N° 8118

Nonlinear Pricing in a Finite Economy

by

*Roger Guesnerie and Jesus Seade**

February 1981

* CEPREMAP, and University of Warwick/El Colegio de Mexico, respectively. Work on this problem was started while Seade was a visitor at CEPREMAP, spring-summer 1979. This author gratefully acknowledges CEPREMAP's hospitality as well as a research grant from the SSRC (UK) and CNRS (France) which made this research possible. We are grateful for comments from members of seminars at Duke and Warwick universities, and the First Latin-American Meeting of the Econometric Society (Buenos Aires, July 1980).

1. *Introduction*

Models of optimal nonlinear taxation or pricing adopt, as a matter of routine, a population forming a continuum. The reason usually advanced for this is a practical one: namely that the optimal schedule would otherwise be too complex and unappealing, presenting each consumer with a tailor-made kink (a sudden increase of the marginal tax) at the point he is programmed to select in the socially optimal equilibrium.

This essentially pragmatic justification of the continuum-approach seems appropriate for applications, but there is nonetheless a clear theoretical case for considering the finite economy. A variety of reasons can be given for this. The most obvious one is that the continuum-approach can only apply when we consider markets or economies with a large number of agents, and the discrete approach fills a gap and has an interest of its own. Certain problems do have consumers which are few and large: such is the case with buyers of large yachts and jewels (price discrimination is only a form of nonlinear pricing) as well as in the upper tails of the income or wealth tax scales. Furthermore even in problems with many agents it may well be that the discrete formulation is the most appropiate or convenient, as we shall presently suggest.

Secondly, the finite economy should be considered for the sake of completeness of the study of (nonlinear-) pricing problems generally, in a different sense from that in the previous paragraph. Even in cases where the continuum-formulation looks *a priori* entirely satisfactory, a full understanding of its results requires, for instance, the study of convergence of the finite case onto the continuous one as the population grows dense. Properties obtained "in the limit" have to be confronted with "limit properties". Furthermore, it may be the case that questions which have not been fully elucidated in the continuous case (such as the incidence of bunching or of discontinuities) can be shed light upon through the study of discrete economies, with the easier access to direct arguments it permits. This takes us to the last and perhaps main justification of the finite approach that we offer.

On account of the two-tier maximization that characterizes it, the nonlinear-tax problem is difficult to grasp, to fully understand in its internal mechanics. The discrete approach provides us with a rather different perspective on how the problem works, and in particular brings to the fore more explicitly the interaction between agents' and principal's choices. In this way this formulation may well be a necessary or at least a helpful step towards the solution of a wider class of nonlinear-pricing problems than have been studied in the literature. In particular, as already mentioned, it can be expected to shed light on the incidence or behaviour of discontinuities or other pathologies of solutions in the continuous case, possibilities on which little is known. Also, it should help to clarify the nature, strength, and purpose of the various assumptions that make the problem workable in the continuum case, providing a different line of attack for their relaxation. In a broader perspective this may be a useful alternative tack to try to understand and study the multidimensional case. In sum, it seems reasonable to expect this fairly different approach to the problem to yield a different set of results, some new, while providing a theoretical underpinning of the standard model.

Clearly, we do not aim in this paper to cover the whole programme suggested by this defence of the discrete approach. Our more modest intention is to provide a basic framework for the discussion of the finite case and to conduct an analysis which, while deriving some results for the general form of the problem, focuses on a detailed exploration of the (discretized) standard model under familiar assumptions. Some of the known main properties of continuous optima are derived (or in one case shown not to hold) by direct arguments, while at the same time we seek to develop a feel as to how the present problem behaves.

The following section introduces the model and some definitions, and sections 3 and 4 discuss some features of solutions at two different levels of generality. We restrict attention to one-tax situations, i.e. two goods, and this is not "for convenience": the restriction is essential. The analysis has been made rigorous most of the time, but the exposition is somewhat informal and the argument largely geometric: it can be undestood intuitively without going into the details. Some concluding remarks are given in section 5.

2. *Model and Definitions*

We consider an economy with two goods. A consumption bundle is denoted $x = (a, b)$. Consumers (households) are indexed by $h = 0, \ldots H$ and consumer preferences are represented by utility functions $u^h(a, b)$.

We take u^h to be defined and continuous on $\mathbb{R}^2_+$, diferentiable and strictly concave in $\mathring{\mathbb{R}}^2_+$, increasing in a and decreasing in b, and to have $u_a \to \infty$ as $a \to 0$, $u_a \to 0$ as $a \to \infty$, $u_b \to q \leqslant 0$ as $b \to 0$, $u_b \to -\infty$ as $b \to \infty$.

We shall for simplicity assume linear technology, permitting the transformation of one unit of one good into one unit of the other. Depending on the specific problem one has in mind, several interpretations can be given to these commodities. In the income tax version of the model, a is the amount of consumption and b the labour supply in "efficiency" units, i.e. earned income; in a pricing context a can be regarded as for example consumption of electricity and b as the required payment, i.e. outlay (loss) of numeraire.

A central feature of this model is that all consumers' net trades with the market are constrained by the same budget set τ, in (a, b) - space (τ is a subset of $\mathbb{R}^2_+$). This covers a variety of situations. In an income-tax problem τ is determined by the tax schedule relating income before and after tax; what one neeeds for this, is the existence of underlying homogeneous efficiency units of labour, i.e. homogeneous before-tax income. Similarly, in a pricing problem, τ reflects the (nonlinear) pricing rule[1].

In this model, a feasible *allocation* consists of a sequence (x^h) of consumption bundles $x^h = (a^h, b^h)$ meeting (i) and (ii):

(i) Feasibility constraint: $\sum_h a^h \leqslant \sum_h b^h + K$;

(ii) Incentive constraint: $x^h \max u^h(x)$, $x \in \tau$.

The first of these is a linearized (at and around the relevant equilibrium) production constraint, which we shall often think of as a profits constraint for the firm in a pricing interpretation or as a budget-balance constraint for the government. We say that bundle (a, b) is *cheaper* (or *less costly*) than (a', b') for the principal, if $a - b < a' - b'$; hence the former bundle represents a greater contribution to profits or tax revenue than the latter. Diagrammatically, "cheaper" means lying on a 45°-line below/to the right of the other point's.

We will consider in this paper the problem of a principal who controls the shape of the budget set τ seeking to optimize relative to his objectives which we come back to below. Now we argue first that there is no loss of generality if we restrict the principal to choose a budget set τ delimited by an increasing step-function, as in figure 1. To show this, we first note that any configuration of actual choices (x^h) by the various agents in (a, b)-space necessarily consists of a finite set of *corners* $\mathcal{C} = \{C_o, \ldots, C_T\}$, which according to our monotonicity assumption $(u_a, -u_b) \gg 0$ must fall in a north-east/south-west direction relative to one another. These points can therefore be

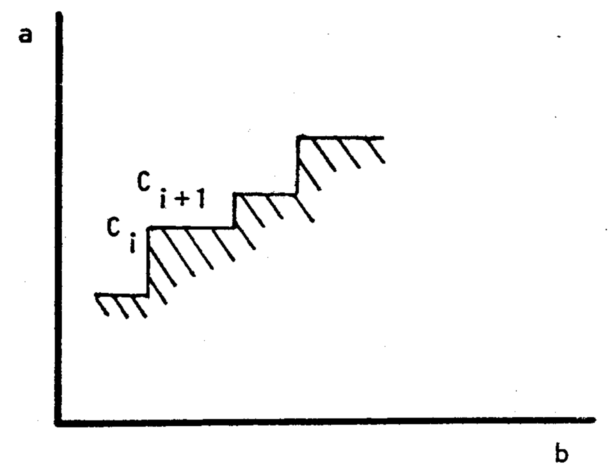

Figure 1

unambiguously arranged and indexed from left to right (SW to NE), from the "bottom" to the "top" of observed demands: $C_0 \ll C_1 \ll \ldots \ll C_T$. Secondly, it is obvious that the choices of the agents will not be changed when whichever schedule giving rise to the above set $\mathcal{C}$ is replaced by the schedule associated with the step function constructed from the C_i's in the obvious manner. Hence, as announced, any result obtained with a general budget set can also be obtained with a step-function-defined set $\{x = (a, b) \mid (a,-b) \leq (a_i,-b_i)$, for some $(a_i,b_i) \in \mathcal{C}\}$. We can thus simply identify the budget set τ with the set $\mathcal{C}$ of existing corners.

Let us now consider a schedule, characterized by the finite set of corners $\mathcal{C} = \{C_o, \ldots, C_T\}$. We shall need a few definitions, aimed at describing the *qualitative features* of the tax schedule and of the associated equilibrium.

Let H_i denote the set of h's selecting the point (or "corner") C_i; that is, $H_i = \{h \mid x^h = C_i\}$. We note here that the choice by h of x^h amongst $\{C_i\}$ need not (in fact will in the optimum normally not) arise as a unique individual optimum, and we allow the principal, by convention, to choose which point from his optimal set a consumer is given.

An allocation (x^h) has *bunching* at C_i if $C_i = x^h = x^{h'}$ for some $h \neq h'$, i.e. if $|\text{card } H_i| > 1$.

Now at the centre of the problem lie individual incentives --the question of how do agents respond to small perturbations of the budget set described by $\mathcal{C}$. These responses in turn depend on the way the different corners $C_i \in \mathcal{C}$, and points in their neighbourhoods, relate to each other in the preferences of consumers involved. Some terminology is required for this analysis.

We say that C_i is incentives-free, or simply *free*, if any sufficiently small variation of its position alone does not lead any h to "jump" to or from any other C_j, i.e. if the induced change in the allocation (x^h) is proportionately small. Formally, C_i is *free* iff $\exists$ a neighbourhood $V(C_i)$ such that if C_i is replaced, as i^{th} corner, by any $C'_i \in V(C_i)$, then H_j does not change, $\forall j$.

Now thinking of the second good a (say consumption) as drawn on the vertical axis, we say that C_i is *upward-* (resp. *downward-*) *free* if the above holds for (at least) increases (resp. decreases) in a alone. Formally, the definition of an upward- (resp. downward-) free corner is similar to the definition of a free corner where the neighbourhood $V(C_i)$ is replaced by an open interval of the vertical axis $](a_i, b_i), (a_i + \varepsilon, b_i)[$, $\varepsilon > 0$ (resp. $\varepsilon < 0$).

Two corners C_i and $C_j^{(2)}$ are *linked together*, or either is linked *to* the other, if they both belong to the optimal set of some agent h, i.e. if there is an indifference curve of h which passes through the two points and is the highest h can reach on $\mathcal{C}$. We way that the said consumer, h, *links* C_i and C_j.

C_i is linked if it is linked to some C_j and is *loose* otherwise.

Now C_i is said to be W-*linked* (W for winner) if a consumer h, linking C_i to some C_j, is allocated C_i, i.e. $x^h = C_i$. (C_i is then W-linked *to* C_j). Equivalently, C_i is W-linked if some agent $h \in H_i$ links C_i to some C_j.

Similarly, C_i is L-*linked* (L for loser) if there exists some consumer h linking C_i to some C_j which is allocated $x^h = C_j$. Equivalently C_i is L-linked if an agent $h \in H_j$ (recall $j \neq i$) links C_i and C_j. (C_i is then L-linked *to* C_j).

It should be clear that the properties of being W-linked and L-linked are not mutually exclusive: a corner could have both.

Lastly, the *relevant* indifference curve of a consumer is that which he/she attains in a given equilibrium.

Let us finally turn to the principal's objective function which we only assume, initially (sec. 3), to be a well-behaved Paretian Social Wefare Function (SWF), i.e. one defined on and increasing in individual utilities. We seek to determine the kinds of properties that this requirement alone imposes on optimal schedules. Later on (sec. 4) we further adopt a common additional requirement, that redistribution in one or other direction along the schedule be unambiguously desirable. The essential features of this problem would not be affected if other (or more general) objectives were considered, such as profit maximization: our discussion and most results below still apply. These are ruled out for expositional convenience.

We shall primarily be interested in features of the second-best optimum for the problem, or its *optimum* for short, i.e. the full problem with both production feasibility and decentralization of consumer demands acting as constraints. We will however not perform the optimization explicitly, but instead derive directly properties of the optimum. We shall at some points also refer to the *first best*, which as usual means the principal's optimum subject to technological feasibility alone.

It will simplify exposition, allowing us to avoid certain taxonomic forms of results, if we adopt the following minimal regularity assumption. We say, by definition, that consumers h' and h'' are *locally different* at x if $\nabla u^{h'}(x)$ and $\nabla u^{h''}(x)$ are not colinear, or equivalently if their marginal rates of substitution $s^h(x) \equiv - u_b^h(x)/u_a^h(x)$ (for $h = h', h''$) between the two commodities are not equal at this point. We assume

Assumption A: All consumers $h \in H_i$ *are locally different from one another at* $C_i \in \mathcal{C}$, $\forall i$, *in the optimum.*

Let us remark that we do not require that consumers be locally different in any possible allocation, which would be rather restrictive (although implied by our further assumption B), but only around their optimal points of demand. Weaker forms of the assumption could be given, but this seems unnecessary, for the assumption can be shown to be generic as it stands, in the sense that it holds for "most" configurations of preferences (of agents and objectives of the principal). Hence it is a reasonable simplification to make at the outset; we take (A) to hold throughout the paper, without further mention.

3. *Some Observations for the General Paretian Case*

In the absence of sufficiently strong assumptions on the nature of individual preferences, the distribution of consumers "along" the corners of the budget set can be very varied and complex, which complicates considerably the analysis. The difficulty lies in that without a natural, given arrangement of consumers on the line, it becomes unclear both what the government should *like* to do (how does deservingness vary with income, say) and more critically what it *can* do, given the complicated nature incentives can then take. Putting these difficulties aside seems to be the main technical contribution of the usual unidimensionality assumption in continuous models, which we essentially adopt (assumption B) in section 4. But some observations can be made before we specialize.

Lemma 1 : *A point* $C_i \in \mathbf{C}$ *is free iff it is loose; it is upward-free iff it is not L-linked; and it is downward-free iff it is not W-linked.*

Proof : Straightforward from the definitions and left to the reader.||

Whether production efficiency is generally desirable in this economy is a basic question upon which most proofs below depend, and one whose answer seems not *a priori* clear to us. The type of argument used by Diamond and Mirrlees (1971) in the (optimal commodity uniform-pricing case, basically transferring a small amount of a desirable commodity to everyone through a suitable small change in prices, does not apply here, for changes in individual demands in response to infinitesimal stimuli will often be finite, depending on the structure of links which in general can be very complicated. This discreteness in responses may well suddenly take us from within the interior of the production set to the outside of this set. But the problem admits of a simple alternative derivation of the result:

Proposition 1 : *Second-best optimality requires production efficiency.*

Proof : Consider the set of relevant (optimal for some h) indifference curves in a given supposedly optimal equilibrium and consider rhe envelope of these curves. Since we have a finite population (and given our assumptions on the functions $u^h(\cdot)$), there exists a *cheapest* point on this envelope, i.e. one lying on a southeasternmost 45°-line. This point is moreover unique by strict concavity, and clearly lies on a differentiable segment of the envelope, corresponding to a segment of some consumer's relevant indifference curve, say that belonging to h'. We now create a new corner denoted C_* at this point. Movements to C_* by any h cannot but (weakly) increase utilities and (weakly) lower costs, so that the supposed optimality of the original allocation is preserved, and we actually set $x^h = C_*$ for any h whose utility is not thereby lowered, which includes h'. Consequently, C_* is a point observed in demands, and is not L-linked. Finally, by L1, consumption and utility of the (nonempty) set H_* of consumers at C_* can be increased, using up any production-slack that there may be, without inducing any other $h \notin H_*$ to change his demands. This would contradict Paretian optimality.||

L-links were referred to in the above argument, but the key rôle they play in the problem comes out more clearly in the following:

<u>*Proposition 2*</u> : *If, in the optimum,* C_i *is not L-linked, then (i) there is no bunching at that point and (ii) the marginal rate of substitution of the "local" consumer is one:* $s^h(C_i) = 1$ for $h \in H_i$.

<u>*Proof*</u> : Suppose $s^{\bar{h}}(x^{\bar{h}}) \neq 1$, for some $\bar{h} \in H_i$, at the corner C_i. Recalling that $s^h(x)$ is measured by the slope of the (differentiable) utility function in x, the set $D(\bar{C}_i) = \{(a,b) \mid u^{\bar{h}}(a,b) > u^{\bar{h}}(\bar{a}_i,\bar{b}_i)\} \cap \{(a,b) \mid a-b = \bar{a}_i-\bar{b}_i\}$ is convex and non empty.

Let us now introduce a new (additional) corner in the budget set, somewhere in $D(\overline{C}_i) \cap V(\overline{C}_i)$ where $V(\overline{C}_i)$ is some small-enough neighbourhood of $\overline{C}_i$. Since $\overline{C}_i$ is not L-linked, this change will induce small shifts, only of agents $h \in H_i$ and actually of $\overline{h}$, and the choices of the agents will remain in $V(\overline{C}_i)$. None of their utilities falls, u^h rises, and all the movements are feasible. The assumption that we are at a (constrained) Pareto optimum is thus contradicted. Part (ii) of the proposition follows directly; so does part (i) given assumption A.||

Figure 2 illustrates the argument.

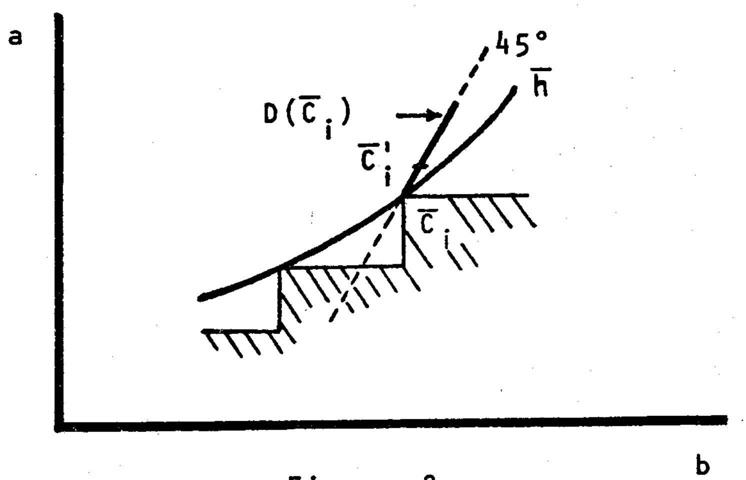

Figure 2

Notice that statements on the value of s^h (as compared with the price ratio, 1) are statements on the discrete counterpart of the *marginal tax* (or mark-up) --the optimal wedge to be put between consumers' and producers' effective prices. From P_2, a distortion at C_i (i.e. $s(x) \neq 1$) can only arise if C_i is L-linked, or equivalently these links emerge as the barriers towards further redistribution which movements towards the first best would require.

It is natural to expect P_2 to actually apply at some points --i.e. that in the optimum there will always be some C_i or C_i's which are not L-linked (upward-free, by L1). This need not *a priori* be the case, for in principle every point could be L-linked to some other, as in fig. 3 (where $h_0 \in H_0$ and $h_1 \in H_1$). The next proposition says that this type of cyclicity on $\mathcal{C}$ can never be optimal, and the corollary following stresses an implication this has.

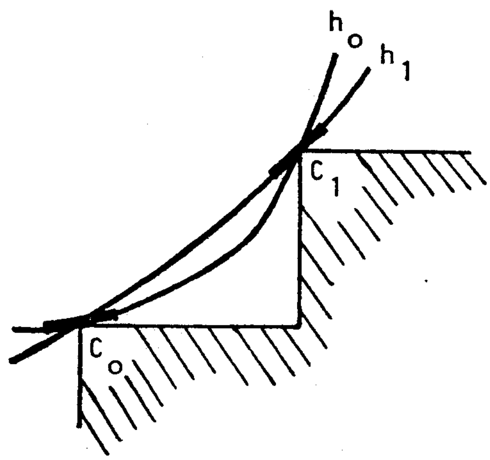

Figure 3

Proposition 3 : *Not all corners can, in the optimum, be L-linked.*

This almost follows from the argument used in the proof of P1 above, but not quite: that argument only said that the optimum can always take the form of an allocation with a C_* that has no L-link, but not the converse, that the absence of such a point imply strict sub-optimality. We thus derive this result separately.

Proof (of P3) : If C_i is L-linked to C_j, it must be the case that the cost of C_j, measured with the production prices (1,1), is not greater than the production cost of C_i : otherwise, it would be possible for the principal to move the consumption bundle of the relevant consumer from C_j to C_i, changing no utilities nor other demands but reducing total costs, i.e. generating a slack in production, which by P1 cannot be optimal.

Consider now the cheapest (at the going prices, (1,1)) element(s) of $\mathcal{C}$. If there is a unique cheapest element, it can be L-linked to no other point, by the argument above.

On the other hand, if there is a set $\mathcal{C}_m = \{C_m, C_n, \ldots\}$ of several cost-minimizing corners, we have to prove that they cannot be L-linked to one another. Let C_m be L-linked to a point or points $\{C_{m'}\} \subset \mathcal{C}_m$. Move to C_m the person or persons who are indifferent between that point and each $C_n \in \{C_{m'}\}$. This leaves all utilities *and* production costs unchanged, while removing all of C_m's L-links (which become W-links). There is then bunching on, and no L-links of, C_m, which by P2 cannot be optimal.||

<u>*Corollary 3.1*</u> : *There is always, in the optimum, a point* C_* *with (no bunching and) no distortion :* $s^{h^*}(x^{h^*}) = 1$ *for the local consumer* h^*.

The next observation is fairly evident, but merits explicit mention for an implication it has on the nature of discrete optima:

Proposition 4 : *If the second-best Pareto optimum that we are considering is not the first best, there is at least one link in it.*

Proof : Otherwise, if all corners C_i were loose, they would be free by L1, with bunching excluded by P2. The entire allocation (x^h) could then be varied locally without any constraint other than the technological one. Hence the equilibrium would be a local first-best Pareto optimum, and as the first-best optimization problem is convex, a global one.||

Corollary 4.1 : *Within the class of problems where the first-best cannot be reached, no linear schedule can be (second-best) optimal.*

Proof : Immediate from P4 and the fact that, under a linear schedule (i.e. one in which the C_i's follow $a_i = \alpha + \beta b_i$) and with a finite number of people, there can be no links, by strict quasi-concavity of preferences.||

The proviso excluding the attainment of the first-best is needed because, without further specification of the principal's objectives, any production-efficient decentralized allocation could in principle be fully optimal. Under the additional assumptions of section 4 this proviso becomes redundant. This is essentially a restatement (for the discrete case and with less assumptions) of results in Seade (1977), where it is shown by construction that any schedule with a non-zero distortion at the top (as any linear one with $\beta \neq 1$ will have) can be Pareto-improved upon, and in Willig (1978), where it is shown, again by construction, that any linear pricing schedule is Pareto-inefficient.

Let us now refer to the 2T + 2 coordinates of the corners C_i as the *direct controls* of the principal, a terminology suggested by the income-tax interpretation of the problem.(3) A change in a_i alone, for example, can be thought of as a change in the marginal tax somewhere between b_{i-1} and b_i, with an offsetting change just above b_i to leave a_{i+1} unchanged; while a change in b_i alone would be a change in that (and, to compensate, in the following) tax bracket's size.

We have the following consequence of P4.

Proposition 5 : *If the second-best social welfare optimum is not a first best, the value of welfare as a function of direct controls is, at the optimum, discontinuous (in some of its arguments).*

The proof is left to an appendix. The intuition behind this result is that as consumers "move" across equally desirable elements of $\{C_i\}$ (there *are* links, by P4) reacting to minute changes in the direct controls, the cost of their consumption, or the tax or profits they contribute, vary in a finite way. The result is somewhat surprising given its generality: it appears to be a feature of discrete incentives problems of the present sort, with little dependence on the nature of the principal's objectives or on other assumptions; it in fact holds in a rather stronger form (discontinuous drop in value when (nearly) any control is moved at all, upwards or downwards) both when the usual redistributive assumptions are built into the objetive function (proposition 8, below) and for the profit-maximization problem subject to incentives, which is not being considered here. Solutions are "infinitely sensitive" to mistakes at the margin.

Some remarks are in order:

First, the discontinuity arises because the x^h's proper do *not* vary slightly when $\{C_i\}$ is perturbed. It is the *solution* of the problem in terms of $\{C_i\}$ which is discontinuous and not the functions (of (x^h)) involved: welfare, costs, behavioural constraints. The Kuhn-Tucker theorem can still in general be applied. Nevertheless, this technical fact will be little comfort to the policy-maker in charge of controlling (x^h) only through $\{C_i\}$. This is a troublesome phenomenon which more generally occurs in problems where the reaction functions of agents are not continuous with respect to the policy variables.

Secondly, however, the size of the discontinuity is essentially commensurate to the discontinuities amongst consumers themselves (or more precisely to the gaps amongst the C_i's) and is expected to become small as the population grows dense. The policy implications of the result will accordingly be quantitatively negligible in most cases of interest, with the possible exception of certain markets or sections of markets with a truly reduced set of consumers.

Lastly, it is interesting to note an alternative form of P5: if the government is *not* given the faculty to select the point a man gets from his optimal set (but, instead, the consumer himself chooses according to some rule), then it is *existence* of an optimum that will generally be lost, although for the reason given in the above paragraph ε-optima will still be there, converging on a true optimum as the population approaches a continuum.

4. *Particular Patterns of Links*

In order to be able to say more on the nature of optima we now introduce restrictions on the way behaviour differs across households, on the one hand, and on the government's desired redistribution on the other. The assumptions we adopt are restrictive but usual, hence at this stage desirable, so as to concentrate on where and how the behaviour of the problem changes as a result of the discretization itself.

Assumption B : Consumers can be indexed in such a way that

$$h > h' \Longleftrightarrow s^h(x) < s^{h'}(x), \forall x, h, h'.$$

This is strong, for it simply cannot hold as a continumm is approached --unless the population is unidimensional (parametrizable by a single attribute and otherwise identical), which is strong all the same.[4] It will be one of our main tasks in future work on this topic to find interesting relaxations of (B) without losing the handle on the problem it provides us with.

Unidimensionality of consumers is not sufficient for (B) to hold, but the additional conditions required would seem to be weak, and are so in various simple special cases of interest: in a "pricing" model, with identical consumers differing only in their transfer incomes (i.e. $u^h = U(a, I^h - b)$; see fn. 1), (B) holds iff the singled-out good ("electricity") is normal. Similarly in the income-tax model of Mirrlees, where consumers differ only in their hourly wage and have identical leisure/consumption preferences (i.e. $u^h = U(a, b/w^h)$, $b \equiv$ gross income), (B) holds if consumption is normal.

The following implication of (B) is rather useful:

<u>*Lemma 2*</u> : *Under* (B), (i) *two corners* C_i, C_j *can only be linked together if they are "neighbours"* ($i - j = \pm 1$) *or have at most one other corner in between, W-linked to them by the same person's equilibrium indifference curve; and* (ii) *all W-and L-links are between neighbours.*

Proof : We first notice that, under assumption (B), pairs of indifference curves of any two agents h, h' $(u^h=K, u^{h'}=K')$ can have at most one intersection. The formal proof of this point, which is diagrammatically straightforward, is left to the reader.

Consequently, supposing that some consumer h^0 is indifferent between corners C_i and C_j, $j > i$, all $h^+ > h^0$ will strictly prefer C_j to C_i and all $h^- < h^0$ will strictly prefer C_i to C_j. Hence any corner C_k between C_i and C_j which would be strictly less good for h^0 than C_i and C_j, would not be chosen by anybody. Such a C_k would then have to be indifferent (for h^0) to C_i or C_j, and actually chosen by h^0 (for it to be a member of $\mathcal{C}$). It is straightforward that there can be only one such C^k. This completes the proof.||

Figure 4 makes intuitively clear that C_i and C_j must either be successive (with h^0 placed on either of them) or have at most a corner C_k in between, with $x^{h^0} = C_k$; this would only depend, for efficiency, on which of C_i, C_j or C_k is least costly in production.

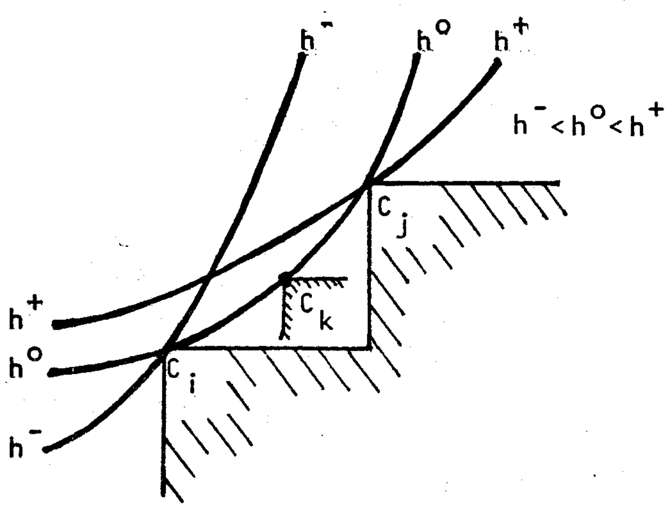

Figure 4

Let us now turn from consumers' to the principal's preferences. One might assume these to be utilitarian, i.e. defining welfare as Σu^h. But without any explicit indication of how the particular cardinalization u^h is to be selected for each h, no conclusion can be reached on the kind of redistribution amongst consumers that may be desired. It seems preferable to directly adopt assumptions of the following kind, which can be checked in particular cases:

<u>*Assumption R*</u> *([strong] redistributive assumption) : For each pair of corners* C_i, C_j, *with* $i<j$, *it is socially desirable to redistribute some amount* $\delta a>0$ *of commodity* a *from (any)* $h' \in H_j$ *to (any)* $h \in H_i$, *provided incentive-effects are ignored, i.e. the* $\{H_i\}$*-structure being held fixed.*

That is, commodity a would be better placed if transfers could be made in a lump sum fashion from any corner C_j to any other C_i down the scale, $i < j$. In the case of a utilitarian social welfare function, assumption (R) is equivalent to : $u_a^h > u_a^{h'}$ if $h \in H_i$, $h' \in H_j$, $j > i$. This latter property would not necessarily have to arise as a result of utilities being higher at higher i, but that would be the natural interpretation or justification for (R). This assumption is usual and perhaps natural, but indeed restrictive, for it could be clear (or agreed) that *utilities* should be redistributed in a given direction without (R) following, for consumption of good b is also varying and affecting u_a in the process. The following relaxation of (R) allows for this possibility, requiring only that it be desirable to redistribute some basket $(\delta a, -\delta b) > 0$ (consumption and/or value-units of leisure, say), hence utility, from top to bottom.

Assumption WR ([weak] *redistributive assumption*) : *For each pair of corners* C_i, C_j, *with* $i<j$, *there is a pair* $(\delta a, -\delta b) > 0$ *which it is socially desirable to tranfer from* (any) $h' \in H_j$ to (any) $h \in H_i$, *provided incentive effects are ignored.*[5]

Having made a case against assumption (R) we now revert to it to simplify the analysis. Most results below seem to follow under (WR) all the same, but the arguments we have are messy, and a careful investigation of the relation between these two assumptions in the optimum merits a more detailed examination than we have so far afforded it.

It is clear that, under (R), the optimum cannot have $(i, j | i < j)$ such that c_i is upward free and c_j is downward free: no incentive effects will arise, by definition, if a positive δa is transferred from c_j to c_i, which by (R) will be a desirable change. This fact will now be generalized to incentive-free transfers between sections (*chains*) of the schedule, say from $(c_i, \ldots, c_T)$ to $(c_0, \ldots, c_{i-1})$.

We define a *chain* $\bar{\mathcal{C}}$ to be a set of corners connected by links. That is, for each two subsets $\mathcal{C}_1$, $\mathcal{C}_2$ such that $\bar{\mathcal{C}} = \mathcal{C}_1 \cup \mathcal{C}_2$, $\exists$ $c_1 \in \mathcal{C}_1$, $c_2 \in \mathcal{C}_2$ s.t. c_1 and c_2 are linked together.

The chain is *simple* if its elements are successive corners (i.e. $\{c_{i_0}, c_{i_0+1}, \ldots, c_{i_f}\}$) and each c_i is linked with c_{i-1} only, for $i = i_0 + 1, \ldots, i_f$.

(Note that any c_i trivially forms a chain by itself).

A chain $\bar{\mathcal{C}}$ is *higher* than another chain $\bar{\mathcal{C}}'$ if $\forall$ $c_i \in \bar{\mathcal{C}}$, $c_{i'} \in \bar{\mathcal{C}}'$, we have $i' > i$ -- i.e. each of its corners lies higher up the schedule than each of the other chain's. *Lower* is similarly defined, and it is clear that neither of these relations must necessarily hold between arbitrary pairs of chains.

A chain is *loose* if none of its elements has links outside the chain, i.e. if it is not a proper subset of a larger chain. It is *W-linked* (resp. *L-linked*) if some of its elements is/are W-linked (resp. L-linked) with some corner *outside* the chain.

A chain is *monotonic* if all its W-links run in the same direction up or down the schedule, i.e. if $C_i \in \overline{\mathcal{C}}$ being W-linked to $C_j \in \overline{\mathcal{C}}$ implies i>j, or implies i<j. It is monotonic *to the left* if the first of these cases holds, i>j, i.e. each $C_i \in \overline{\mathcal{C}}$ being W-linked to a C_j further down the schedule. Monotonicity *to the right* is similarly defined.

Lastly, a chain is *upward-free* (resp. *downward-free*) if for some small-enough $\varepsilon>0$ there exists a non-zero sequence (δa_i), with $0 \leqslant \delta a_i < \varepsilon$ (resp. $0 \geqslant \delta a_i > -\varepsilon$) in all its elements, such that moving the chain from $\overline{\mathcal{C}} \equiv \{C_i\} = \{(a_i, b_i)\}$ to $\overline{\mathcal{C}}^+ \equiv \{C_i^+\} = \{(a_i + \delta a_i, b_i)\}$ does not affect the sets H_i, $\forall$ i = 0, ..., T.

One might expect that properties previously obtained for corners generalize for chains -- in particular that a chain which is not L-linked (resp. W-linked) be upward-free (resp. downward-free). In fact this turns out not to be correct for the general case, as the reader will convince himself by examining the configuration of fig. 3 above. But we have the following:

Lemma 3 : *A monotonic chain* $\bar{\mathfrak{C}}$ *which is not L-linked (resp. W-linked) is upward-free (resp. downward-free).*

Proof : As the chain is monotonic, not both of its end-points (corners with highest and lowest indices within the chain) are L-linked (resp. W-linked) to other corners in the chain; nor to corners not in $\bar{\mathfrak{C}}$, by hypothesis. It follows that either of these end-points is upward-free (resp. downward-free), by L1. The chain itself is therefore upward-free (resp. downward-free) by definition.||

The following obvious fact is stated for easy reference:

Lemma 4: *Under (R), in the optimum, there cannot be two monotonic chains one of them* $\bar{\mathfrak{C}}'$ *not L-linked and the other one* $\bar{\mathfrak{C}}''$ *higher than* $\bar{\mathfrak{C}}'$ *and not W-linked.*

Proof : Straightforward from lemma 3 given (R).||

Lemmas 2 and 4 adopt assumptions (B) and (R) respectively. Both of these assumptions seem natural in applications, and underlie a number of the main results in models of income tax or nonlinear pricing with populations forming a continuum. Let us now suppose that both (B) and (R) hold. This allows us to combine the results above and obtain the following simple qualitative picture of the pattern of links in the optimum.

__Proposition 6__ : Suppose that (B) *and* (R) *hold. In the optimum, the budget set is a simple monotonic chain to the left. That is,* (i) *each pair of successive corners will be linked, and* (ii) *these links will all be W-links of* C_i *to* C_{i-1}*, for* $i = 1, \ldots, T$.

__Proof__ : Recall from lemma 2 that, under (B) a corner cannot be W-linked or L-linked but to a neighbouring corner.

We first prove that C_T is W-linked to C_{T-1}. Suppose the contrary. Consider $\bar{\mathcal{C}}'' = \{C_T\}$; it is clearly a monotonic chain, not W-linked, and higher than any other chain that may be formed not including C_T.

Now consider C_o. If C_o is not L-linked, put $\bar{\mathcal{C}}' = \{Co\}$.

Otherwise, C_o must be L-linked to its only neighbour C_1, and by (B) (or, independently, for optimality, as in proposition 3), C_1 cannot be L-linked back to C_o. Hence either C_1 is not L-linked at all, or is L-linked to C_2. In the first case, put $\bar{\mathcal{C}}' = \{C_1\}$.

In the second case, with C_1 L-linked to C_2, C_2 cannot be L-linked to C_1, just as above. If C_2 is not L-linked, put $\bar{\mathcal{C}}' = \{C_2\}$. If it is L-linked to C_3 we proceed as before examining subsequent C_i's. But C_{T-1} is not L-linked to C_T, by hypothesis. Hence we eventually come to a C_p, $0 \leq p \leq T-1$, which is not L-linked. Put $\bar{\mathcal{C}}' = \{C_p\}$.

The two chains $\bar{\mathcal{C}}'$ and $\bar{\mathcal{C}}''$ meet the conditions stated in lemma 4 and thus contradict optimality. It follows that C_T is W-linked to C_{T-1}. Furthermore, again by (B) as above, C_{T-1} cannot be W-linked to C_T. The pair $\{C_{T-1}, C_T\}$ forms a simple monotonic chain to the left.

Now proceed by induction: supposing $\{C_k, \ldots, C_T\}$ to be a simple monotonic chain to the left we prove that C_k must be W-linked to C_{k-1} (for k>1), or equivalently that $\{C_{k-1}, \ldots, C_T)$ is a simple monotonic chain to the left too. To do this we consider $\bar{\mathcal{C}}'' = \{C_p\}$ as above, now for some p: $0 \leq p \leq k-1$. ||

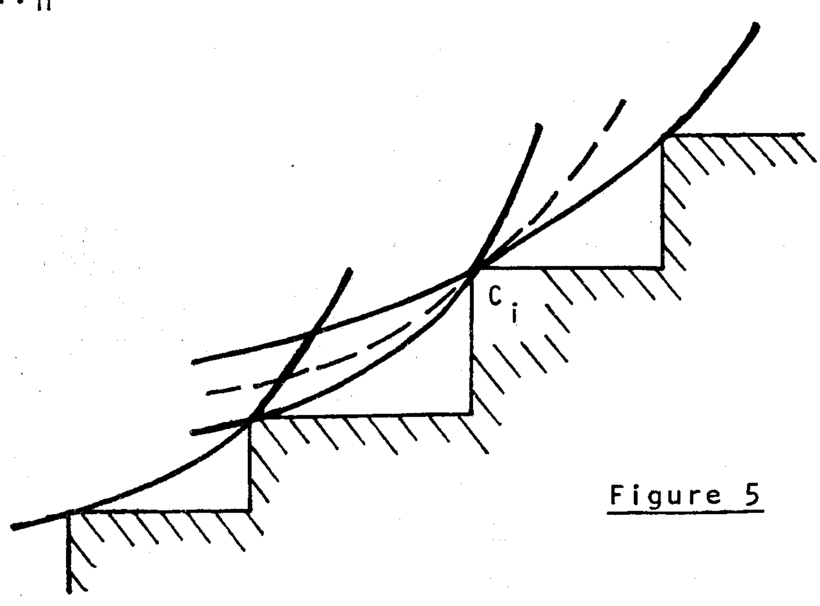

Figure 5

Hence the structure of links in the optimum, under these assumptions, is rather fully specified, and simple, as shown in figure 5. (Bunching is not excluded, however, as in C_i in the figure). This is the central result that follows from (B) and (R); the counterparts of familiar results (for the continuum case), some re-confirmed and others refuted for the finite case, follow from here.

Corollary 6.1 : *Under* (B) *and* (R), $s^h(x^h) \leq 1, \quad \forall h$.

That is, the "apparent marginal tax" (price ratio less s^h) faced by all consumers is non-negative.

Proof : Consider C_i, $i > 0$, and the consumer h' who W-links C_i to C_{i-1}. Clearly h' is the smallest $h \in H_i$, so that $s^h(C_i) \leq s^{h'}(C_i)$ $\forall h \in H_i$, by (B). Suppose $s^{h'}(C_i) > 1$. Keep the corner C_i fixed but create a new corner C_i' "just"

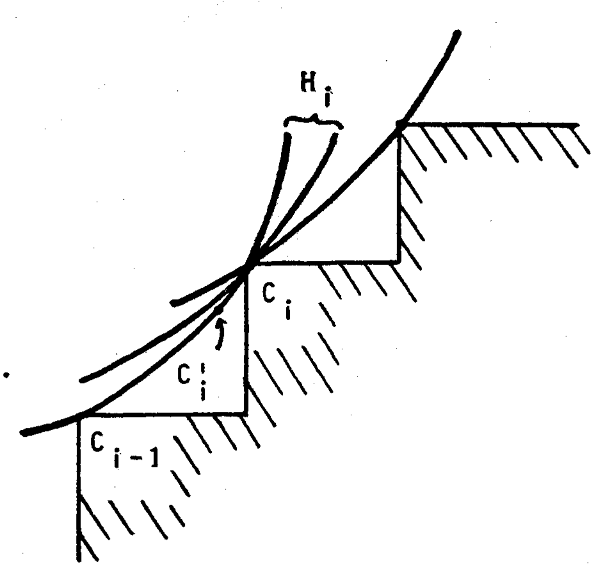

Figure 6

below C_i along the relevant indifference curve of $x^{h'}$, as in fig. 6. Moving $x^{h'}$ to this C_i' decreases the cost of the bundle $x^{h'}$ without changing any utility. By continuity points exist in a neighbourhood of C_i' which are strictly preferred by h', and still cheaper than C_i. On the other hand no other x^h will change: in particular no-one else links C_i to C_{i-1}, and the person L-linking C_i to C_{i+1} (if i<T) as well as other $h \in H_i$ have flatter indifference curves trough C_i than that of h', hence will not react to the changes described.||

<u>*Corollary 6.2*</u> : *Under* (B) *and* (R), *there is no bunching nor distortion at the top* : $s^H(C_T)=1$.

<u>*Proof*</u> : Direct from propositions 2 and 6.||

Hence C_T is, under these conditions, the (and the only) point C_* of corollary 3.1.

A similar no-distortion result holds at the *bottom* of the scale in the continuum-model and one might expect the same to be true here. This is not implied by P5 as directly as the above corollary is but neither is it ruled out --just as in the continuous case, the no-distortion result might be immediate at the top and less so, but still hold, at the bottom. However, it is perhaps somewhat surprising that this is not the case here. This is contained within the next result, strong form of corollary 6.1.

Proposition 7 : *In the optimum, under* (B) *and* (R), $s^h(x^h) < 1$, $\forall$ $h<H$. *That is, there is a strictly positive distortion at all points but the top.*

This further implies, clearly, that total tax-liability or cumulative mark-up are strictly increasing functions of income or purchases. This follows from the observation that the *entire* arc of the indifference curve linking any given pair of corners has gradient less than one, for it does at the *higher* of the two corners, by P6 and P7. Hence successive corners always lie on lower and lower iso-cost 45°-lines.

For the sake of simplicity we shall only give a heuristic proof of this proposition, using second- *vs* first-order arguments informally. The argument can easily be made rigorous.

Proof (*of P7*). Consider the bottom-point and suppose for simplicity that there is no bunching there. Suppose that, contrary to what is stated in P7 (and using corollary 6.1), the bottom-man faces no distortion: $s^o(x^o) = 1$. The situation is as shown in fig. 7, where the relevant indifference curves for the bottom two consumers are shown: $h = 0 \in H_0$ and $h = 1 \in H_1$. A direct incentives-free redistribution to C_o is impossible, for C_o has an L-link. But this we achieve as follows. First replace C_o by C_{o-} as shown, setting $x^o = C_{o-}$. Utilities have not changed and the cost of the move is second-order relative to $|C_o - C_{o-}|$, for the shift is along an iso-cost line, by $s^o(C_o) = 1$. But now C_{o-} is upward free (not linked at all) and C_1 is downward free (not W-linked): a desirable redistribution from the latter to the former can be effected, improving welfare to first order.

The argument applies essentially unchanged if there initially is bunching at C_o, and it equally applies (given the pattern of links of P6) to any C_i which has a C_{i+1} to its right --i.e. all observed demands but the top.||

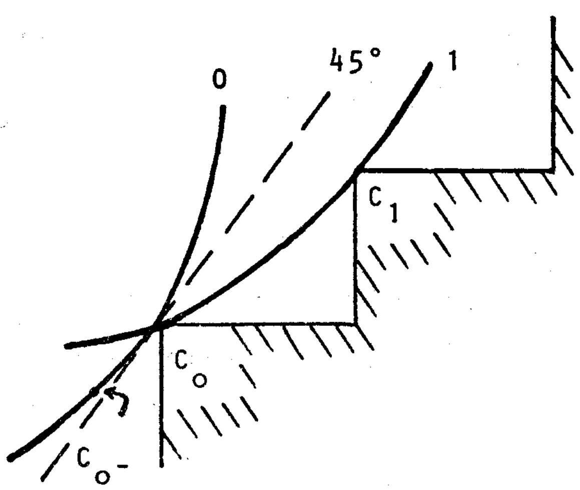

Figure 7

Hence, as announced, the distortion at the bottom is positive, contradicting the fairly general property of solutions for the continuous modal, that distortions be zero at both end-points when these are interior (Seade (1977)).

In order to make more apparent the way in which top and bottom differ in the discrete case, we can put this result in the following form. Suppose (B) holds, so that the system has a modicum of good behaviour, and that either (R) or its converse (increasing "deservingness" as one moves up the scale) holds. Then at the end-point which is a *source* of tax there will be no bunching nor distortion, whereas at the end-poind which is the *final recipient* of benefit the tax is positive at the margin --and bunching, so it seems, cannot be excluded either. Intuitively, allowing this "final recipient" to freely vary his tradings with producers at the margin (no distortion) will have a redistribution cost, for others will try to take his place (his benefits);[6] whereas free tradings permitted to the "source" will either attract no-one, for people would have to first become top-payers too, or else attract some of them --increasing revenues all the same.

A third direct consequence of P6 is the following strong form the discontinuity result P5 takes.

Proposition 8 : *Under* (B) *and* (R), *the value of the problem is, at the optimum, a discontinuous fuction of* (*all*) *the direct controls; in fact discontinuous from "right" and "left" in each coordinate of all but the top-and bottom-points.*

Proof : Consider any "interior" point C_i, $0 < i < T$. Given the structure of links described in P6, an increase in a_i above its value at the optimum will induce the consumer W-linking C_{i+1} to C_i to "jump" from the former to the latter point, which by P7 is strictly more costly than C_{i+1}. Utilities (including those of other $h \in H_i$) vary to first-order only, but costs rise in a finite way, discontinuously. The same is true if a_i is, instead, decreased : it is now the consumer W-linking C_i to C_{i-1} who will immediatly have the incentive to shift his consumption to the latter, more expensive, point. Similarly for increases or decreases in b_i and again similarly, but only in one direction, for each a and b at the end points.||

The discussion of P5 (pp. 20f) applies, of course, again here. We may only add a remark on why this discontinuity property of the optimum is after all what one would expect on economic grounds. Incentives apart, and under redistributive assumptions, one whishes to levy as much tax as possible from a man at the "income" level given by C_i, to transfer it to lower points. That is, one wishes to lower that man's utility till the point comes where this is not possible any longer without suffering incentive losses. This is why we always want each h (or the lowest h in each corner's group H_i) to be *indifferent* between taking the larger burden he has been assigned and dropping to the next income-category down, and that is also why any extra load causes him to react discretely.

5. *Concluding Remarks*

A central purpose of this paper has been to study the non-linear tax or pricing problem using basically no formal tools, replacing these by direct arguments throughout, which are both easier to grasp and often useful in permitting the analysis of situations where more formal approaches fail or are too hard to apply.

As far as specific results are concerned, we have found, for the general case (imposing Paretianism only), (i) that production efficiency is generally desirable; (ii) certain features of optima, including the fact that there will always be someone at *some* point in the scale facing undistorted producers' prices; and (iii) a basic discontinuity of the (value of the) optimum as a function of the policy tools of the principal. We then imposed further assumptions, on the way consumers differ amongst themselves and that the principal has a specific (usual) kind of redistributive preferences, and found a rather full characterization of the qualitative nature of the optimum (Prop. 6), from which some results from the (continuum) taxation literature were re-obtained or disproved for the finite case: amongst these are (iv) the requirement that the marginal tax be strictly positive throught the schedule (Mirrlees (1971), Seade (1980)[(7)]), (v) except at the top where it is zero (Sadka (1976)), noting in particular that (vi) the distortion at the "most deserving" end-point, usually the bottom, is *not* zero in the discrete case, unlike the result that obtains with a continuum (Seade (1977)). Lastly, (vii) it was noticed that the discontinuity result of (iii) above, takes a somewhat disturbingly strong form here, that the value of the optimum is discontinuous from both sides in the two coordinates of basically all points in the optimal price or tax schedule.

It is worth noting finally that the Paretian or the redistributive assumptions that we have adopted can be replaced by other objectives of the principal: the methods used and behaviour of the problem remain pretty much the same. In particular, *all* results in the paper hold in just the same form when the objective is the maximization of profits[(8)].

Appendix : proof of P5

By P3, $\exists$ C_j, C_k linked by some h', with $x^{h'}$ = either C_j or C_k. Let j<k.

Two possibilities arise, the main one being the case when C_j and C_k are *not* equally costly. Suppose C_j is the cheaper of the two, as in fig. 8, so that for optimality $x^{h'} = C_j$. Small vertical shifts of either of C_j and C_k will only change utilities to first order (at most), i.e. proportionately to the shift δa introduced. But the revenve *losses* (not gains, if the allocation *is* an optimum) from some such moves will be large, finite, as h' takes his consumption to or from C_j or C_k, evading a fall of "his" point or attracted by an improvement in the other one as the case may be. Hence, *at* the optimum, the value of welfare net of (shadow) production costs, $W = \sum_h\{u^h - p\cdot x^h\}$, will move discontinuously with $\{C_i\}$: it will at most be upper-semicontinuous in a_j (given $x^{h'} = C_j$), and at most lower-semicontinuous in a_i for any C_i L-linked to C_j. Similar remarks apply if horizontal (b-) shifts are considered.

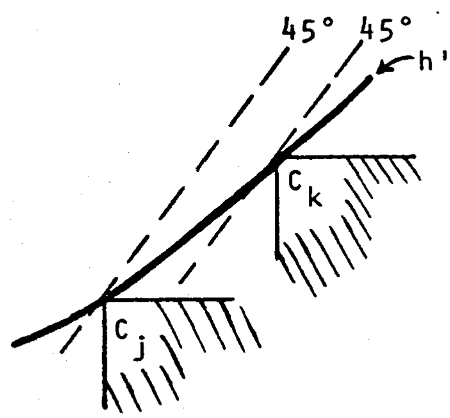

Figure 8

The second case has C_j and C_k equally costly. Say the former is W-linked (by h') to the latter, i.e. $x^{h'} = C_j$. If this is the only link C_j has, it can have no bunching, by P2. Hence, in that case, the relevant (optimal) indifference curve of h' is the only relevant indifference curve through C_j. This cannot be optimal: moving $x^{h'}$ to a point up along the 45°-line towards C_k but sufficiently near C_j will induce no changes by other consumers, keep costs constant, and raise $u^{h'}$. On the other hand, C_j may have other links (to C_k itself-by some other h- or to other corners). If any such link is with a $C_{k'}$ wich is *not* as costly as C_j, the discontinuity result obtains directly, as in the previous paragraph.

But if these links are with corners which are as costly as C_j, we can move the corresponding consumers' bundles all to C_j, changing no costs or utilities and leaving C_j with bunching but with no L-links, which by P2 cannot be optimal.||

Footnotes

(1) To illustrate, suppose utility depends on the consumption of electricity a and of a Hicksian-composite good c, the former being subject to nonlinear pricing according to the total-outlay function $P(a)$. The problem $\max U^h(a, c)$ s.t. $P(a) + c \leqslant I^h$, with both preferences and incomes varying across h, can be rewritten $\max U^h(a, I^h - b)$ s.t. $P(a) \leqslant b$, where b is defined as $b \equiv I^h - c$, i.e. the "slack" left, to buy a, after buying numeraire. The budget constraint has become the same $\forall h$, and income-differences have been translated into the nature of the utility function: $u^h(a, b) \equiv U^h(a, I^h - b)$.

(2) Here and later the use of pairs of different indices such as i and j or h and h' is directly taken to mean *different* elements of the corresponding sets, i.e. $i \neq j$, $h \neq h'$.

(3) It would be more natural to take the *controls* in the technical sense to be the x^h's; hence the qualifier "direct" used above, to emphasize that these are the actual tools the principal chooses directly and announces. These controls should also of course include the *number* of corners C_i (hence whether there is bunching or not) and not only their positions, but that is mere convention on terminology.

(4) In the continuum, demands at each point on the *line* $\mathcal{C}$ will generically correspond to a sub-population of dimension one less than that of the population itself (codimension one, relative to the latter).

(5) Assumption WR could be weakened further, permitting that the vector $(\delta a, - \delta b)$ be different for each pair h, h' being considered.

(6) But if those who react, following a slight fall in the distortion, are "infinitely close" to our end-point recipient, the weight of this distributive effect goes to zero, and so does the tax. Hence the key difference between the continuum and the present model is, in this regard, the *measure* of consumers reacting at the margin.

(7) The former obtains the result (in weak form --non-negative distortion) under somewhat stronger assumptions; the latter extends it basically to the form P7.

(8) In fact the arguments used in the proofs do themselves apply in most cases perhaps suitably adapted: we often found that unless something was true, additional net revenue could be exacted from consumers without lowering their utilities (hence without inducing them to opt out from the scheme, a central

feature to take explicit account of under profit maximization). This added net revenue has been used to raise utilities in this paper, but would itself be an addition to profits in the case being discussed.

References

Diamond, P.A. and J.A. Mirrlees, 1971, "Optimal taxation and public production: I-production efficiency", American Economic Review 51, 8-27.

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

Sadka, E., 1976, "On income distribution, incentive effects and optimal income taxation", Review of Economic Studies 43, 261-267.

Seade, J., 1977, "On the shape of optimal tax schedules", Journal of Public Economics 7, 203-235.

Seade, J., 1980, "On the sign of the optimum marginal income tax", Warwick Economics Discussion Paper no. 166 (April).

Willig, R.D., "Pareto-superior nonlinear outlay schedules", Bell Journal of Economics 9, 56-59.