ. Since the left hand side of (9) is readily shown to be decreasing in n^*, a reduction in the revenue requirement can thus be expected (informally) to shift the point at which $\mu(n)$ peaks to the right. Thus $\mu(n)$ will continue to increase – and, to that extent, the optimal marginal tax rate also tend to increase – further into the distribution than would

[11] To see this, rewrite (3) as

$$-T'(\lambda/u_x s_n)f(n) = \mu(n), \tag{A.1}$$

and, for simplicity, suppose preferences to be of the form $u(x,y) = x + v(y)$, so that (5) becomes

$$\mu(n) = \int_{\underline{n}}^{n} G' f dn - \lambda F(n). \tag{A.2}$$

The left hand side of (A.1) can be thought of, loosely, as the marginal deadweight loss from slightly increasing the marginal tax rate at z, which is to be equated to (loosely speaking, again) the distributional gain from doing so, measured by μ. In (A.2), this distributional advantage is shown to be the difference between the social welfare gain from a unit poll subsidy to ability levels lower than n and the social cost of financing that subsidy by increasing the marginal tax rate at z: for since the subsidy costs $F(n)$, and since at the optimum, raising the marginal tax rate at z is as good a way of financing it as any, that social cost is just $\lambda F(n)$.

otherwise be the case.

Put crudely, the intuition is thus that the lower is the revenue requirement the more the government can afford to support the poor by a generous poll subsidy, recouping at least part of this by a pattern of rising marginal rates on the better-off. The simulations show that it takes revenue requirements far lower than those which have usually been considered for this effect to dominate. But the orders of magnitude required are plausible ones for beneficiary groups of categorical schemes. Note too that this emphasis on the role of the revenue requirement is consistent with the important role played in the DKM intuition – as described in Section 2 above – of the desire to make inter-group transfers to the poorest of the poor.

There are of course a variety of further issues in the optimal use of categorical and income information that can be explored by simulation, concerning, for instance, the sensitivity of the optimal inter-group transfer to the nature and magnitude of the differences between the groups. While we do not pursue these here, it should be emphasised that the findings stressed above recur in all simulations we have performed. Optimal marginal tax rates continue to increase with income in the poorer group if one takes lower values of ρ_1, for instance, or if one introduces inequality aversion by taking $G(u) = -e^{-u}$. That conclusion, in particular, appears robust.

4 Summary and conclusions

This paper has sought to extend understanding of the optimal design of tax/transfer schemes when both categorical and income information are fully utilised. This has been formalised as a conceptually straightforward two-dimensional non-linear income tax problem. As is typically the case with optimal non-linear taxation, few general analytical results are available. Simulations, however, point to two important features of the solution. First, the gains from the use of categorical information can quite plausibly be significant. Second, and more strikingly, the qualitative pattern of the optimal group-specific schedule is likely to be entirely different across groups: for the richer group, marginal tax rates optimally decrease with income; for the poorer, on the other hand, they optimally increase with income (over most of the income range, at least). This latter stylised fact is one of some potential importance in evaluating and designing tax/transfer schemes. It tends to confirm the earlier informal argument to the same effect of DKM, an argument that runs counter to the previous conventional wisdom that optimal marginal tax rates decline with income. The reconciliation, we have argued, lies in the role played in optimal tax design by the revenue requirement: inter-group transfers are liable to imply group-specific revenue requirements for poorer groups that are very much lower than simulations have conventionally considered, and optimal tax design in such circumstances is liable to involve (very loosely speaking) a generous poll subsidy recovered from the better-off by a pattern of increasing marginal tax rates over much of the income range.

Appendix: The computational procedure

The optimal control problem was solved by the FORTRAN program MISER3, described in detail in Jennings *et al* (1990). This program has been developed to solve a general class of optimal control problems with constraints. The constraints are allowed to be of equality as well as inequality type. The program is based on the concept of the control parameterisation technique. The main idea of this method is to convert the optimal control problem into a sequence of approximate finite-dimensional mathematical programming problems. The problem is reduced to the mathematical programming problem by approximating each component of control by a piecewise control function on $[\underline{n}, \overline{n}]$. Thus a grid is chosen and the values of the control between two grid points (control parameters) are the variables on the nonlinear programming problem. The problem is then solved using e.g. SQP-method.

We chose the range of n such that $\underline{n}$ = 0.05 and $\overline{n}$ = 2.2 when α= -1.0 , -0.85 or -0.8 and σ = 0.5 and $\underline{n}$ = 0.04, $\overline{n}$ = 1.6 when α = -1.4, -1.0 or -0.8 and σ = 0.4, 0.45 or 0.55. This implies that the integrated value of $f(n)$ from $\underline{n}$ to $\overline{n}$ is more than 0.999 in each case. A majority of individuals were distributed to the interval $[\underline{n}, 0.9]$ (i.e. frequencies $f_i(n)$ are higher) so we used a more dense grid in that interval. In order to improve the numerical accuracy, we weighted the objective (social welfare) function (1) and the constraint (2) by some positive constant which was chosen so that in different cases the order of the absolute value of the objective function was 100 in a solution point.

We used piecewise constant approximation for controls y_i, i =1,2 and this implies piecewise linear approximation for the state functions u^i, i = 1,2. Thus the control functions y_i, i = 1,2, are discontinuous and this gives piecewise continuous approximation for consumption x_i, gross income z_i, the tax rates T_i and the marginal rates T_i', i = 1,2. However, in Figures 2-4 we used piecewise linear approximation for them by calculating $x_i(n)$, $z_i(n)$, $T_i(z_i(n))$ and $T_i'(z_i(n)), i$ = 1,2, at the points $n = \underline{n}, n = \overline{n}$ and at the midpoints of each subinterval $[n_i, n_{i+1}], i = 1, ..., n_p - 2$, from the solution obtained by MISER3 using the discretisation described above, and approximating them linearly between those points (n_p being the number of control subintervals). We solved the problem using $n_p = 10$, 20 and finally $n_p = 40$. Using different initial guesses the problem converged to the same optimum point. The optimization convergence criterion that we used was 10^{-7} for constraints and 10^{-5} for a zero of gradient. In numerical computations we used the bound constraint $y_i(n) \leq 0.99 \; \forall \; n \in [\underline{n}, \overline{n}]$. When the utility function was defined as in (9) the optimal control problem proved to be difficult to solve since $u^i(n), i = 1,2$, can change rapidly and due to the discontinuities of constraints. In this case we used additional constraint $0 \leq x_i(n) \leq 2.2 \;\; \forall n \in [\underline{n}, \overline{n}]$ that $y_1(n) \leq 0.8 \;\; \forall n \in [\underline{n}, 0.8]$, and we required that $y_1(n) \leq 0.9 \;\; \forall n \in [0.8, \overline{n}]$ and $y_2(n) \leq 0.9 \;\; \forall n \in [\underline{n}, \overline{n}]$. In minimum point these constraints were inactive.

References

Akerlof, G.A. (1978), 'The economics of "tagging" as applied to the optimal income tax, welfare programs and manpower planning', *American Economic Review* **68**, 8-19.

Atkinson, A.B. (1992), 'On targeting social security: the experience of western countries with family benefits', World Bank Conference on Public Expenditure and the Poor: Incidence and Targeting, Washington D.C.

Atkinson, A.B. and F. Bourguignon (1982), 'The comparison of multi-dimensioned distributions of economic status', *Review of Economic Studies* **49**, 183-201.

Besley, T. (1990), 'Means testing versus universal provision in poverty alleviation programmes', *Economica* **57**, 119-129.

Cowell, F. (1986), 'Welfare benefits and the economics of take-up', ESRC Programme on Taxation, Incentives and the Distribution of Income Working Paper No. 89, London School of Economics.

Dilnot, A.W., Kay, J. and C.M. Morris (1984), *The Reform of Social Security* (Clarendon Press, Oxford).

Jennings, L.S., M.E. Fisher, K.L. Teo and C.J. Goh (1990), *MISER3, Optimal Control Software: The Theory and User Manual* (EMCOSS pty Ltd, Australia).

Kanbur, R. (1987), 'Transfers, targeting and poverty', *Economic Policy* **4**, 112-136 and 141-147.

Kanbur, R. and M. J. Keen (1989) 'Poverty, incentives and linear income taxation'. In A.Dilnot and I.Walker (eds) *The Economics of Social Security* (Oxford University Press, Oxford).

Kanbur, R., Keen, M.J. and M.Tuomala (1994). 'Optimal non-linear income taxation for the alleviation of income poverty', *European Economic Review* **38**, 1613-1632.

Keen, M.J. (1992), 'Needs and targeting', *Economic Journal* **102**, 670-679.

Lambert, P.J. (1993), 'Inequality reduction through the income tax', *Economica* **60**, 357-365.

Lydall, H.F. (1968). *The Structure of Earnings* (Clarendon Press, Oxford).

Mirrlees, J.A. (1971), 'An exploration in the theory of optimum income taxation', *Review of Economic Studies* **38**, 175-208.

Stern, N. (1976), 'On the specification of optimum income taxation', *Journal of Public Economics* **6**, 123-162.

Stern, N. (1982), 'Optimum taxation with errors in administration', *Journal of Public Economics* **17**, 181-211.

Stiglitz, J. E. (1982). 'Self-selection and Pareto-efficient taxation.' *Journal of Public Economics* **17**, 213-240.

Tuomala, M. (1990), *Optimal Income Tax and Redistribution* (Clarendon Press, Oxford).