SOME THEORETICAL ASPECTS OF OPTIMUM REDISTRIBUTION

William Anthony Jackson

Ph.D. thesis submitted to the University of Warwick

Research conducted in the Department of Economics, University of Warwick

Date of submission: July 1987

CONTENTS

Page

ACKNOWLEDGEMENTS

My period of full-time research at Warwick University was supported by a two-year E.S.R.C. Research-Linked Studentship, for which I am grateful.

I would also like to thank Professors N. H. Stern and K. W. S. Roberts for their helpful and expert supervision.

V. A. J.

SUMMARY

As its title suggests, the thesis considers various theoretical aspects of optimal government policy. The method is described in Chapter One: it is essentially that of 'normative' public economics and provides the common theme of the remaining chapters.

Chapters Two and Three are concerned with redistributive policy in general. Chapter Two discusses the use of observed information in redistributive policy. It derives the appropriate optimality conditions in simple models of redistribution, comparing the outcome with the existing theoretical literature and actual redistributive policy. Chapter Three considers quantities as redistributive tools, in contrast with the more usual concentration on incomes and prices.

Chapter Four addresses the optimal taxation of wealth. To tax current wealth under optimal life-cycle saving may imply negative marginal tax rates at some point of the tax schedule, an outcome avoided when lifetime wealth is taxed directly. The principal theoretical obstacle to redistributing wealth is found to be the anticipation of tax implementation.

Chapters Five and Six are both concerned with unemployment benefits. Chapter Five discusses them within the optimal policy framework. Attention is first concentrated on the optimal level of benefits and then on their optimal time pattern, in cases where they can vary with the duration of unemployment. Chapter Six digresses from the optimal policy format to discuss the macroeconomic role of unemployment benefits, arguing that the replacement ratio deserves a more explicit inclusion in the Keynesian income/expenditure analysis.

Chapter Seven applies the optimal policy method to pension and retirement practices. The initial concern is whether or not formal pensions and retirement are theoretically justifiable as part of an optimal approach to policy. The discussion is then broadened to consider redistributive issues, within and between generations.

Chapter Eight concludes.

CHAPTER 1 : GENERAL INTRODUCTION

The common theme of the following chapters is the redistribution of income and resources between individuals. This presupposes some advantage to be gained from redistribution: anyone who is indifferent to the distribution of resources would find little of interest in the discussion below. Such views are probably rare, and anything less than the extreme position allows at least some scope for redistributive policies. More common opinions are that redistribution should be limited to special circumstances or that it is subordinate to other policy measures. The former idea arises when the objective is the maintenance of minimum living standards; redistribution is valuable in guaranteeing these but not otherwise. Much of the welfare state seems to have originated on these principles, and they find formal expression in libertarian writings (Friedman (1962), Hayek (1960)). The role of redistribution is curtailed under this philosophy, although it is not removed entirely. A different set of arguments sees redistribution as subordinate to other economic concerns, and in particular to economic growth (see, for example, Beckerman (1979)). The discussion here does not aim to contradict this view: a nation's prosperity has more to do with its long-term growth rate than its distribution of income. All that is implied below is that some redistribution may be desirable, giving reason to implement it through a redistributive policy.

This introduction firstly sets out the background to the models used, then looks at their particular nature, and finally summarises the topics and issues to be covered.

Background and Rationale

Most of the following analysis is utilitarian in form, using a social welfare objective defined in terms of individual utilities. Redistributive policy is implemented by an outside agency (the 'government') in such a way as to maximise social welfare. The approach is 'consequentialist', in the sense that it is the consequences of actions that matter, rather than the actions themselves (on this issue see, for example, Smart and Williams (1973)). It follows that any feasible means of redistribution is legitimate in these models, ruling out objections based on fundamental individual rights (as put forward on a broad front by Nozick (1974)). In practice some redistributive tools are quite likely to be seen as undesirable on principle (those based on certain kinds of discrimination, for instance), but there is no room for such views within the theoretical structure of the models below. The desire for redistribution arises formally from two main sources:

(a) Diminishing marginal utility of income or goods implies that equalisation of income or consumption raises social welfare. This applies even where the social welfare function is strictly indifferent to equality, as under pure Benthamite utilitarianism.

(b) The social welfare function may also entail inequality aversion, as summarised in its degree of convexity to the origin. A standard example of this is the isoelastic form, where

$$V = \frac{1}{1-\theta} \sum_h [(U_h)^{1-\theta} - 1]$$

and θ represents the degree of inequality aversion. Setting $\theta=0$ gives an inequality neutral objective, while the limiting case as $\theta \to \infty$ implies Rawlsian maximin social preferences (following the ideas of Rawls (1971)).

One or other of these features is assumed below to provide a reason for redistributive measures. The result can be seen as normative public economics, which assesses alternative policies on the basis of a particular social objective. More will be said later about the nature of the modelling; initially it is appropriate to consider the rationale for this line of approach.

In modelling the redistribution of personal incomes, the simplest possible starting point is to define a population of individuals and an agency for redistribution which has specific social objectives. A utilitarian method provides such a framework and in that sense is an 'obvious' way of proceeding. Its main virtue is that it can look in a formal way at the implications of well-defined social objectives, and thereby give an explicit account of the interrelationship with policy measures. Against this directness has to be set a number of possible disadvantages, which are discussed below.

An issue which immediately arises is the nature and origin of the social welfare function. It is assumed here that social welfare is individualistic in form, defined as a function of individual utilities. Since individual preferences are respected, this reduces the degree of paternalism implied in the method. Other types of social welfare are also possible, and a relaxation of the individualistic approach would not alter the structure of the modelling. The precise functional form of social preferences does not have to be specified, allowing a variety of attitudes towards inequality within the general utilitarian framework (on the lines of Sen (1973) or Atkinson and Stiglitz (1980), Chapter 11). Hence the use of social welfare maximisation is not particularly restrictive, and leaves room for considerable variation in detail. A more substantial set of problems concerns the origin of the social welfare function. To attempt to derive an agreed view of social welfare from individual preferences leads to the questions addressed in social choice theory; much has been written on this subject since the original contribution of Arrow (1951) (later work being summarised in Sen (1970)). In general it is not possible to apply a given social choice rule to move from individual preferences to a unique formulation of social welfare. This means that a particular view of social welfare cannot always be given a democratic justification in terms of majority voting or some other form of decision rule. To use a social welfare function in defining 'optimal' policy measures, one has to accept that the social preferences used may not have a firm grounding in the wishes of the population. Where this is the case social welfare becomes paternalistic in tone, imposed from without instead of arising spontaneously from within. The choice of social welfare can be seen literally as that of a reforming, paternalistic government or alternatively as the opinion of an outside observer assessing different policy measures. These conditions are not necessarily restrictive or extreme. In practice some choice of policy has to be made, and so some notion of social welfare is always being imposed; it seems as well to acknowledge this by adopting an expressly chosen set of social preferences.

Another general difficulty is in the role of the state. The models below represent the state as an independent agent (the 'government'), which implements policy in the interests of society, as reflected in the social welfare objective. There is no notion of the self-interest of the state, nor is there a supposition that the state defends the sectional interests of part of the population. Impartiality of the government is an assumption frequently questioned, with two main lines of criticism:

(i) On the one side is the libertarian view that state activity is inevitably inefficient and cannot be relied on to pursue society's interests. Instead a growing bureaucracy tends to emerge, which is concerned with defending its own position and fails to fulfil its intended role. Such views are put forward, for example by the 'Virginia School' (Buchanan and Tullock (1962)) and lead to calls for a diminution of state intervention.

(ii) A contrasting set of arguments is based on the Marxian principle that the state supports the interests of the dominant class in society. In a capitalist economy the government is always biased in favour of the owners of capital, preventing the achievement of any thoroughgoing redistribution. Only a change of economic system can offer any hope of improving the relative position of the non-capitalist classes.

These ideas produce very different conclusions, but share a common scepticism about the workings of the state. There is certainly some truth in the belief that the state will never be a perfect, impartial agent, but the conclusions to be drawn from this are less clear. One can support virtually anything with claims about the inadequacies of the state: Pareto used similar reasoning in defence of fascism, and the arguments fit naturally into an advocacy of anarchy. In all cases the solutions put forward are less convincing than the criticism on which they are based. Libertarians seem unconcerned at leaving society to be dominated by private commercial interests - in this environment even the self-interested state depicted in their theories would be an alternative power source, producing a slightly more pluralistic outcome. The Marxian view has little to say about post-capitalist society - if the state still exists it is not clear why the hitherto universal scepticism suddenly ceases; if the state has withered away, it would be desirable to know what follows. These are difficult and controversial issues, with no prospect of being resolved in the near future. They are avoided in the following analysis by assuming an abstract and idealised 'government', which always acts to maximise social welfare. Nothing is said about the nature of the government, nor about the problems which might arise in the course of state economic intervention. This leaves two alternative ways of interpreting the models. Where state activity is acknowledged as having some social value, the 'government' in the model can be seen as an approximation to a possible actual state. The policies in the models then represent state activity which might potentially occur in some form. Under a more critical view of state activity, the analysis is not necessarily worthless - it becomes a hypothetical exercise, which can be contrasted with policies observed in reality. The differences can serve to support the various assorted critiques of the state.

Associated with the previous point is the lack of any institutional detail in the models. Apart from the government and a population of individuals, there are no other agents which have any influence on the outcomes of different policies. No explicit mention is made of production or the form of social organisation it entails. The type of society being discussed is not made specific, although the topics covered suggest a developed capitalist economy. The outcomes obtained are therefore on the condition that the details of social institutions can legitimately be ignored. In some of the areas (such as wealth taxation) this may not be true, and the results of the models have to be assessed accordingly. One institutional aspect that is always present is the question of policy administration. Costs of administration are assumed to be zero in the models below, but in practice they will always be present. For state intervention to be worthwhile, it must be true that the social benefits outweigh the administrative costs. The general framework used below addresses certain common features that can always be expected, but does not necessarily give an adequate representation of cases where institutional details are important.

Many other, less central, issues can be raised. If individuals always dislike being constrained in their behaviour, then the case for interventionist policy is weakened (Hahn (1982)). There are general difficulties associated with the definition of 'utility'. In some cases the use of a static formulation may not be desirable (as perhaps in the context of unemployment). Assumptions made about the extent of government information (and use of perfect foresight in intertemporal models) can be viewed as unsatisfactory. This list could undoubtedly be lengthened.

The basic point is that the models used are inevitably stylised and simplified. They are only a caricature of reality, and the best that can be achieved is a 'good' caricature, which identifies and highlights the most important sides of its subject. Whether the method used here accomplishes this is a matter for personal judgement; as with any economic modelling, the formal theory aims only to assist an intuitive understanding of reality.

Structure of the Models

Once within this particular 'vision' of the economy, the treatment of policy issues has a common pattern. There is a two-tier structure, with the government implementing redistributive (and other) policies which affect the population and individual behaviour influencing the government's policy decision. The position is as below:

	Objective	Activity
Government	$W(U_1,\ldots,U_H)$	max W by income, price, quantity controls subject to revenue/resource constraint and individual choice
	Redistributive policy ↓ Disincentives ↑	
Information	------------------------------	
Individuals ($h=1,\ldots,H$)	$U_h(X_1^h,\ldots,X_N^h)$	Max U_h by choice of $X_1^h,\ldots,X_N^h$ subject to budget constraint

At the top level comes the government, which has the maximisation of individualistic social welfare as its objective. It can use a variety of policy instruments in trying to achieve this, which can be categorised as income, price and quantity controls. Individuals seek to maximise their own utilities defined in the conventional manner as a function of their consumption of a range of goods, i=1,...,N. Their choice is limited by the usual monetary budget constraint, and by any relevant policy measures imposed by the government. Decisions made by individuals and government constantly interact; individuals must adhere to the government's requirements of them (with taxation, etc., legally enforced), while the government's policy is subject to the behavioural responses of the population (which appear as 'disincentives'). There are four main obstacles facing redistibutive policy, namely, the economy's resource constaint, the availability of policy tools, the extent of information, and the presence of disincentives.

The underlying constraint facing the economy is its scarcity of currently available resources. Without such a constraint a bliss point would be attained, and there would be no economic problem to be faced. This appears in the present framework as the government's resource or revenue constraint, which represents the limits within which redistribution must operate. In most of the following discussion this is expressed as a revenue constraint, such that the government has to balance its budget subject to a given revenue requirement. Since the economic constraint is real rather than monetary in nature, this is intended to represent a limited general resource availability. Changes in the precise form of the constraint would not dramatically alter the character of the analysis. The target is to achieve an allocation at which no further redistribution of goods can increase social welfare. Formally, let the function

$$G(\bar{X}_1,\ldots,\bar{X}_N) = 0$$

denote the general resource constraint, where $\bar{X}_i = \sum_{h=1}^{H} X_i^h$ i=1,...,N. A first-best is obtainable by choosing X_i^h, $\forall$ h, i, to maximise social welfare subject to the resource constraint above. Solution by Lagrangian gives

$$L = V(U_1,\ldots,U_H) + \mu\, G(\bar{X}_1,\ldots,\bar{X}_N)$$

$$\frac{\partial L}{\partial x_i^h} = \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial x_i^h} - \mu \frac{\partial G}{\partial \bar{X}_i} = 0$$

$$\Rightarrow \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial x_i^h} = \frac{\partial V}{\partial U_k} \cdot \frac{\partial U_k}{\partial x_i^k} , \qquad \begin{array}{l} h,k = 1,\ldots,H \\ i = 1,\ldots,N \end{array}$$

Hence the marginal social utility of consuming each good is equated across individuals. Any redistributive policy in the current framework is aiming ultimately at a first-best allocation of this type.

The government's success in implementing redistribution depends on the policy tools it uses. It is rational to deploy as many instruments as possible, although theoretical discussion often concentrates on particular measures in isolation. Instruments may be classified as income, price and quantity controls, corresponding to the three main sets of variables in the analysis. None of these is unambiguously superior to the others, in the sense that any one of them can (hypothetically) be used to attain the ideal first-best outcome. The position is as below:

'First best'	Lump-sum income transfers and efficient pricing Perfect price discrimination Direct quantity allocation
'Second best'	Any other policies

A first best is most often linked with lump-sum income transfers, which are regarded as the best redistributive tool. Their efficacy depends on 'efficient' pricing, so that there is a uniform set of prices that satisfy the usual efficiency requirements (equating marginal rates of substitution for all individuals with the marginal rate of transformation between goods). Such pricing does occur in the special case of perfect competition, although it cannot be relied upon in reality. The same first-best position can also be reached by centralised control of prices or quantities. Under perfect price discrimination a separate price is charged for each good to each individual. This gives the government a total of HN instruments, which is enough to choose the ideal quantity distribution. A similar outcome follows from perfect central planning, where the planning body itself implements the allocation of quantities. Again the government has HN instruments at its disposal, allowing the first best to be selected. The standard view in welfare economics is to emphasise the benefits of lump-sum income transfers; as long as the economy is competitive, the first best is attainable by the use of H instruments (as compared with the HN instruments in the other two methods). The benefits of lump-sum transfers are less clear in the absence of perfect competition, when they are insufficient to implement a first-best position. There is consequently little justification for preferring certain redistributive tools to others. An efficient government would make the maximum use of policy tools, and would not allow any possible redistributive instruments to lie dormant. Where a policy is genuinely not available, then there exists a true constraint on redistributive policy. This issue is considered again in Chapter 3.

A third obstacle to redistribution is the lack of necessary information. Theoretical models are often based on one- or two-dimensional populations, which make full information appear a feasible proposition. In reality the population has so many relevant and unmeasurable dimensions that it is difficult to envisage what is meant by complete information. Omniscience can be represented theoretically by imagining a characteristic vector, $\underline{a}$, of enormous size, which depicts the 'absolute truth', unknown to mortal observers, but enshrining each individual's identity. Without knowing $\underline{a}$ the first best is not attainable, and any redistribution must depend on a subset of characteristics. Suppose that $\underline{a}$ is partitioned to give $\underline{a} = [\underline{a}^*, \underline{a}^\dagger]$, where $\underline{a}^*$ are measurable, observed characteristics; uncertainty must somehow be based on the unobserved vector

$\underline{a}^{\dagger}$. One possibility is that $\underline{a}^{\dagger}$ is utterly unknown, ruling out any optimisation. A less pessimistic (and perhaps more plausible) view is that we are neither omniscient nor bewildered, and that some information on $\underline{a}^{\dagger}$ is available. An example of partial knowledge is 'assignment uncertainty', where exact values of $\underline{a}$ cannot be ascribed to individuals, but the distribution of $\underline{a}$ is known. This is equivalent to a stochastic knowledge of $\underline{a}$ based on the true distribution of characteristics in the population. Optimisation can proceed through an expected maximand, although the outcome for any particular individual remains unknown. An informational barrier of this sort is partially but not wholly surmountable, enough to prevent a first best but not to prevent optimisation. Individual characteristics are further discussed in Chapter 2, which considers redistribution under partial information.

The final barrier is the question of disincentives, which are found in many discussions of optimal policy. A 'disincentive' occurs whenever policy measures provoke an accommodating response from the individuals concerned, resulting in an outcome different from that which the government had initially intended. The issue is addressed in a general form in the 'principal-agent' problem (Shavell (1979), Grossman and Hart (1984)), where the 'principal' seeks to maximise an objective conditionally on the behaviour of the 'agent'. An optimum is achieved when the disincentive effects of the agent's behaviour are minimised or removed entirely. The redistributive models set out below are particular cases of principal-agent problem, with the government as principal and the population as agents. Optimality then requires the individual behavioural response to be included in the government's policy design from the outset. Disincentives can easily be associated with the informational question of the previous paragraph, as occurs whenever the government's information depends in some way on individual behaviour. Many other types of disincentive are also possible and examples appear in Chapters 2, 4, 5 and 7 below.

The following chapters address particular policy issues within the general redistributive framework set out above. Each policy question raises difficulties which can be categorised in terms of the four main obstacles identified here.

Topics Covered in Following Chapters

The remaining analysis is divided between different policy questions, all seen from the viewpoint of optimal redistribution. The chapters are not intended to be exhaustive treatments of their particular subjects, but it is hoped that they identify some of the most important relevant issues. Discussion proceeds as follows:

Chapter 2 considers the optimal use of observed information when making redistributive income transfers between individuals. The outcome is contrasted with previous theoretical discussion and with policies observed in practice.

Chapter 3 looks at quantity constraints and non-monetary allocation schemes, and compares them with other types of redistributive policy.

Chapter 4 considers wealth taxation within the optimal policy framework.

Chapters 5 and 6 are both concerned with unemployment benefits and the position of the unemployed in general. Chapter 5 adopts a utilitarian method, and considers its implications for unemployment benefits. Chapter 6 relaxes the optimal policy framework, emphasising instead the macroeconomic impact of unemployment benefits.

Chapter 7 deals with the issues of retirement and pension provision. It is not easy to identify an optimal approach in this area, but some of the main relevant aspects are discussed.

Finally, Chapter 8 makes some brief concluding comments.

CHAPTER 2 : REDISTRIBUTION BASED ON OBSERVABLE INDIVIDUAL CHARACTERISTICS

(1) Introduction

The most immediate way to set about redistribution is to make use of observed information on individual characteristics, exploiting to the maximum the knowledge that one holds with certainty. Measures based on such information are 'direct' (in the taxation sense), since they depend on known features of the population. A redistributive policy does not have to take this form, and 'indirect' measures (such as commodity taxes) can also be applied to the same objectives. There is no conflict between the two approaches, and indirect measures can be seen as supplementary, to be introduced when the possibilities for direct redistribution are exhausted. This chapter considers direct measures, concentrating particularly on income transfers.

Before the theoretical discussion can begin, it is necessary to incorporate individual characteristics in the standard utility framework. The conventional model of consumer demand is based on a single individual's preferences, with utility a function only of the quantities of the particular goods consumed. Any differences between individuals are assumed to be incorporated in the functional form of preferences, without explicit mention of the differences involved. Such an approach is adequate in considering consumer demand behaviour, but it is not very helpful in representing specific individual characteristics. An alternative method is to enumerate the different characteristics and set them out as a vector, $\underline{a}$. In reality a comprehensive list of personal characteristics would be almost infinite in extent, and virtually impossible to compile in a definitive way. Some of the items included would have no bearing on utility levels, and in many cases would not even be known to the person involved; for example, a well-defined physical characteristic like the number of hairs on a person's head is not common knowledge and has little effect on utility (unless perhaps when visibly approaching zero). The notion of a complete characteristic set $\{\underline{a}\}$ is an intangible entity, although it seems reasonable to suppose that such a thing exists, even if it is not readily identifiable. To introduce this into a utilitarian analysis, it is necessary to incorporate the characteristic vector $\underline{a}$ in individual utility functions, so that utility depends on characteristics as well as consumption of goods. Since the full vector $\underline{a}$ encompasses all differences between individuals, it is not necessary in this case to have variation in the functional forms of utilities, which can be assumed uniform (A view also retained in the following chapters, although the analysis there can still proceed with variation in preferences if the structure of preferences is known to the government). The following discussion is based mainly on simplified cases with only two characteristics, but behind this lies the idea of a very large list of characteristics representing 'full information' on a particular individual.

When the entire characteristic vector $\underline{a}$ is accurately observed by the government, there is no obstacle to the achievement of a 'first-best' redistribution of income. Formally, suppose that there are H individuals, $h=1,\ldots,H$, with associated characteristics $\underline{a}_h$. The government freely adjusts ex post incomes, M_h, so as to maximise individualistic social welfare, $V=V(U_1,\ldots,U_H)$, subject to a total income constraint

$$\sum_{h=1}^{H} M_h = M^{\dagger}$$

where $M^{\dagger}$ denotes the aggregate income level in the economy. Solution by the Lagrangian method with M_h, $h=1,\ldots,H$, as instruments gives

$$L = V(U_1, \ldots, U_H) + \mu(M^{\dagger} - \sum_{h=1}^{H} M_h)$$

$$\frac{\partial L}{\partial M_h} = \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h} - \mu = 0$$

$$\Rightarrow \quad \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h} = \mu \qquad h=1,\ldots,H$$

μ is the marginal social utility (MSU) of government revenue, and is equated at the optimum with the MSU of income to each individual. Hence there is a first-best income distribution, at which it is impossible to increase social welfare by implementing a marginal redistribution of income between individuals. All redistribution based on income transfers is seeking to reach this position.

Within the present theoretical approach, there are two main reasons why the first-best might not be attained. Firstly, the government could fail to observe some of the elements of $\underline{a}$. If no information at all is available on some relevant characteristics, then an exact optimisation cannot be undertaken, and the existence of a precise policy optimum of any sort is ruled out. For most characteristics playing a major role in policy questions, however, there is not presumably total ignorance on the part of the government. A lesser degree of ignorance may be represented by assignment uncertainty, where the government is aware of the distribution of the characteristics in the population, but is unable to assign a particular value to a particular individual. Under assignment uncertainty it is possible to carry out a policy optimisation, although the impact of this on any particular individual remains uncertain. In the usual distinction between 'risk' and 'uncertainty' (Knight (1921)), assignment uncertainty corresponds to a calculated risk as to the effects on particular individuals, while full ignorance corresponds to genuine uncertainty. The optimisations in the rest of this chapter are all dependent on assignment uncertainty to allow them to proceed, ruling out a complete lack of information. When the characteristics distribution is known, it is normally necessary to assume that the number of households exceeds the number of distinct characteristics; otherwise characteristics are not in any way summarising information (one could redefine households as 'characteristics'), and the first-best is attainable given the requisite policy tools. Such an assumption is implicit in the models below, which can be seen as discussing a more limited number of 'relevant' characteristics. As was mentioned above, the number of personal characteristics in reality approaches the infinite, far exceeding the size of any population. This is perhaps fortunate for humanity: it is why no two persons are identical, and why we think of people as discrete individuals, rather than as points on a continuum. The discussion below is not on this level of complexity, concerning a more restricted set of characteristics, about which the government has considerable, but not complete, information.

The second reason for not attaining the first-best is the possible presence of disincentives, interpreted in the broad sense of any behavioural response by individuals to the government's policy measures. When income transfers are related to characteristics, individuals can benefit by changing or misrepresenting their level of the characteristic so as to increase their return from the associated payments scheme. Such behaviour would not be possible if characteristics were fixed, well-defined and freely observable by the government. In many cases, however, they will not satisfy these requirements, and will be subject to manipulation by the individuals concerned (either by a genuine change in the characteristic, or by misinforming the government about its true value). When disincentives are present, a full policy optimum will only be attained by including them in the optimisation problem, in the form of an additional set of constraints.

This would prevent the achievement of the first-best outcome given above, even if redistributive payments were related to the full set of 'observable' characteristics.

The remainder of this chapter discusses the points raised above in more detail. Section (2) sets out a basic model of redistributive income transfers related to individual characteristics, and Section (3) considers how these measures interact with optimal taxation. Section (4) extends the model of Section (2) to allow for disincentive effects arising from the manipulation of characteristics by individuals. In the final section the theoretical discussion is related to more practical issues, such as the implementation of actual social security and redistributive policies.

(2) Basic Model

The policy instruments discussed in this section are income payments/taxes related to an observable characteristic, a_h. In order to facilitate optimisation, it is assumed throughout that a_h is a continuous scalar variable, although this is not crucial to the analysis; the characteristic could alternatively take a discrete form, or even be qualitative, represented by a binary type 'dummy' variable. For a_h to be a viable basis for income transfers it should satisfy the following four requirements:

(i) It should be observable;

(ii) It should be objectively measurable;

(iii) It should either have a direct influence on utility, or be correlated with something that does (or both);

(iv) It should offer a socially acceptable means of discriminating between individuals. (Discrimination by characteristics such as race is often seen as intrinsically undesirable.)

If these conditions are satisfied, then there is nothing within the model to prevent the government from carrying out redistributive income transfers.

In part (a) below individuals vary in two respects, the wage rate, W_h, and the characteristic, a_h, on which policy is based. The format of the model reflects that in optimal taxation models, with the wage rate not directly observable by the government, but with the assumption of assignment uncertainty. Hence the government's available information is the data on the observable characteristic, a_h, and knowledge of the joint distribution of a_h and W_h. The introduction of W_h into the models is not essential, and they can be alternatively expressed in more general terms, as in part (b).

The income transfers may be related to a_h in either a linear or non-linear fashion. Although the former case is less general, it can be rationalised by appealing to the administrative difficulties associated with highly non-linear payments or taxes (they are rarely found in practice). Direct administrative costs are taken to be zero, and it is assumed that a_h is observed without error. The linear and non-linear cases are discussed successively below.

(a) Linear Payments

Let there be H individuals, h=1,...,H, (H>2) with identical preferences, and associated characteristics W_h, a_h. Following the standard practice with household composition effects, preferences can be depicted by the consumer cost function, of the form

$$C_h = C(\underline{p}, W_h, a_h, U_h)$$

where U_h is the utility level and $\underline{p}$ is the vector of consumer prices (this is equivalent to setting out the indirect utility function, incorporating

a_h). Each individual has T hours of available working time and receives no unearned income.

The government is assumed to pay out a uniform lump sum, α, to all individuals, and a payment/tax which depends linearly on a_h, βa_h, so that individual h receives $\alpha+\beta a_h$ (where α or β may be negative). Denoting the revenue requirement by R, the government's revenue constraint is

$$R + H\alpha + \beta \sum_{h=1}^{H} a_h = 0$$

where α and β are arbitrarily regarded as payments rather than taxes; it follows that at least one out of α and β must be negative to satisfy the revenue constraint for R>0. The government's problem is to choose α and β so as to maximise individualistic social welfare, $V=V(U_1,\ldots,U_H)$, subject to the revenue constraint and individual utility maximisation. The U_h terms denote the indirect utility functions of the individuals concerned, for a fixed set of prices common to all consumers. Optimising gives the following conditions:

$$L = V(U_1,\ldots,U_H) + \mu(-R-H\alpha-\beta \sum_{h=1}^{H} a_h)$$

$$\frac{\partial L}{\partial \alpha} = \sum_{h=1}^{H} \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h} - \mu H = 0 \quad (\text{since } U_h = U(\underline{p}, W_h, W_h T+\alpha+\beta a_h))$$

$$\frac{\partial L}{\partial \beta} = \sum_{h=1}^{H} \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h} \cdot a_i - \mu \sum_{h=1}^{H} a_h = 0$$

Let $\lambda_h \equiv \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h}$ = MSU of income to individual h.

The first-order conditions can then be rewritten as

$$\mu = \sum_{h=1}^{H} \frac{\lambda_h}{H} = \bar{\lambda}$$

and $$\sum_{h=1}^{H} \lambda_h a_h - \bar{\lambda} \sum_{h=1}^{H} a_h = 0$$

The first equation implies that the MSU of government revenue, μ, is equated with the mean MSU of income to individuals.

Dividing the second equation by H yields

$$\frac{1}{H} \sum_{h=1}^{H} \lambda_h a_h - \bar{\lambda}\,\bar{a} = 0 \qquad \text{where} \qquad \bar{a} = \sum_{h=1}^{H} \frac{a_h}{H}$$

i.e. $\quad \mathrm{cov}(a_h, \lambda_h) = 0$

This zero covariance condition is the basic optimality result for a linear payments scheme, and implies that at the optimum the social welfare function does not discriminate between individuals on the basis of the characteristic a_h. In other words, any systematic relationship between a_h and the social value placed on income received by individuals is eliminated. If the condition is not satisfied, then it is possible to increase social welfare by adjusting the payments scheme in favour of individuals with a high λ_h.

The nature of the optimal payments scheme is influenced by the interrelationship between W_h and a_h. Suppose, for example, that the wage rate is constant for all individuals, so that all variation is due to a_h. The effect of a_h on the MSU of income will tend to have the same sign as $\frac{\partial C}{\partial a_h}$, given that the marginal utilities of income decline with the income level. To satisfy the zero covariance condition it follows that payments must have the opposite effect to a_h on the MSU of income, implying that they are positive if $\frac{\partial C}{\partial a_h} > 0$ and negative if $\frac{\partial C}{\partial a_h} < 0$. The optimal β should in general be approximately at the level required to offset the direct effect of a_h on consumers' costs; this corresponds to the usual view of welfare benefits related to household composition (such as family allowances), where benefits are directly linked to the estimated effect of the composition variable on the household's cost of living.

In practice, however, variation in W_h is certainly present, and any correlation between W_h and a_h effectively introduces a further avenue through which a_h is seen to influence cost. For example, consider a situation in which W_h and a_h are highly positively correlated, and the effect of W_h on the cost function dominates the direct effect of a_h. In such a case, even if $\frac{\partial C}{\partial a_h} > 0$, the optimal β is liable to be negative, since a high a_h is usually accompanied by a high W_h, giving a negative net effect on the cost function. a_h has therefore become a proxy for W_h, providing a means of identifying individuals with a high income level (in this capacity it resembles the 'tagging' of Akerlof (1978)). Because almost all populations involve non-zero correlation between W_h and a_h, some distortionary effect of this nature must virtually always occur, implying that the optimal payments scheme cannot be viewed narrowly as a compensation for the immediate effects of the characteristic on costs. Indeed, it would be possible to have successful redistributive policies based solely on the 'proxy' role of characteristics which themselves have no influence on utilities.

(b) Non-linear Payments

In order to generalise the discussion in (a), it will now be assumed that there is a vector $\underline{a}$ of different individual characteristics, of which only a subset $\underline{a}_1$ is observable. $\underline{a}$ can therefore be partitioned in the form $\underline{a} = \{\underline{a}_1, \underline{a}_2\}$, and the wage rate included in (a) above would be an element in the vector of unobservable characteristics, $\underline{a}_2$. Viewing the population as a continuum, in contrast with the discrete population of (a), the distribution of individuals over the vector $\underline{a}$ can be represented by the joint density function $f(\underline{a})$. Assignment uncertainty with respect to $\underline{a}_2$ implies that the government knows the function $f(\underline{a})$, but cannot assign definite values of the vector $\underline{a}_2$ to individuals. The remaining characteristics in $\underline{a}_1$ can be used as the basis of a redistributive policy.

The government's optimisation problem is a straightforward generalisation of that in (a). Redistributive payments are based on the observed characteristics, and take the general functional form $\beta(\underline{a}_1)$. Social welfare is given by

$$\int_{\underline{a}} V(U(\underline{p},\underline{a},\beta(\underline{a}_1))f(\underline{a})d\underline{a}$$

where the summation is over the full vector $\underline{a}$, and the revenue constraint is

$$R + \int_{\underline{a}_1} \beta(\underline{a}_1)g(\underline{a}_1)d\underline{a}_1 = 0$$

with $g(\underline{a}_1)$ denoting the joint marginal density of $\underline{a}_1$. The policy problem is therefore to set the function $\beta(\underline{a}_1)$ so as to maximise social welfare subject to the revenue constraint and individual utility maximisation. The Lagrangian and first-order conditions are:

$$L = \int_{\underline{a}} V(U(\underline{p},\underline{a},\beta(\underline{a}_1))f(\underline{a})d\underline{a} + \mu(-R- \int_{\underline{a}_1} \beta(\underline{a}_1)g(\underline{a}_1)d\underline{a}_1)$$

$$\frac{\partial L}{\partial \beta(\underline{a}_1)} = \int_{\underline{a}_2} \frac{\partial V}{\partial U} \cdot \frac{\partial U}{\partial \beta(\underline{a}_1)} f(\underline{a})d\underline{a}_2 - \mu g(\underline{a}_1) = 0 \qquad \forall \underline{a}_1$$

Rearranging yields

$$\mu = \frac{1}{g(\underline{a}_1)} \int_{\underline{a}_2} \frac{\partial V}{\partial U} \cdot \frac{\partial U}{\partial \beta(\underline{a}_1)} f(\underline{a})d\underline{a}_2$$

$$= \frac{1}{g(\underline{a}_1)} \int_{\underline{a}_2} \lambda f(\underline{a})d\underline{a}_2 \qquad = \bar{\lambda}|_{\underline{a}_1} \qquad \forall \underline{a}_1$$

where $\lambda \equiv \frac{\partial V}{\partial U} \cdot \frac{\partial U}{\partial \beta}$ = MSU of income to individuals.

The optimum requires that, for any value of $\underline{a}_1$, the mean MSU of income taken over all the possible values of $\underline{a}_2$ is equal to a constant, μ (which is the MSU of government revenue). Hence, in a similar way to (a) above, any systematic relation between λ and the variables in $\underline{a}_1$ is eliminated. The outcome is analogous to the equating of MSU's which occurs at the first-best, but applies only over the subset $\underline{a}_1$ of the full vector $\underline{a}$; the first-best is therefore a special case of this model in which $\underline{a}_1=\underline{a}$.

It is not possible to say anything definite about the nature of the optimal payments schedule, which could be quite complex in form. The β function obtained will depend on the nature of individual and social preferences and on the distribution of characteristics, $f(\underline{a})$; exactly the same influences are present as in (a), so that the optimal redistributive payments are governed by the interrelationship between the observable characteristics $\underline{a}_1$ and the unobservable ones $\underline{a}_2$. In general it will not be accurate to view the optimal β as a compensatory payment for the effects of the characteristics $\underline{a}_1$ on utility.

The central feature of the models in (a) and (b) is that they make the best possible use of a limited information set. At the policy optimum there is on average no relationship between the observed characteristics and the MSU of income to individuals. The impact on any particular individual is unknown, and depends on the unobserved characteristics; those with atypical values may end up either better or worse off than the expected outcome for recipients of the same income transfer. Certain individuals could suffer badly in the interests of raising the average welfare level in the population, a situation which conflicts with the objectives of many social security policies. If the model is intended to represent welfare payments, then it may be necessary to make suitable modifications (see Section (5) below). Nevertheless, income transfers are highly efficient redistributive tools, involving no disincentive effects. The next section considers how they can be used in conjunction with (and to replace) redistributive taxes, and Section (4) discusses the manipulation of characteristics by individuals.

(3) Redistributive Income Transfers combined with Taxation

Much of the theoretical literature on optimal redistribution has concentrated on taxation rather than income transfers. The problem with taxes as policy instruments is that they may influence individual behaviour, causing a disincentive which constrains policy optimisation. Such effects do not occur with income transfers, and the optimal tax literature usually assumes that the information needed for lump-sum income redistribution is not accessible; in other words it assumes a unidimensional set of consumers with assignment uncertainty. When individuals are multidimensional (which is closer to reality), there arises the possibility that assignment uncertainty applies only to a subset of characteristics, with others directly observable. In this position it is rational for the government to implement income transfers related to the observable characteristics, in addition to taxation. The presence of income-based redistributive transfers in general leads to a different set of optimal taxes from the standard model, and the possible interaction between the different policy instruments is considered below, in part (a). A further possibility is that the tax rates themselves can vary with the observed characteristics, and this is allowed for in part (b).

(a) Non-discriminatory Taxes

The model used is that of Section (2), part (a), where H individuals differ in the wage rate, W_h and the observable characteristic, a_h,

$h=1,\ldots,H$. Taxation takes the form of linear direct and indirect taxes, as in Diamond (1975) (an extension of Ramsey (1927)). There are N commodities, with producer prices normalised to unity and the tax on employment income normalised to zero. The government's tax instruments are denoted by t_i, $i=1,\ldots,N$, so that consumer prices are equal to $1+t_i$, $i=1,\ldots,N$. In addition to the tax rates, there is also a set of income transfers based linearly on a_h, of the form taken in Section (2), part (a), with each individual receiving $\alpha+\beta a_h$. The appropriate revenue constraint is

$$R + H\alpha + \beta \sum_{h=1}^{H} a_h = \sum_{h=1}^{H} \sum_{i=1}^{N} t_i x_i^h$$

where x_i^h is the consumption of the i^{th} commodity by the h^{th} individual. The optimal policy problem is therefore to set $t_1,\ldots,t_N$, α and β to maximise social welfare, subject to the revenue constraint and individual utility maximisation. This is soluble by the Lagrangian method, as below:

$$L = V(U_1,\ldots,U_H) + \mu\left(\sum_{h=1}^{H} \sum_{i=1}^{N} t_i x_i^h - R - H\alpha - \beta \sum_{h=1}^{H} a_h\right)$$

$$\frac{\partial L}{\partial \alpha} = \sum_{h=1}^{H} \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h} - \mu \sum_{h=1}^{H} \sum_{i=1}^{N} t_i . \frac{\partial x_i^h}{\partial M_h} - \mu H = 0$$

$$\frac{\partial L}{\partial \beta} = \sum_{h=1}^{H} \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h} a_h + \mu \sum_{h=1}^{H} \sum_{i=1}^{N} t_i . \frac{\partial x_i^h}{\partial M_h} \cdot a_h - \mu \sum_{h=1}^{H} a_h = 0$$

$$\frac{\partial L}{\partial t_i} = \sum_{h=1}^{H} \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial t_i} + \mu \left[\sum_{h=1}^{H} x_i^h + \sum_{h=1}^{H} \sum_{j=1}^{N} t_j \frac{\partial x_j^h}{\partial t_i} \right]$$

$$= - \sum_{h=1}^{H} \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h} \cdot x_i^h + \mu \left[\sum_{h=1}^{H} x_i^h + \sum_{h=1}^{H} \sum_{j=1}^{N} t_j \left(S_{ji} - x_i \frac{\partial x_j^h}{\partial M_h} \right) \right] = 0$$

substituting from Roy's identity and the Slutsky equation (where S_{ji} is the j, ith element of the Slutsky matrix).

Let $\lambda_h \equiv \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial C_h} + \mu \sum_{i=1}^{N} t_i \frac{\partial x_i^h}{\partial M_h}$ = net MSU of income to individual h

Rearranging the optimality conditions gives

$$\mu = \sum_{h=1}^{H} \frac{\lambda_h}{H} = \overline{\lambda}$$

$$\sum_{h=1}^{H} \lambda_h a_h - \overline{\lambda} \sum_{h=1}^{H} a_h = 0 \qquad \Rightarrow \qquad \text{cov}\,(\lambda_h, a_h) = 0$$

$$\sum_{j=1}^{N} t_j \overline{S}_{ij} = \frac{1}{H} \sum_{h=1}^{H} \left(\frac{\lambda_h}{\overline{\lambda}}\right) x_i^h - \overline{x}_i \qquad \forall i$$

where a bar denotes taking mean over h.

The zero covariance condition and the optimal tax equation are unchanged from the separate optimisation problems, but they must now hold simultaneously (S_{ij}, x_i^h and λ_h are in general functions of a_h). As is usual with optimal tax models, it is not possible to generalise about the resulting tax rates, although it is clear that they will not usually coincide with the outcome of the optimal tax problems treated in isolation.

An impression of the interrelationship between redistributive income payments and taxation can be gained by considering models with demand functions linear in a_h. The hypothetical extreme case is where the cost function satisfies

$$\frac{\partial C}{\partial p_i} = \chi_i = \gamma_i(\underline{p}) + \delta_i(\underline{p})a_h \qquad i=1,\ldots,N$$

implying that the compensated demands are the same linear function of a_h for all individuals. Substitution in the tax formulae above establishes that

$$\sum_{j=1}^{N} t_j \bar{S}_{ij} = 0 \qquad i=1,\ldots,N$$

so that there is no need for taxation. This confirms that whenever consumption patterns and the observable characteristic provide identical information on inequality, it is preferable to use income transfers in place of taxation as they involve no efficiency cost. A slightly more general case is a cost function satisfying

$$\frac{\partial C^h}{\partial p_i} = x_i^h = \gamma_i^h(U_h, \underline{p}, W_h) + \delta_i(\underline{p})a_h \qquad i=1,\ldots,N$$

that is, where the compensated demands vary in an identical linear fashion with a_h, but are otherwise general. Here the tax equation reduces to

$$\sum_{j=1}^{N} t_j \bar{S}_{ij} = \frac{1}{H} \sum_{h=1}^{H} \left(\frac{\lambda_h}{\bar{\lambda}}\right) \gamma_i^h - \bar{\gamma}_i \qquad i=1,\ldots,N$$

so that taxation is offsetting only that part of the inequality in the consumption of the i^{th} commodity which is not due to a_h. Assuming separability of goods from leisure and parallel linear Engel curves gives the uniform commodity tax result of Deaton and Stern (1986).

With more complex and realistic demand functions the position is less straightforward, but as a rule one would expect both the proportionate reduction in compensated demand and the tax rate to be lower on goods whose demand is significantly influenced by a_h. The intuitive reason for this is that the more efficient payments scheme takes over the redistributive role of the taxation of goods closely associated with a_h, leaving taxation to mitigate the residual inequality unconnected with a_h. It should be remembered that the 'proxy' effect of correlation between a_h and W_h is still present, and that, if this dominates the direct effects of a_h, the remarks above will apply to W_h rather than a_h. For instance, with positive correlation the effect of the benefit scheme is likely to be to reduce the differential taxation of goods associated with W_h (viz. luxuries), and instead to penalise a_h by setting a negative β. It remains possible to base payment schemes entirely on the 'proxy' effect, using characteristics which have no independent influence on utility.

(b) Discriminatory Taxes

Besides income transfers, information on observed characteristics can also serve as the basis of discriminatory taxation. By imposing a different level and pattern of commodity taxes for different groups in the population, the government can increase the number of policy instruments and thereby (in general) improve on the optimum of part (a). A government would rationally seek to introduce consumer price discrimination wherever feasible, although in practice the scope is likely to be rather limited. The obstacles to such a policy resemble those facing price discrimination by firms: in particular, it must be possible to enforce the tax rates in question, and prevent the resale of goods between individuals. An additional difficulty is that tax discrimination might be opposed on principle as contravening the spirit of 'fair' taxation (despite its redistributive intent, it appears at face value to break Adam Smith's first canon of taxation). Taxes based partly on observed characteristics and partly on individual consumption decisions do not fit easily into the usual direct/indirect classification, and do not feature in existing tax policies. These problems notwithstanding, it remains the case that tax discrimination can contribute to redistributive policy, and a discriminatory tax optimum is set out below.

Let the model take the same form as in part (a), except that taxes are now differentiated on the basis of the observed characteristic a_h. Instead of assuming a continuous set of tax rates defined as a function of a_h, it seems more reasonable to have a finite number of tax regimes based on groupings of the parameter a_h. Suppose, therefore, that a_h values can be assigned to one of R sets, Y_r, r=1,...,R, each containing H_r individuals, such that $\sum_r H_r = H$. If all commodities are included in tax discrimination, then a separate set of linear tax rates is levied on each Y_r set, increasing the number of tax instruments from N to RN; in practice discrimination might be restricted to a subset of commodities, producing a smaller number of tax rates. An optimum is derivable in a similar way to part (a), and linear income transfers (where present) lead to the same zero covariance condition

$$\mathrm{cov}(\lambda_h, a_h) = 0$$

The only difference is in the optimal tax equations, which become

$$\sum_{j=1}^{N} t_j^r \bar{S}_{ij}\Big|_{h\varepsilon Y_r} = \frac{1}{H_r}\sum_{h\varepsilon Y_r}\left(\frac{\lambda_h}{\bar{\lambda}}\right)x_i^h - \bar{x}_i\Big|_{h\varepsilon Y_r} \qquad \begin{matrix} r=1,\ldots,R \\ i=1,\ldots,N \end{matrix}$$

where $\bar{\lambda}$ is still evaluated over the whole population. For each of the R sets of tax rates enforced, the tax equations have an interpretation similar to that for the standard non-discriminatory case. The compensated reduction in demand (on the left) is greater, the lower is the (negative) correlation between λ_h and consumption of the commodity by the relevant Y_r set (on the right); 'luxuries' will tend to be taxed more highly than other goods. With tax discrimination, however, this relation is tailored according to observed a_h values, allowing alternative tax treatment of different groups in the population. A good can potentially be a 'luxury' for one Y_r set but not for another, leading to variations in tax treatment. Hence, although the nature of optimal taxation is not radically changed, it is now localised to smaller sub-populations, allowing an increase in its effectiveness.

There are two main situations in which tax discrimination is likely to be useful:

(i) Under certain individual preferences the value of λ_h may be strongly influenced by other individual characteristics, in addition to income levels. This will tend to hamper the ability of income transfers to reduce variation in λ_h, leaving scope for other redistributive tools. Tax discrimination based on a_h provides another means by which information on a_h can be brought to bear on λ_h, and thus may be able to improve on income transfers alone.

(ii) Discriminatory taxes also become valuable when relative consumption levels vary significantly with a_h (either directly or via 'proxy' effects). The conventional tax optimum can distinguish appropriate tax rates for the whole population, but cannot guarantee that these are equally appropriate to sub groups.

As in part (a), nothing can be said in general about tax rates at the optimum. The importance of discrimination based on a_h depends on the importance of a_h, both as an influence on individual preferences and as a source of information on other characteristics. Where a_h does play a significant role, there are no grounds to expect discriminatory tax regimes to resemble each other: effectively the comparison is between two separate tax optima. The same is true for comparison between discriminatory tax regimes and the standard tax optimum, where no systematic relationship can be expected.

Differing standard and discriminatory tax optima are illustrated by the following algebraic example:

Example

Let the characteristic a_h take two possible values, $\tilde{a}$, $a^\dagger$, such that individuals h=1,...,G have $\tilde{a}$ and h = G+1,...,H have $a^\dagger$, where G<H. The only other characteristic is the wage rate, W_h, which is freely variable over the population.

The government wishes to implement optimal discriminatory taxes, $\tilde{t}_i$, $t^\dagger_i$, for the two commodities i=1,2, coupled with a uniform lump-sum payment, α. Social preferences are utilitarian, such that $V=\Sigma U_h$.

Individual utilities take the form

$$U_h = a_h \ln x_1^h + \ln x_2^h + l$$

where l is leisure. The resulting demand functions are

$$x_1^h = \frac{a_h W_h}{1+t_1}, \quad x_2^h = \frac{W_h}{1+t_2}, \quad l_h = (T-a_h-1) + \frac{\alpha}{W_h}$$

where l>0 is assumed to hold. These preferences simplify calculation of a tax optimum, since income and cross-substitution effects are zero; they also increase the influence of relative prices on consumption patterns, and thus enhance the role of tax discrimination.

Calculating the optimal tax rates for the standard and discriminatory cases yields the following:

Standard tax optimum

$$\left.\begin{aligned} \frac{t_1}{1+t_1} &= 1 - \frac{1}{\bar{\lambda}}\left(\frac{\sum_{h=1}^{H} a_h}{\sum_{h=1}^{H} a_h W_h}\right) \\ \frac{t_2}{1+t_2} &= 1 - \frac{1}{\bar{\lambda}}\left(\frac{H}{\sum_{h=1}^{H} W_h}\right) \end{aligned}\right\} \quad t_1 \gtreqless t_2 \text{ as } \mathrm{cov}(a_h, W_h) \gtreqless 0$$

$$\text{where } \bar{\lambda} = \frac{1}{H}\sum_{h=1}^{H}\left(\frac{1}{W_h}\right) \quad = \text{ MSU of government revenue}$$

$$= \text{ mean MSU of income to individuals}$$

Discriminatory tax optimum

$$\left.\begin{aligned} \frac{\tilde{t}_1}{1+\tilde{t}_1} &= \frac{\tilde{t}_2}{1+\tilde{t}_2} = \frac{\tilde{t}}{1+\tilde{t}} = 1 - \frac{1}{\bar{\lambda}}\left(\frac{G}{\sum_{h=1}^{G} W_h}\right) \\ \frac{t_1^\dagger}{1+t_1^\dagger} &= \frac{t_2^\dagger}{1+t_2^\dagger} = \frac{t^\dagger}{1+t^\dagger} = 1 - \frac{1}{\bar{\lambda}}\left(\frac{H-G}{\sum_{h=G+1}^{H} W_h}\right) \end{aligned}\right\} \quad \begin{aligned} &\tilde{t} \gtreqless t^\dagger \quad \text{as} \\ &\tilde{W} \equiv \frac{\sum_{h=1}^{G} W_h}{G} \gtreqless \frac{\sum_{h=G+1}^{H} W_h}{H-G} \equiv W^\dagger \end{aligned}$$

The government in this example wishes to tax the unobserved W_h, on which the population's inequality is based. Because consumption patterns vary with a_h, this can be achieved at the standard tax optimum by setting differentiated tax rates. Consumption of the first good increases with a_h, so $t_1 > t_2$ if a_h happens to be positively correlated with W_h. At the discriminatory optimum, however, it is possible to link tax rates directly to a, removing the need to have differential taxation of commodities. The outcome is two uniform tax regimes, in which the higher tax rate falls on the group with the higher average wage. Hence tax discrimination in this case permits non-uniform redistributive commodity taxes to be replaced by separate uniform tax regimes for each a_h group.

One can conclude that there will in general be a case for tax discrimination wherever it is feasible. This is simply a matter of using information to the best advantage, and it is rational to base as many policy instruments as possible on observed characteristics. The practical value of tax discrimination is rather less apparent. Compared with, say, income transfers, it is likely to be difficult to enforce and administer on a large scale: feasible schemes are probably limited to a few commodities and a small number of tax regimes. Even when discrimination is possible it faces a further hurdle of public acceptability, as it appears to break conventional notions of horizontal equity.

This section has argued that standard tax optima can be improved on by using directly observed information, as the basis of income transfers or tax discrimination. The impact of redistributive income transfers is especially significant in this respect: it is difficult to generalise about the interrelationship between policy tools, but the presence of income transfers can be expected to lessen the general level of redistributive taxation at the social optimum. Although commodity and income taxes remain valuable redistributive tools, their relative importance is somewhat overstated by the usual formulation of optimal taxation models.

(4) Redistributive Income Transfers based on Variable Characteristics

The previous two sections have assumed that the characteristic in question is outside the control of the individual and accurately observable by the government. When these conditions are satisfied the individual cannot manipulate the characteristic so as to increase net receipts from redistributive transfers. In practice, however, manipulation may take place, and its effects on the policy optimum are considered in this section.

The model is again that of Section (2), part (a), where individuals vary in the wage rate, W_h, and an observable characteristic, a_h, and receive a payment $\alpha+\beta a_h$. There are two main cases to be considered, depending on whether the true value or the reported value of the characteristic is altered.

(a) True Value

In this case the characteristic is not rigidly fixed, and can be altered by the individual in response to government policy. Any such change is liable to involve a cost to the individual (monetary, psychological, etc.), which is weighed against the additional income received. The government, given that it is aware of the individual responses taking place, must allow for them in its policy formulation if it is to achieve a full social optimum. An example of this type of characteristic is family size, when it forms the basis of family allowance payments. One possible view is that family size is not very responsive to monetary incentives, and that it can be treated as a fixed characteristic in the manner of previous sections; this presumably is not far from the truth in a downward direction for an already existing family. However, monetary payments have sometimes been used in the attempt to influence family size in the long run, and some theoretical discussions of family allowances have been based on their population effects (see, for example, Mirrlees (1972) and Cigno (1983)). If one accepts the latter view, then it is necessary to incorporate the family size responses in any optimisation model.

In order to include these effects in the theoretical model, it is assumed that the characteristic a_h is one of the individual's decision variables, implying that W_h is the only exogenously fixed characteristic. Preferences are defined by the indirect utility function.

$$U_h = U(\underline{p}, W_h, a_h, W_h T+\alpha+\beta a_h)$$

and a_h is chosen by the individual to maximise this. If $\hat{a}_h$ denotes the optimalchoice of a_h, then it satisfies

$$\frac{\partial U_h}{\partial \hat{a}_h} + \beta \frac{\partial U_h}{\partial M_h} = 0 \qquad h=1,\ldots,H$$

and can be expressed as $\hat{a}_h = \hat{a}(\underline{p}, W_h, W_h T, \alpha, \beta)$. For the manipulation problem to be meaningful it is necessary to assume that this optimum exists. Letting $\hat{U}_h$ denote the associated utility level, it follows that

$$\hat{U}_h = U(\underline{p}, W_h, \hat{a}_h, W_h T+\alpha+\beta\hat{a}_h) = \hat{U}(\underline{p}, W_h, W_h T, \alpha, \beta)$$

where $\hat{U}$ is a modified form of the indirect utility function incorporating the effects of variation in a_h under the linear payments scheme. Given this situation, the government can undertake its policy optimisation in a similar way to Section (2), but allowing for the effects of α and β on $\hat{a}_h$. The Lagrangian and first-order conditions are:

$$L = V(U_1,\ldots,U_H) + \mu(-R-H\alpha-\beta\sum_{h=1}^{H}\hat{a}_h)$$

$$\frac{\partial L}{\partial \alpha} = \sum_{h=1}^{H}\frac{\partial V}{\partial U_h}\left[\frac{\partial U_h}{\partial M_h} + \left(\frac{\partial U_h}{\partial \hat{a}_h} + \beta\frac{\partial U_h}{\partial M_h}\right)\frac{\partial \hat{a}_h}{\partial M_h}\right] - \mu H - \mu\beta\sum_{h=1}^{H}\frac{\partial \hat{a}_h}{\partial M_h}$$

$$= \sum_{h=1}^{H}\frac{\partial V}{\partial U_h}\cdot\frac{\partial U_h}{\partial M_h} - \mu H - \mu\beta\sum_{h=1}^{H}\frac{\partial \hat{a}_h}{\partial M_h} = 0$$

$$\frac{\partial L}{\partial \beta} = \sum_{h=1}^{H}\frac{\partial V}{\partial U_h}\left[\frac{\partial U_h}{\partial M_h}\cdot\hat{a}_h + \left(\frac{\partial U_h}{\partial \hat{a}_h} + \beta\frac{\partial U_h}{\partial M_h}\right)\frac{\partial \hat{a}_h}{\partial \beta}\right]$$

$$- \mu\left[\sum_{h=1}^{H}\hat{a}_h + \beta\sum_{h=1}^{H}\frac{\partial \hat{a}_h}{\partial \beta}\right]$$

$$= \sum_{h=1}^{H}\frac{\partial V}{\partial U_h}\cdot\frac{\partial U_h}{\partial M_h}\cdot\hat{a}_h - \mu\left[\sum_{h=1}^{H}\hat{a}_h + \beta\sum_{h=1}^{H}\frac{\partial \hat{a}_h}{\partial \beta}\right] = 0$$

(using the above individual optimality condition for the choice of a_h).

Let λ_h denote the net MSU of income to individual h, so that

$$\lambda_h = \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h} - \mu\beta\frac{\partial \hat{a}_h}{\partial M_h}$$

This differs from its equivalent in Section (2) in that it is a net expression involving an extra term to represent the effects on government revenue of the induced changes in $\hat{a}_h$.

Consider also the term $\frac{\partial \hat{a}_h}{\partial \beta}$ in the equation $\frac{\partial L}{\partial \beta} = 0$. If β is viewed as (minus) the 'price' attached to $\hat{a}_h$, then this resembles an uncompensated price derivative in demand theory, and can be decomposed in a similar way. Thus, noting that $\hat{a}_h$ is expressible in the particular form $\hat{a}_h = \hat{a}(\underline{p}, W_h, \beta, W_h T + \alpha + \beta\hat{a}_h)$, one can write

$$\frac{\partial \hat{a}_h}{\partial \beta} = \left.\frac{\partial \hat{a}_h}{\partial \beta}\right|_{U_h} + \hat{a}_h \frac{\partial \hat{a}_h}{\partial M_h} \qquad h=1,\ldots H$$

where the first term is the positive 'substitution' effect, holding income and utility constant, and the second term is the income effect, which could be positive or negative.

Using these expressions for λ_h and $\frac{\partial \hat{a}_h}{\partial \beta}$, the first-order conditions become

$$\mu = \sum_{h=1}^{H} \frac{\lambda_h}{H} = \bar{\lambda}$$

and $$\sum_{h=1}^{H} \lambda_h \hat{a}_h - \bar{\lambda} \sum_{h=1}^{H} \hat{a}_h = \mu\beta \sum_{h=1}^{H} \left.\frac{\partial \hat{a}_h}{\partial \beta}\right|_{U_h}$$

$\Rightarrow$ $\mathrm{cov}(\lambda_h, \hat{a}_h)$ = mean MSU of revenue losses from compensated responses of $\hat{a}_h$ to β

The equation $\mu=\overline{\lambda}$ is unchanged from the standard case, apart from the revised definition of λ_h. The covariance condition differs from the standard case in the presence of the extra term $\mu\beta \sum_{h=1}^{H} \frac{\partial \hat{a}_h}{\partial \beta}\bigg|_{U_h}$, which determines whether or not λ and $\hat{a}$ are positively or negatively correlated. It is known that the Lagrange multiplier, μ, denoting the MSU of government revenue is positive. One can also say that the compensated adjustment $\frac{\partial \hat{a}_h}{\partial \beta}\bigg|_{U_h}$ is positive, implying that an increase in the return to the characteristic leads to a rise in its level (although the net response could be negative, via the income effect). Hence the sign of $\text{cov}(\lambda_h,\hat{a}_h)$ depends entirely on the sign of β, in such a way that

$$\text{cov}(\lambda_h,\hat{a}_h) \gtreqless 0 \qquad \text{as} \qquad \beta \gtreqless 0$$

This outcome can be explained intuitively as follows. If β is positive, then the characteristic a_h is being favoured by the redistributive policy, and the direct adjustment effect $\frac{\partial \hat{a}_h}{\partial \beta}\bigg|_U$ causes an increase in $\hat{a}_h$. This 'disincentive' is an obstacle to redistribution, resulting in a failure to eliminate totally the positive correlation between λ_h and $\hat{a}_h$, so that $\hat{a}_h$ continues to be indicative of the less well off (those with a high value of λ_h). In other words, for β to be positive a high $\hat{a}_h$ must have been associated on average with the relatively disadvantaged in the no policy situation, and the redistributive policy has not been able to offset this relationship entirely. When β is negative exactly the opposite is true, so that $\hat{a}_h$ is being penalised, the adjustment $\frac{\partial \hat{a}_h}{\partial \beta}\bigg|_U$ leads to a fall in $\hat{a}_h$ (compared with the status quo), and redistribution does not entirely remove the favoured position of those with a high $\hat{a}_h$. The desired zero covariance condition of Section (2) is not achieved because of the disincentive constraint arising from variation in $\hat{a}_h$.

The case discussed here is rendered slightly artificial by the presence of only two characteristics, W_h and a_h, of which only W_h is exogenously determined. Individuals are therefore essentially unidimensional, which removes any informational deficiencies and implies that policy is constrained only by the movements of $\hat{a}_h$ and the linearity of the income transfers. In models with more than two variables, however, the same informational considerations apply as in Section (2), with $\hat{a}_h$ providing an imperfect indication of the levels of the other characteristics.

(b) Reported Value

Even if the true value of a characteristic is fixed, it may still be possible for individuals to increase their net income from redistribution by altering the value reported to the government. This presupposes that the government cannot directly observe the characteristic, or can only do so at an excessive cost, a situation which may well be a reasonable representation of certain actual characteristics. If the government is totally unaware of manipulation, then it will proceed to design its policy on the lines of the previous sections but using the misreported values of characteristics; this will generally lead to a suboptimal outcome, with some individuals possibly better off than at the true optimum, but society as a whole worse off. Alternatively it may be that the government is aware of the presence and extent of manipulation, without being able to prevent its occurrence. In this situation the manipulation is an additional constraint on policy formulation, which can be included in the formal optimisation problem. As in (a), the act of false reporting of characteristics probably involves a cost to the individual, to be set against the monetary returns. In practice the misrepresenting of personal characteristics to the taxation or social security authorities is usually declared illegal, so the potential costs to individuals of being discovered are quite high.

In order to represent this theoretically it is necessary to specify the costs involved in manipulation of the characteristic (a similar situation is considered in Section (3) of Roberts (1984)). Unlike (a), there is no direct effect on utility through movements in the true value of the characteristic, and it is assumed here that the cost of manipulation can be expressed in monetary terms. Let the money cost of manipulation be given by $g[(a_h-\hat{a}_h)^2]$, where a_h is the true characteristic value, $\hat{a}_h$ is the reported value, and g is an increasing function (here taken to be common to all individuals, although this does not have to be assumed). The cost involved therefore rises with the size of the absolute deviation of the reported value from the true value of the characteristic. Since welfare payments are based on $\hat{a}_h$, an individual's utility can be expressed as

$$U_h = U(\underline{p}, W_h, a_h, WT+\alpha+\beta\hat{a}_h - g[(a_h-\hat{a}_h)^2]$$

so that $g[(a_h-\hat{a}_h)^2]$ is a compensating variation measure of the costs of manipulation. The optimal $\hat{a}_h$ is chosen to maximise net income, satisfying

$$\frac{\partial M_h}{\partial \hat{a}_h} = \beta + 2g'[(a_h-\hat{a}_h)^2]\ (a_h-\hat{a}_h) = 0$$

so that $\beta = -2g'[(a_h-\hat{a}_h)^2]\ (a_h-\hat{a}_h) \qquad h=1,\ldots,H$

and the return from the welfare payments is equated with the marginal cost of adjusting $\hat{a}_h$. From this equation it is clear that

$$\hat{a}_h \gtreqless a_h \qquad \text{as} \qquad \beta \gtreqless 0$$

which means that the characteristic is exaggerated when it is favoured by the redistribution scheme and understated when it is penalised. Since $\hat{a}_h$ is an increasing function of β alone, the relation can be summarised as $\hat{a}_h = \hat{a}_h(\beta)$, where $\frac{\partial \hat{a}_h}{\partial \beta} > 0$, h=1, ...,H, and the function $\hat{a}_h$ depends on the (fixed) true characteristic value, a_h. The government's policy problem proceeds as in Section (2), but allowing for the $\hat{a}_h(\beta)$ relationship. The Lagrangian and optimality conditions are:

$$L = V(U_1,\ldots,U_H) + \mu(-R-H\alpha-\beta\sum_{h=1}^{H}\hat{a}_h(\beta))$$

$$\frac{\partial L}{\partial \alpha} = \sum_{h=1}^{H}\frac{\partial V}{\partial U_h}\cdot\frac{\partial U_h}{\partial M_h} - \mu H = 0$$

$$\frac{\partial L}{\partial \beta} = \sum_{h=1}^{H}\frac{\partial V}{\partial U_h}\cdot\frac{\partial U_h}{\partial M_h}\cdot\hat{a}_h - \mu\sum_{h=1}^{H}\hat{a}_h - \mu\beta\sum_{h=1}^{H}\frac{\partial \hat{a}_h}{\partial \beta} = 0$$

Following Section (2), one can define

$$\lambda_h = \frac{\partial V}{\partial U_h}\cdot\frac{\partial U_h}{\partial M_h} = \text{MSU of income to individual } h$$

so that $\mu = \sum_{h=1}^{H}\frac{\lambda_h}{H} = \bar{\lambda}$

and $\sum_{h=1}^{H}\lambda_h\hat{a}_h - \bar{\lambda}\sum_{h=1}^{H}\hat{a}_h = \mu\beta\sum_{h=1}^{H}\frac{\partial \hat{a}_h}{\partial \beta}$

Hence the covariance between λ_h and $\hat{a}_h$ at the optimum depends on β in such a way that

$$\text{cov}(\lambda_h,\hat{a}_h) \gtreqless 0 \qquad \text{as} \qquad \beta \gtreqless 0$$

The outcome is similar to that in (a), and has a similar interpretation, namely that the adjustments in individual behaviour prevent redistributive policy from reaching the desired zero covariance condition. The chief points of difference are that the optimality conditions are in terms of the reported values, $\hat{a}_h$, rather than the true values, a_h, and that, unlike (a), there is no dependence of $\hat{a}_h$ on income, so that revenue effects do not appear in the expression for λ_h. Again the essentially unidimensional nature of this particular example does not carry over to models with a greater number of characteristics.

For actual examples of cases (a) and (b) there will generally be a time lag involved in an individual's adjustment to a policy change. If adjustment is fast and the policy is well anticipated, then a static model constrained by the individual's response is the appropriate one to use. On the other hand, if the response is slow and/or the policy is not anticipated, then in the short run the government may be justified in ignoring the individual behavioural constraint when designing its initial policy (although possibly making suitable changes later on). The distinction between models with and without variable characteristics may not always be as clear cut as suggested, perhaps in some cases requiring a more unified view. Certainly for situations where a protracted individual adjustment to the policy is a central element in its total impact (as could be the case with, say, family allowances), it would be desirable to set up an ad hoc intertemporal model of optimal policy.

In conclusion, the ability of individuals to vary their observed characteristics presents an additional barrier to redistributive policies; indeed, if undetectable it stops exact policy optimisation from taking place. When the government knows the extent of individual responses, however, it can adjust its policy formulation accordingly, yielding constrained optima on the lines of those above.

(5) Policy Applications

In practice the closest equivalents to the policy instruments discussed in this chapter are the monetary payments included in social security systems (such as family allowances, pensions etc.). It is consequently worth considering whether theoretical models of redistribution can be viewed as representations of, or prescriptions for, actual welfare policy. This section considers some of the issues which arise when trying to relate redistributive theory to practice.

The basis of the models is the inclusion of a vector of characteristics $\underline{a}$ in the utility functions of the population. Merely defining and quantifying $\underline{a}$ leads to difficulties, as mentioned in Section (1), and knowledge of the distribution of $\underline{a}$ is also necessary before policy optimisation can occur. These would be real problems in any policy application, and could probably be 'solved' only in an approximate manner. Assuming that $\underline{a}$ is adequately defined, it then has to be introduced into preferences, an issue which has received relatively little attention in the economic literature (presumably because it concerns utility effects which are not based on consumption of material goods and fall outside the traditional subject area of economics). The matter is mainly raised in the context of household composition effects, with comparisons made by index numbers termed 'equivalence scales' (see Deaton and Muellbauer (1980), Chapter 8). The main types of equivalence scale are based on either income or price effects, and their impact on redistributive policy optima can be summarised as below:

(i) Income Effects : Engel Equivalence Scales

This is the simplest case, originating in the observations of Engel, and estimated, for example, by Muellbauer (1977). Preferences are assumed to take the form

$$C(\underline{p},W_h,a_h,U_h) = k(a_h)\ C^*(\underline{p},W_h,U_h)$$

where $k(a_h)$ is a multiplicative constant dependent on a_h, and U_h is a per capita value; the impact of the characteristic a_h is then limited to an income effect, raising or lowering the cost of attaining a given utility level. Introducing this functional form into the tax/benefit optimum of

Section (3), part (a), yields the following. The MSU of income to household h is such that

$$\lambda_h(a_h) = \frac{\partial V}{\partial U_h} \cdot k(a_h) \frac{\partial U^*(\underline{p}, W_h, \frac{M_h}{k(a_h)})}{\partial M_h} + \sum_{i=1}^{N} t_i k(a_h) \frac{\partial x_i^h(\underline{p}, W_h, \frac{M_h}{k(a_h)})}{\partial M_h}$$

$$= \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h} + \sum_{i=1}^{N} t_i \frac{\partial x_i^h}{\partial M_h} = \lambda_h$$

and hence λ_h is not a function of a_h. The zero covariance condition remains unchanged from Section (3), but in this case must be seen as removing the indirect association between λ_h and a_h, given that there is no systematic functional link between them. Adapting the general tax equations of Section (3), part (a), to the present case yields

$$\frac{1}{H} \sum_{j=1}^{N} t_j \left(\sum_{h=1}^{H} k(a_h) S_{ij}^{*h} \right) = \frac{1}{H} \sum_{h=1}^{H} \left(\frac{\lambda_h}{\bar{\lambda}} \right) k(a_h) x_i^{*h} - \bar{x}_i \qquad \forall i$$

where S_{ij}^{*h} and x_i^{*h} are defined in per capita terms. Although a_h explicitly appears in these equations, it does so in a relatively simple way, based solely on income effects. An Engel equivalence scale therefore prevents a_h from having much visible impact on the optimality conditions, but still leaves room for strong 'proxy' effects, which would be apparent in the tax rates obtained rather than the tax equations.

A modification of the above is termed the Prais-Houthakker equivalence scale by Deaton and Muellbauer (1980), and is used in Prais and Houthakker (1955) and Muellbauer (1980). This allows for a series of functions $k_i(a_i)$, i=1,...,N, varying over commodities, but simultaneously forces price substitution effects to be zero. It is less restrictive to define equivalence scales based on prices, as in (ii).

(ii) Price Effects : Barten Equivalence Scales

The main alternative method is to define separate equivalence scales acting on the price of each commodity, as originally suggested by Barten (1964) and applied by Muellbauer (1977). Extensions of this approach are the models of Gorman (1976) and Pollak and Wales (1979). With a standard set of Barten equivalence scales, preferences are such that

$$C_h = C(p_1 m_1(a_h), \ldots, p_N m_N(a_h), W_h, U_h)$$

where the functions $m_i(a_h)$, $i=1,\ldots,N$, represent the impact of a_h on the consumption of individual commodities. These preferences leave the functional form of λ_h and the zero covariance condition unchanged from Section (3), although prices are now adjusted by the relevant $m_i(a_h)$ factors. The associated optimal tax equations are

$$\frac{1}{H} \sum_{j=1}^{N} t_j \left(\sum_{h=1}^{H} m_i(a_h) m_j(a_h) S_{ij}^{*h} \right) = \frac{1}{H} \sum_{h=1}^{H} \left(\frac{\lambda_h}{\bar{\lambda}} \right) m_i(a_h) x_i^{*h} - \bar{x}_i \qquad \forall i$$

where S_{ij}^{*h}, x_i^{*h} are again per capita values, calculated at the adjusted prices $p_i^* = m_i(a_h) p_i$. Compared with (i) it can be seen that the composition effects now enter the optimality conditions in a slightly more complex fashion, allowing for interaction with the substitution effects on the left-hand side of the equation.

The two cases above are merely specific examples of how a_h can enter the functional form of the optimality conditions. Their main significance is that they are simple enough to permit empirical estimation of household composition (and related) effects. When such estimates are available, it is feasible to calculate optimal tax/benefit rates, and thus to derive a numerical solution to the government's policy problem. It is sometimes claimed that this is the purpose of optimal tax theory (Deaton (1981 a)) whereas others take a more cautious line (Atkinson and Stiglitz (1980),

p. 12). Considering the (inevitable) gulf between reality and theory, it seems unrealistic to expect theoretical modelling to produce meaningful policy prescriptions. If this ambition is abandoned, the particular cases in (i) and (ii) above lose much of their attraction, and it is preferable to return to the more general formulations of previous sections.

Beyond the question of modelling a, there is a need to reconcile the optimisation present in theoretical models with the objectives apparently attached to existing welfare policies. In the models of this chapter the government is assumed to decide its policy measures on the basis of the constrained maximisation of a social welfare function. The implication is that the objective is truly redistributive, aiming for a comprehensive readjustment of the income distribution at all levels. In reality, despite the occasional labelling of some taxes as redistributive, welfare policies based on income payments are rarely framed in these terms. Instead they are often described as 'social security', with the more cautious objective of preventing hardship at the bottom end of the income distribution; in other words, of guaranteeing a certain minimum income or utility level. It is rather difficult to explain this in terms of policy optimisation. There are two possibilities: that the welfare policy is an accurate reflection of the true social welfare function, or that the government is failing to implement an optimal policy, whether by irrationality or by a fundamental set of constraints (institutional, informational, etc.). The former case is conceivable, but it would be a strange view of social welfare, at odds with the forms of social welfare function which are usually put forward. It is more plausible to suppose that actual policy is subject to additional complexities and constraints above those featured in the theoretical analysis. A prime example is the political process to which policy making is subjected, and the fact that governments do not always act in the best interests of society. Accepting the presence of additional constraints allows the usual picture of social welfare to be retained, while regarding the simple theoretical models of redistribution as an inadequate representation of actual policy formulation. It remains possible that the model offers a potential policy prescription, the fate of which depends on whether the additional constraints on policy are removable or genuine facts of economic life. The basic point here is that there is a marked contrast between the theoretical efficiency of income transfers as redistributive tools and their inhibited use in actual welfare policy; this suggests that either the theory or policy (or probably both) are at fault and that the true social optimum lies somewhere between their outcomes.

In view of these difficulties it is not really possible to regard the redistributive models above as representations of social security payments in the usual sense of the term. Nevertheless, it is worth noting that (as remarked in Section (2)) the policy optima covered so far do not guarantee any minimum income or utility level,often an explicit objective of social security policy. In particular, if any initially disadvantaged individuals have atypical observable characteristics which are correlated with high income levels, then they will fail to benefit from the income redistribution, and may even become considerably worse off. This makes little difference in a social welfare maximisation, which ensures that society in general is better off than it was before, but it sits uneasily with a desire toprevent hardship. It may thus be desirable to impose an explicit minimum income or utility constraint on the redistributive model in order to safeguard against unpalatable outcomes (although one could also argue that this aim should already be embodied in the form of the social welfare function). In the context of the model of Section (2), part (a), it is preferable to impose a minimum utility, rather than income, guarantee, of the form

$$U_h = U(\underline{p}, W_h, a_h, W_hT + \alpha + \beta a_h) \geqslant \bar{U} \qquad h=1,\ldots,H$$

where $\bar{U}$ is some exogenously determined 'acceptable' minimum utility level. If the inequality constraints are notbinding at the optimum, then there is no change from the cases above. But if, say, the constraints on the first J individuals, h=1,...,J, are binding, then the optimality conditions become

$$\frac{\partial L}{\partial \alpha} = \sum_{h=1}^{H} \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h} - H\mu + \sum_{h=1}^{J} \nu_h \frac{\partial U_h}{\partial M_h} = 0$$

$$\frac{\partial L}{\partial \beta} = \sum_{h=1}^{H} \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h} \cdot a_h - \mu \sum_{h=1}^{H} a_h + \sum_{h=1}^{J} \nu_h \frac{\partial U_h}{\partial M_h} a_h = 0$$

$$\frac{\partial L}{\partial \nu_h} = U_h - \bar{U} = 0 \qquad h=1,\ldots,J$$

where ν_h (>0) is the Lagrange multiplier attached to the inequality constraint on the h^{th} individual, h=1,...,J. From this it is apparent that μ is no longer equal to the mean MSU of income to individuals, and takes a lower value owing to the constrained nature of the optimum. Similarly, the zero covariance condition is also disrupted. The presence of a binding set of utility guarantees is therefore a barrier to achieving the maximum possible degree of redistribution, emphasising the differing nature of the redistributive and income support objectives. The importance of the constraints tends to be greater where there are highly atypical individuals (regardless of the number of them), although the effect should be less important when there are many observable characteristics. If a small number of eccentric cases does turn out to be hampering redistribution, it is to the mutual benefit of the government and the individuals concerned to remove the informational deficiency in their particular case, so that the constrained optimum may not be stable in the long run. Generally speaking, this kind of redistribution plus social security framework has theoretical shortcomings (not least the determination of the threshold $\bar{U}$), and cannot be proffered as an ideal means of analysing policy design.

Much discussion of social security has concentrated on the choice between an income-related and needs-related payments scheme (Meade (1978), Kay and King (1983), Chapter 8). A pure income-related scheme gives out payments based solely on an individual's income level; social dividend schemes and negative income taxes come into this category. Conversely, pure needs-related payments take account only of an individual's needs, as reflected in some observable characteristic, such as family size, sickness, employment status, etc.; this in the U.K. largely corresponds to the original Beveridge approach. Both of these have been suggested as the ideal social security system, but in practice most benefits tend to be a hybrid of the two. Hence there are income support payments which vary with an individual's observed needs, and needs-related payments which are means-tested and thereby depend partly on income. This compromise is strongly supported by the theoretical models above, where the choice between characteristics as a basis for payments is something of a non-issue. In the absence of administrative or practical difficulties it is always preferable to maximise the information used to determine payments, by considering as many relevant observable characteristics as are available. If income happens to be one of these characteristics, then the result is a scheme which is partly income-based and partly needs-based. The bias towards one or the other approach depends on the structure of characteristics in the population and their effects on utility, but it is almost certain that any socially optimal scheme involves elements of both. Of course in practice there are other complicating factors (administration costs, social attitudes to means testing, etc.), and these have to be allowed for in formulating policy.

(6) Conclusion

This chapter has looked at some simple models of redistribution, in which income payments are related to the observable characteristics of multidimensional individuals. The approach is slightly different from many redistributive models, which are based on a unidimensional population differing in a characteristic not directly observable. Such models (originating in the optimal income taxation framework of Mirrlees (1971)), lead to a constrained policy optimum in which income transfers provoke disincentive responses by individuals. In practice it is clear that individuals are (emphatically) multidimensional, possessing a combination of observable and unobservable characteristics. Given that some characteristics are observable and that income transfers are the most efficient redistributive tool, it is natural for a government to ask whether income transfers can be related to the observable characteristics, leading to the type of model considered above. Provided that assignment uncertainty applies, the optimum is derivable in a perfectly straightforward fashion, with no difficulties arising from disincentives; the only fundamental constraint on policy is the inability to observe a subset of characteristics. It remains the case that further policy tools (such as indirect taxes, discriminatory taxes, or rations) could improve on the policy optimum, although income transfers are fairly clearly the first choice of instrument. In situations where the characteristic is subject to known systematic variation by individuals, the optimisation can still go ahead, yielding a modified optimality condition. Within the theoretical framework used, there is consequently no overriding obstacle to achieving a significant degree of redistribution.

Such simplicity does not carry over to the implementation of actual redistributive policies. There are two principal difficulties involved:

(a) The enumeration of personal characteristics on the lines assumed in the models is not feasible in reality, and the same can also be said of the accurate specification of both individual tastes and the distribution of characteristics in the population. Any actual redistributive measures are perforce an approximate exercise, perhaps better described as policy reform than optimisation. While the prospects for redistribution are not nullified by these difficulties, it has to be conceded that a high degree of exactitude attached to policy proposals would be spurious.

(b) In practice the presence of social institutions has an important influence on the chance of redistribution, and this element is neglected by the theoretical models above. The modest degree of redistribution accomplished by existing social security systems seems to be a reflection of this.

Consequently these models cannot really be advanced as a precise policy prescription or as a representation of conventional social security measures. Their main value is that they give a clear picture of the basic informational constraint facing redistributive policies, and therein represent an upper bound to a government's redistributive ambitions. In terms of policy practice, one ventures to say only that the models raise some doubts about current redistributive policies, considering that income transfers are physically feasible policy tools and that the present income distribution is an improbable candidate for a social optimum.

CHAPTER 3 : QUANTITY REDISTRIBUTION

(1) Introduction

In economics utilitarian discussion has a materialistic flavour, centring on the physical consumption of goods and services. Individuals are assumed to have utility functions defined in terms of quantities linked to market transactions; all other facets of utility are omitted from the analysis. Although 'utility' is a nebulous concept, one can safely say that it has more ingredients than the consumption of goods. Material prosperity is only one element, albeit an important one, in the full set of influences on an individual's welfare. Economics does not deny this, of course, and can defend its approach as a choice of emphasis rather than a restrictive assumption. Non-material influences on utility can be either embodied in the functional form of utility (if they interact with material influences) or regarded as an additively separable component of utility (if there is no such interaction). There is consequently no need to mention them explicitly whenever a utilitarian formulation is being used. Many non-material influences on utility are based on fixed personal characteristics, which are not feasible targets for the government's social welfare maximisation policies. In other cases, however, variation is possible, implying that utilitarian policy may go further than the production and distribution of goods; an example is manipulation of legal and political rights, where these are considered flexible. Policies based on such parameters are not a standard concern of economics, and are absent from most theoretical discussion. Economics generally lives up to its role as the material social science by restricting its attention to the consumption of observable goods and services.

This places the burden squarely on quantities as the foundation of individual and social welfare. A large proportion of economic theorising starts with the assumption that individual utility has the form $U = U(X_1,\ldots,X_N)$, where $X_1,\ldots,X_N$ are the quantities of goods 1 to N consumed in a given period. All subsequent conclusions depend on what happens to these X variables, even if this is not always apparent from the terms of the discussion. Anything that is said about prices and incomes must boil down to quantity effects when the analysis is framed in these terms. The highlighting of quantities means that the modelling is 'real' as opposed to 'monetary', and that all welfare results have to be based on the consumption of physical commodities. This is probably entirely justified, but it does mean that the monetary variables of prices and incomes are secondary where welfare issues are concerned. One is therefore prompted to ask why relatively little is said about quantities when redistribution is being discussed. The neglect is such that redistribution is often equated with income redistribution, and other redistributive policies are seen as contributing to that end (for example, wealth or commodity taxes). This view is misplaced whenever utility is derived from consumption of goods; income redistribution is a means to the end of redistributing consumption, not an end in itself. There may well be good reasons for concentrating on incomes and prices, but they do not emerge from the body of the theory itself and they are not so obvious as to be self-evident. It is consequently of interest to consider quantity redistribution.

This chapter looks at some issues relevant to the direct allocation of quantities. Section (2) is a generalisation of rationing theory to allow for any type of quantity constraint; the outcome is used in Section (3) to consider the desirability of quantity allocation.

(2) Generalised Rationing

The polar cases in allocating goods are a complete reliance on prices, income and consumer choice (as is usual in practice), and a direct centralised distribution of commodities. Between them there is a wide range of mixed situations, involving a combination of allocation by incomes and quantities. In consumer theory these appear as rationing problems, where the consumer faces both income and quantity constraints. This section generalises the conventional rationing model to allow for a variety of restrictions on quantities.

Rationing theory originated in Rothbarth (1940), with the notion of 'virtual' prices at which the ration is voluntarily chosen by consumers. This was elaborated by Tobin and Houthakker (1950), who compared the demand elasticities of rationed and unrationed demand functions. In recent years the theory has been refined by the application of duality methods (Neary and Roberts (1980), Deaton (1981b)), and the resulting rationed demand functions have a general applicability equivalent to that of ordinary demand functions. The keystone of the modern theory is the definition of virtual prices as a function of other prices, utility and the ration level. If good 1 is rationed to a level $\bar{X}_1$, then the virtual price $\tilde{p}_1$ is defined such that

$$\left.\frac{\partial C}{\partial p_1}\right|_{\tilde{p}_1} = \bar{X}_1 \Rightarrow \tilde{p}_1 = \tilde{p}_1(\underline{p}\ ,\ \bar{X}_1\ ,\ U)$$

Hence the derivative of the consumer's cost function with respect to p_1 is equated with X_1, yielding an equation for the virtual price $\tilde{p}_1$ in terms of the remaining commodity prices $\underline{p}$, the ration level $\bar{X}_1$ and utility, U. When the function $\tilde{p}_1$ is substituted for p_1 in conventional demand and cost functions, the outcome is a set of rationed functions which can be used in a similar way to standard consumer theory. Details of the relation between rationed and unrationed demand functions are given in Neary and Roberts (1980).

Besides restrictions on a single good, it is also possible to have quantity constraints applying to several goods (for example, in the allocation of time). In discussing quantity constraints in general, it is consequently desirable to extend rationing theory to cover any type of quantity constraints. The discussion below firstly looks at additive constraints, and then at the general case.

(i) Additive Constraints

Suppose that there are N goods, with associated prices $p_1,\ldots,p_N$. A consumer has preferences defined over these goods, with utility $U = U(X_1,\ldots,X_N)$, where X_i is consumption of the i^{th} good. The first M goods are subject to a binding additive constraint

$$\sum_{i=1}^{M} \beta_i X_i = T$$

where T is some given constant. Otherwise any bundle of goods can be chosen within the budget constraint set by the income level.

To treat this as a rationing problem, it is necessary to obtain the relevant virtual prices for goods 1 to M. The first requirement is that they must satisfy the quantity constraint, so that

$$\sum_{i=1}^{M} \beta_i \left.\frac{\partial C}{\partial p_i}\right|_{\tilde{p}} = T$$

Hence, at the virtual prices, the weighted sum of the price derivatives of the cost function must equal T. On its own this relation is not enough to define a unique set of virtual prices for the M goods inside the constraint. There is a range of potential virtual prices, reflecting the fact that the consumer can choose how to distribute the ration between the M goods. The appropriate virtual prices are those which minimise the cost of attaining a given utility level, subject to the relation above. The consumer's choice of $X_1,\ldots,X_M$ within the ration can be represented by a choice of the corresponding virtual prices $\tilde{p}_1,\ldots,\tilde{p}_M$. Let $\underline{q}$ denote the prices of the N-M goods outside the ration and $X_j^c(\underline{p},\underline{q},U)$ the compensated demand function for good j. The relevant cost objective is

$$\tilde{C} = C(\tilde{\underline{p}},\underline{q},U) + \sum_{j=1}^{M}(p_j-\tilde{p}_j)\, X_j^c(\tilde{\underline{p}},\underline{q},U)$$

that is, the consumer's cost at virtual prices $\tilde{\underline{p}}$ plus an adjustment to allow for the difference between $\tilde{\underline{p}}$ and $\underline{p}$. The problem and solution are as below:

Choose $\tilde{p}_1,\ldots,\tilde{p}_M$ to minimise $C(\tilde{\underline{p}},\underline{q},U) + \sum_{j=1}^{M}(p_j-\tilde{p}_j)\, X_j^c(\tilde{\underline{p}},\underline{q},U)$

subject to $\sum_{j=1}^{M}\beta_j\, X_j^c(\tilde{\underline{p}},\underline{q},U) = T$

$$L = C(\tilde{\underline{p}},\underline{q},U) + \sum_{j=1}^{M}(p_j-\tilde{p}_j)\, X_j^c + \phi\left(T - \sum_{j=1}^{M}\beta_j\, X_j^c\right)$$

$$\frac{\partial L}{\partial \tilde{p}_i} = \frac{\partial C}{\partial \tilde{p}_i} - X_i^c + \sum_{j=1}^{M}(p_j-\tilde{p}_j)\frac{\partial X_j^c}{\partial \tilde{p}_i} + \phi\sum_{j=1}^{M}\beta_j\frac{\partial X_j^c}{\partial \tilde{p}_i} = 0 \qquad i=1,\ldots,M$$

From the properties of cost functions it follows that

$$\sum_{j=1}^{M}(p_j-\tilde{p}_j)\, S_{ji} + \phi\sum_{j=1}^{M}\beta_j\, S_{ji} = \sum_{j=1}^{M}(p_j-\tilde{p}_j+\phi\beta_j)\, S_{ij} = 0 \quad i=1,\ldots,M$$

where S_{ij} is the i,j^{th} term of the Slutsky matrix. The equations above are satisfied if

$$\tilde{p}_i = p_i + \phi\,\beta_i \qquad i=1,\ldots,M$$

such that the virtual price of the i^{th} rationed good is the sum of its actual price and the 'price' incurred by its presence in the ration. Eliminating ϕ yields the M-1 conditions

$$\frac{p_i-\tilde{p}_i}{p_M-\tilde{p}_M} = \frac{\beta_i}{\beta_M} \qquad i=1,\ldots,M-1$$

treating good M as the common basis of comparison. At the optimum the relative impact on cost of adjusting any two rationed goods is equated with their marginal rate of substitution within the ration.

Combining the quantity constraint with the M-1 independent conditions above is enough to define a unique set of virtual prices for the rationing problem. These can be used in a manner closely resembling that in the standard case. In practice it may not be possible to derive explicit expressions for virtual prices, but the properties of the resulting implicitly defined functions remain the same. The rationed cost function is defined as

$$\tilde{C}(\underline{p}, \underline{q}, U, T) = \min \sum_{i=1}^{M} p_i X_i + \sum_{j=M+1}^{N} q_j X_j \quad \text{s.t.}$$

$$\sum_{i=1}^{M} \beta_i X_i = T, \; U = \bar{U}$$

If X^c denotes a compensated demand, one can then write

$$\tilde{C} = \sum_{i=1}^{M} p_i \tilde{X}_i^c (\underline{p}, \underline{q}, U) + \sum_{j=M+1}^{N} q_j \tilde{X}_j^c (\underline{p}, \underline{q}, U)$$

$$= \sum_{i=1}^{M} p_i X_i^c (\tilde{\underline{p}}, \underline{q}, U) + \sum_{j=M+1}^{N} q_j X_j^c (\tilde{\underline{p}}, \underline{q}, U)$$

$$= C(\tilde{\underline{p}}, \underline{q}, U) + \sum_{i=1}^{M} (p_i - \tilde{p}_i) X_i^c (\tilde{\underline{p}}, \underline{q}, U)$$

as is also the case with the standard theory. The cost function has the usual properties, and in particular the relation

$$\frac{\partial \tilde{C}}{\partial p_i} = X_i \qquad i=1,\ldots,M$$

still holds true. The only real difference from the usual case is the effect of a change in the ration level, T. With a single good ration this has a known effect on the consumption of the good concerned, whereas in the present case a change in T can produce a reallocation of the ration. The effect of T on the virtual prices therefore has to be taken account of. Differentiating $\tilde{C}$ with respect to T yields

$$\frac{\partial \tilde{C}}{\partial T} = \frac{\partial}{\partial T} \left[C(\tilde{\underline{p}}, \underline{q}, U) + \sum_{i=1}^{M} (p_i - \tilde{p}_i) X_i^C(\tilde{\underline{p}}, \underline{q}, U) \right]$$

$$= \sum_{i=1}^{M} \left(\frac{\partial C}{\partial \tilde{p}_i} - X_i^C \right) \frac{\partial \tilde{p}_i}{\partial T} + \sum_{i=1}^{M} \left[(p_i - \tilde{p}_i) \sum_{j=1}^{M} \frac{\partial X_i^C}{\partial \tilde{p}_j} \cdot \frac{\partial \tilde{p}_j}{\partial T} \right]$$

$$= \sum_{i=1}^{M} (p_i - \tilde{p}_i) \left(\sum_{j=1}^{M} S_{ij} \frac{\partial \tilde{p}_j}{\partial T} \right) = (\underline{p} - \tilde{\underline{p}}) \, \underline{S} \frac{\partial \tilde{\underline{p}}}{\partial T} = \phi$$

where $\underline{S}$ refers to the section of the Slutsky matrix relevant to the first M goods (evaluated at virtual prices). The equivalent expression for a single good ration is simply $p - \tilde{p}$, since there are no substitution possibilities. Remaining properties of the rationed functions vary from the standard case to the extent that this relation is relevant. They can be summarised as follows:

<u>Change in T</u>

$$\tilde{X}_i^C(\underline{p}, \underline{q}, U, T) = \tilde{X}_i(\underline{p}, \underline{q}, C, T) \qquad \forall i$$

$$\frac{\partial \tilde{X}_i^C}{\partial T} = \frac{\partial \tilde{X}_i}{\partial T} + \frac{\partial \tilde{X}_i}{\partial M} \cdot \frac{\partial \tilde{C}}{\partial T} = \frac{\partial \tilde{X}_i}{\partial T} + \frac{\partial \tilde{X}_i}{\partial M} \left[(\underline{p} - \tilde{\underline{p}}) \, \underline{S} \frac{\partial \tilde{\underline{p}}}{\partial T} \right]$$

$$\Rightarrow \frac{\partial \tilde{X}_i}{\partial T} = \frac{\partial \tilde{X}_i^C}{\partial T} - \frac{\partial \tilde{X}_i}{\partial M} \left[(\underline{p} - \tilde{\underline{p}}) \, \underline{S} \frac{\partial \tilde{\underline{p}}}{\partial T} \right]$$

<u>Change in Income, M</u>

$$\tilde{X}_i(\underline{p}, \underline{q}, M, T) = \tilde{X}_i(\underline{p}, \underline{q}, M, \sum_{j=1}^{M} \beta_j X_j(\tilde{\underline{p}}, \underline{q}, M)) = X_i(\tilde{\underline{p}}, \underline{q}, M)$$

$$\frac{\partial \tilde{X}_i}{\partial M} + \frac{\partial \tilde{X}_i}{\partial T} \sum_{j=1}^{M} \beta_j \frac{\partial X_j}{\partial M} = \frac{\partial X_i}{\partial M}$$

$$\Rightarrow \frac{\partial \tilde{X}_i}{\partial M} = \frac{\partial X_i}{\partial M} - \frac{\partial \tilde{X}_i}{\partial T} \left(\sum_{j=1}^{M} \beta_j \frac{\partial X_j}{\partial M} \right) \qquad \forall i$$

where $\partial \tilde{X}_i / \partial T$ is as above, and ordinary demand functions, X_i, are evaluated at virtual prices.

Changes in Prices, q,p

$$\tilde{X}_i^C(\underline{p}, \underline{q}, U, T) = \tilde{X}_i(\underline{p}, \underline{q}, M, T) = \tilde{X}_i(\underline{p}, \underline{q}, \sum_{j=1}^{N} p_i\tilde{X}_i, T)$$

$$\Rightarrow \frac{\partial \tilde{X}_i^C}{\partial q} = \frac{\partial \tilde{X}_i}{\partial q} + \frac{\partial \tilde{X}_i}{\partial M} \cdot \tilde{X}_i$$

$$\Rightarrow \frac{\partial \tilde{X}_i}{\partial q} = \frac{\partial \tilde{X}_i^C}{\partial q} - \frac{\partial \tilde{X}_i}{\partial M} \cdot \tilde{X}_i \qquad \forall\ i$$

Hence there exists a standard Slutsky equation for the rationed demand functions, with $\partial\tilde{X}_i/_{\partial M}$ as defined above. Changes in $\underline{p}$ are analogous, with $\underline{p}$ in place of $\underline{q}$. For a general ration there exists a direct substitution effect for $\underline{p}$, arising from the allocation of goods within the ration.

As with single good rationing, there is a complete set of rationed/unrationed, compensated/uncompensated demand functions, with associated derivatives. The effect of the ration on the various derivatives is to introduce a 'quantity' effect (operating through changes in T), to set alongside the income and price (substitution) effects. A change in T has both a direct impact on demand and an income effect, while a change in M has a direct impact and a 'quantity' effect. Changes in the prices of unrationed goods have a rather complicated mixture of all three effects, and for rationed goods the position is similar, with the presence of direct substitution effects. The intuitive rationale for the 'quantity' effect is fairly clear. A movement in prices or income serves to 'tighten' or 'slacken' the ration, according to whether the imposed constraint becomes closer or further from what the individual would choose voluntarily. The perceived effect on the ration is therefore relevant in considering the net impact of price and income changes.

Before discussing applications of this theory, it is worth extending it to the general case.

(ii) General Constraints

When several goods are involved quantity constraints need not always be linear, and could well take more complicated non-linear forms. To allow for this, suppose that a constraint is again imposed on the first M goods, but that it now takes the general form $g(X_1,\ldots,X_M) = 0$. Discussion can follow the same procedure as in (i), firstly defining a set of virtual prices, and then using them to obtain rationed cost and demand functions.

The appropriate virtual prices must induce the consumer to satisfy the quantity constraint voluntarily, implying the condition

$$g\left(\frac{\partial C}{\partial P_1}, \frac{\partial C}{\partial P_2}, \ldots, \frac{\partial C}{\partial P_M}\right)\Bigg|_{\underline{\tilde{p}}} = 0$$

As in (i), this is not enough on its own to identify a unique set of virtual prices. Allocation of goods within the ration can be represented by a cost minimising choice of virtual prices, such that

$$L = C(\underline{\tilde{p}},\underline{q},U) + \sum_{j=1}^{M}(p_j-\tilde{p}_j)X_j^C + \phi(g(X_1^C,\ldots,X_M^C))$$

$$\frac{\partial L}{\partial \tilde{p}_i} = \frac{\partial C}{\partial \tilde{p}_i} - X_i^C + \sum_{j=1}^{M}(p_j-\tilde{p}_j)\frac{\partial X_j^C}{\partial \tilde{p}_i} + \phi\sum_{j=1}^{M}\frac{\partial g}{\partial X_j}\frac{\partial X_j}{\partial \tilde{p}_i} = 0 \qquad i = 1,\ldots,M$$

From the properties of cost functions one obtains

$$\sum_{j=1}^{M}(p_j-\tilde{p}_j)\,S_{ji} + \phi\sum_{j=1}^{M}\frac{\partial g}{\partial X_j}S_{ji} = \sum_{j=1}^{M}\left(p_j-\tilde{p}_j+\phi\frac{\partial g}{\partial X_j}\right)S_{ij} = 0 \qquad i = 1,\ldots,M$$

These equations are satisfied if

$$\tilde{p}_i = p_i + \phi\frac{\partial g}{\partial X_i} \qquad i = 1,\ldots,M$$

so that $\tilde{p}_i$ is a sum of an actual price, p_i, and a shadow price, $\phi\ {}^{\partial g}/_{\partial x_i}$. The sign of ϕ depends on the value of the function g in the unconstrained consumption decision : let this be denoted by g*. If g* > 0, then the ration reduces the value of g, and $\phi > 0$ must hold, so that a higher permitted value of g reduces the consumer's cost. The converse also applies, and hence $\phi \gtreqless 0$ as $g^* \gtreqless 0$. From the expression for $\tilde{p}_i$ it follows that if ${}^{\partial g}/_{\partial x_i} > 0$, then $\tilde{p}_i \gtreqless p_i$ as $\phi \gtreqless 0$. The virtual price of x_i exceeds its actual price when the level of g is reduced ($\phi > 0$), and vice versa. If ${}^{\partial g}/_{\partial x_i} > 0$, then the relation is reversed, so that $\tilde{p}_i \gtreqless p_i$ as $\phi \lesseqgtr 0$. Eliminating ϕ yields

$$\frac{p_i - \tilde{p}_i}{p_M - \tilde{p}_M} = \frac{{}^{\partial g}/_{\partial x_i}}{{}^{\partial g}/_{\partial x_M}} \qquad i = 1,\ldots,M=1$$

with good M as the arbitrary basis of comparison. Thus the relative cost of substituting any two rationed goods is equated with their marginal rate of substitution within the function g. Otherwise the same utility can be achieved at lower cost by reallocating the ration.

Rationed cost and demand functions are derived by substitution of virtual prices (as functions of other prices, utility and the parameters of the quantity constraint) into the ordinary cost and demand functions. It is not possible to analyse demand derivatives in the same way as in (i), since there is no longer an analogy for the 'quantity', T. The goods involved in the ration are not necessarily measurable by a common quantity unit, and the constraint may involve many different coefficients. Thus, although a set of demand derivatives exists, it will always depend on the particular coefficients of the constraint function g, and consequently a general form cannot be given.

In most cases it will not be possible to obtain explicit equations for virtual prices, but the same principles still apply to a set of implicitly defined virtual prices. One can always therefore interpret quantity constrained demand in terms of rationing theory.

There are numerous instances where demand may be influenced by quantity constraints, and where a general rationing theory may be appropriate. Some examples are:

(a) Allocation of Time

Time is the classic example of a scarce commodity subject to a well defined quantity constraint. Any activity that takes time must be included in a constraint on the total time available. This is a common feature of household production theory, where time may be an input into many different activities (Becker (1965)).

(b) Physical Constraints

In some situations demand for goods is subject to a physical constraint on the amount that can be consumed. An example is where storage is involved and spatial capacity is limited; in stocking, say, a refrigerator or freezer it is necessary to make allowance for the available space, which may impose a binding constraint on demand. Another case is where previous decisions restrain current consumption, as when the purchase of a given type and size of house dictates the choice of other consumption activities over subsequent years (choice of furnishings, range of activities compatible with size and location of house, etc.).

(c) Non-physical Constraints

The individual may also face non-physical constraints on consumption. One possibility (increasingly common) is limits imposed for health reasons, as when dietary rules restrict the consumption of food. If the constraint is self-imposed, it can potentially be integrated with an unconstrained utility function, although it remains the case that the consumer's 'spontaneous' choice has probably been influenced by external information and persuasion. A similar situation arises with advertising, where 'voluntary' consumption decisions can be seen as guided by the activities of producers (a major theme of Galbraith (1958, 1967)). In some cases there are legal restrictions on consumption, as when it generates excessive noise or pollution.

(d) Rations

Quantity constraints can be imposed deliberately as a tool of redistributive policy, appearing either as rations or some other form of non-monetary allocation. Specific restrictions on particular goods are depicted by standard rationing theory. Voucher schemes or other restrictions require the generalised theory given above.

(e) Consumer Psychology

Consideration of consumer psychology can easily lead to notions of constrained consumption. In household production theory, for example, utility depends on unobserved characteristics or qualities, which may be psychological in nature; consumption of characteristics is then constrained partly by the household production technology, in addition to the need to purchase goods (Lancaster (1971)). Alternatively, in trying to analyse what is meant by utility, one can envisage diminishing returns imposed (or strengthened) by constraints; an absolute 'psychological' valuation may be compromised by 'physical' restrictions, as occurs with over-consumption of food. Another example is the possibility of 'positional' goods (Hirsch (1977)), which can lead to quantity constaints if, say, consumers wish to match the (exogenous) consumption patterns of their neighbours (or otherwise to collective constraints on all consumers). These issues can often be subsumed in preferences, but in some situations it may be instructive to look at them more explicitly.

The above are merely a few examples of where quantity restrictions may be relevant to consumer demand. Generally speaking, there is little cause to believe that consumption decisions are universally explicable by a monetary constraint alone. The choice of many goods appears to be influenced by other restrictions, requiring a more complex demand theory subject to several different non-monetary constraints. In this sense quantity restricted theory is more general than the usual representations of consumer demand. Rationing is also applicable in other areas, such as the constrained decisions of firms or other agents.

The discussion above has set out an extended version of rationing theory, applicable to general quantity constraints. It differs from the standard version in the formulation of virtual prices, which need an additional set of conditions in their definition. For the general constraint $g(X_1,\ldots,X_M) = 0$, virtual prices $\tilde{p}_1,\ldots,\tilde{p}_M$ are defined by

$$\text{(a)} \qquad g(X_1^C(\tilde{\underline{p}}, \underline{q}, U), \ldots, X_M^C(\tilde{\underline{p}}, \underline{q}, U)) = 0$$

$$\text{(b)} \qquad \frac{p_i - \tilde{p}_i}{p_M - \tilde{p}_M} = \frac{\partial g / \partial X_i (\tilde{\underline{p}}, \underline{q}, U)}{\partial g / \partial X_m (\tilde{\underline{p}}, \underline{q}, U)} \qquad i = 1, \ldots, M-1$$

(a) ensures the satisfaction of the constraint, corresponding to the condition that uniquely defines the virtual price in single good rationing. The M-1 conditions in (b) arise from the allocation of goods within the ration, and can be characterised as the consumer choosing the optimal set of virtual prices. The theory of Neary and Roberts (1980) is then a special case of the above, where there is no scope for allocation within the constraint and the second set of conditions is not required.

A generalised rationing theory is useful in considering the redistributive role of quantity constraints, a question addressed in Section (3).

(3) Quantity Allocation

Welfare economics customarily emphasises the desirability of lump-sum income transfers, coupled with unrestricted consumer choice under efficient pricing. Quantity constraints are then viewed as distortionary, upsetting the marginal efficiency conditions that emerge from unconstrained behaviour. Under these circumstances, the value of quantity redistribution depends on special conditions outside the usual assumptions made in economic theory. An example is 'specific egalitarianism', where social preferences diverge from an individualistic form and depend also on the consumption of a particular commodity; it may then be justifiable for the government to intervene in redistributing the said commodity (Tobin (1970)). In practice there is little prospect of efficient pricing, so the impact of the usual conclusions is accordingly diminished. When pricing is inefficient, an optimum optimorum cannot be achieved by lump-sum income redistribution, and there is scope for the use of quantity redistribution. It is premature to dismiss intervention in quantity allocation on the basis of principles derived from idealised competitive equilibria. In suboptimal 'second-best' situatins there is a much stronger case for the use of quantity controls, which persists even when lump-sum income transfers are available. One can therefore say that the value of non-monetary allocation is underrated by conventional economic discussion. Far from needing to be justified by special conditions, the opposite is true; quantity allocation has nothing to contribute only in the special case of a fully efficient monetary allocation system.

As practical policy measures, rationing and quantity controls are traditionally associated with a definite shortage of one or more goods relative to others. In these conditions rationing by price may be viewed as inequitable (given an unequal income distribution), and it may be felt desirable to withdraw the goods from the usual monetary allocation methods. The main observed example of a distributional 'crisis' is in wartime, when the supply of consumer goods is curtailed by a redirection of resources towards military production. Any other situation which significantly reduced the supply of certain goods could be used as a basis for similar arguments. The main feature is that there is a perceived aggregate constraint on certain goods which is more severe than that applicable to others. This does not mean that the value of rations and other allocation tools is limited to extraordinary conditions, as was pointed out above. A more general case for non-monetary allocation schemes could certainly be made, although the outcome would not fit in very easily with the customary set of policy tools. The discussion here relates to those positions commonly recognised as requiring an alternative to price allocation.

Quantity allocation can be divided between 'direct' and 'indirect' schemes, in the sense of Chapter (2) above. A 'direct' scheme distinguishes individuals either fully or by some observed characteristic, whereas 'indirect' allocation cannot predict the impact on any particular individual. The former cases are more efficient redistributive tools, but may be ruled out by the absence of the required information. Both types of scheme are considered below.

(i) Direct Allocation

As with income redistribution, one can distinguish between cases where full information is available and those where only a limited subset of individual characteristics is observed.

(a) Full Information

The government is here assumed to have complete knowledge of individual preferences and characteristics, allowing it to redistribute quantities in a manner resembling a lump-sum income redistribution. Suppose that a single good, X_1, is in short supply, such that the aggregate availability is equal to $\bar{X}_1$. All remaining goods, i=2,...,N, are allocated in the usual way, with consumers allowed to purchase as much as they wish within their budget constraints. The government's problem is therefore to choose the quantities $\bar{X}_1^h$, h=1,...,H, to maximise social welfare subject to the condition

$$\sum_{h=1}^{H} \bar{X}_1^h = \bar{X}_1$$

where $\bar{X}_1^h$ is the amount of the first good allocated to individual h. Individual preferences can be depicted by the rationed indirect utility functions $\tilde{U}_h = \tilde{U}_h(\underline{p}, M_h, \bar{X}_1^h)$, where $\underline{p}$ is the vector of consumer prices for goods j=1,...,N and M_h is the income of the h^{th} individual. There is nothing to be gained from charging a price for the first good, so p_1 is assumed to be zero. The rationed good is therefore allocated by administrative fiat, entirely outside the monetary system. The government's policy problem and solution are as below:

$$L = V(\tilde{U}_1,\ldots,\tilde{U}_H) + \mu\,(\bar{X}_1 - \sum_{h=1}^{H} \bar{X}_1^h)$$

$$\frac{L}{\bar{X}_1^h} = \frac{\partial V}{\partial \tilde{U}_h} \cdot \frac{\partial \tilde{U}}{\partial \bar{X}_1^h} - \mu = 0$$

$$\Rightarrow \mu = \frac{\partial V}{\partial \tilde{U}_h} \cdot \frac{\partial \tilde{U}_h}{\partial \bar{X}_1^h} \qquad i=1,\ldots,H$$

At the optimum the MSU of raising $\bar{X}^h$ for each individual is equated with μ, which is the MSU of raising the aggregate quantity $\bar{X}_1$. There is no further scope for raising social welfare by reallocating the first good, and the economy has attained a 'first best' with respect to non-monetary allocation. The optimality condition can be rewritten using the relation

$$\tilde{p}_1^h = \frac{\partial \tilde{U}_h / \partial \bar{x}_1^h}{\partial \tilde{U}_h / \partial M_h}$$

where $\tilde{p}_1^h$ is the virtual price of the first good for individual h; this is a general property of rationed demand functions. Substituting into the conditions above one obtains

$$\mu = \lambda_h \, \tilde{p}_1^h \qquad i=1,\ldots,H$$

where $\lambda_h \equiv \frac{\partial V}{\partial U_h} \cdot \frac{\partial \tilde{U}_h}{\partial M_h}$ = MSU of income to individual h

Hence the optimum equates the virtual prices of the first good weighted by the relevant MSU of income. Individuals with a high MSU of income will have a lower virtual price of the first good, and vice versa. This means that X_1 is distributed so as to compensate partially for the inequalities of the income distribution. In general the tendency will be for individuals on low incomes (and thus usually with a high MSU of income) to receive a relatively high allocation of the good (through their having a low virtual price). At a first-best income distribution the λ_h values are by definition equal, so that the optimal X_1 allocation requires an equalisation of the virtual prices for all individuals.

In strict theoretical terms the rationale for this policy is not clear. A government with full knowledge of individual preferences could reach a first-best position by redistributing income and allowing X_1 to be allocated by price. Alternatively, the same outcome could be attained by an administered non-monetary distribution of all N goods. There is consequently no compelling reason for a mixture of monetary and non-monetary allocation; the case above does not even attain the first best, since a lump-sum income redistribution is assumed not to take place. The model is theoretically justifiable only if there is something blocking the implementation of income redistribution. This could simply be an institutional constraint, as for example when the need for redistribution is only perceived in 'crisis' periods related to a subset of goods. The obstacle could otherwise be built into the model as a disincentive effect, in the manner of optimal taxation models. Since full scale income redistribution does not occur in practice, the case for a direct quantity allocation is stronger in reality than it appears in this particular framework.

An example of the scope for quantity allocation is the distribution of employment. Individuals often have little immediate control over their working time, in which case they can be regarded as constrained in their hours of work. Efficient allocation of work would be on the lines of the model above, producing an outcome where the weighted virtual wages of all individuals are equated (presumably with some sub-division between types and locations of employment). Such a result is a long way from the current method of organising employment, with some individuals fully unemployed and others working a fixed number of hours. If working time were to be allocated in the same way as other goods, the resulting inequality of hours worked would depend largely on the inequality in non-employment incomes - income redistribution together with efficient pricing could produce a socially optimal result. Whether optimal or not, the outcome would be very unlikely to coincide with the situation found in practice, rigidly differentiating the unemployed from the fully employed. Instead there would be a continuous range of working hours, and no unemployment in the standard sense of the term. This fits in with a pure neoclassical labour market theory, which presents employment as an unconstrained equilibrium arising from the interaction of employers' and workers' behaviour. Observed employment distributions appear, in contrast, to be non-monetary in origin, implying a set of constraints placed on the individuals involved. They are not the optimal non-monetary allocations described above, and cannot be justified by the theory of this chapter. To explain why they arise one would have to delve deeper into the institutional side of the question.

(b) Partial Information

The government may have insufficient information to implement a fully optimal quantity redistribution of the type set out in (a). It will not usually have zero information, so a feasible quantity allocation is liable to fall somewhere between the case in (a) and the no policy situation. This can be treated in the same way as in Chapter 2 with respect to monetary redistribution.

Let the policy problem remain as before, with the first good in short supply subject to an aggregate constraint $\bar{X}_1$. The only difference is that individuals are now distinguished by an observable characteristic, a_h, as well as their incomes, M_h. Preferences are such that

$$\tilde{U}_h = U(\underline{p}, M_h, a_h, \bar{X}_1^h) \qquad h=1,\ldots,H$$

where the functional form of utility is assumed to be identical for all individuals. The government either cannot observe incomes, or for some other reason does not make use of the M_h values in allocating X_1. Information on the joint distribution of M_h and a_h is available, however, permitting a_h to be used as a basis for redistribution. If a linear functional form is used, the allocation of X_1 is such that

$$\bar{X}_1^h = \alpha + \beta a_h$$

where α and β are the policy parameters to be chosen by the government. The constraint on the first good then becomes

$$\bar{X}_1 = \alpha H + \beta \sum_{h=1}^{H} a_h$$

The policy problem is to choose α, β to maximise social welfare subject to the restriction above, producing the following solution

$$L = V(\tilde{U}_1,\ldots,\tilde{U}_H) + \mu(\bar{X}_1 - \alpha H - \beta \sum_{h=1}^{H} a_h)$$

$$\frac{\partial L}{\partial \alpha} = \sum_{h=1}^{H} \frac{\partial V}{\partial \tilde{U}_h} \cdot \frac{\partial \tilde{U}_h}{\partial \bar{X}_1^h} - \mu H = 0$$

$$\frac{\partial L}{\partial \beta} = \sum_{h=1}^{H} \frac{\partial V}{\partial \tilde{U}_h} \cdot \frac{\partial \tilde{U}_h}{\partial \bar{X}_1^h} \cdot a_h - \mu \sum_{h=1}^{H} a_h = 0$$

Let ν_h denote the MSU of raising individual h's allocation of X_1, so that

$$\nu_h \equiv \frac{\partial V}{\partial \tilde{U}_h} \cdot \frac{\partial \tilde{U}_h}{\partial \bar{X}_1^h}$$

The first-order conditions can then be rewritten as

$$\mu = \sum_{h=1}^{H} \frac{\nu_h}{H} = \bar{\nu}$$

$$\sum_{h=1}^{H} \frac{\nu_h a_h}{H} - \bar{\nu} \sum_{h=1}^{H} \frac{a_h}{H} = 0 \qquad \Rightarrow \qquad \text{cov}\,(\nu_h, a_h) = 0$$

Hence the MSU of relaxing the aggregate constraint is equated with the mean MSU of increasing the allocations of X_1 received by individuals. When combined with the equation for β this yields a zero covariance condition, such that there is no systematic relationship between ν_h and a_h at the optimum. The values of α and β obtained will depend on the joint distribution of M_h and a_h, and their interrelationship within the functional form of preferences. If individuals started out with an initial suboptimal allocation of X_1, then their net gain or loss is equal to the optimal allocation minus the initial X_1 holding.

The optimality conditions are identical in form to those derived in Chapter 2 relating to income redistribution. One can repeat the question raised in (a) on the government's use of information; an ability to redistribute X_1 using knowledge of the distribution of M_h and a_h suggests that income could be redistributed in the same manner. Doing so would produce the zero covariance condition of Chapter 2, such that

$$\text{cov}\ (\lambda_h,\ a_h) = 0$$

where $\lambda_h = \dfrac{\partial V}{\partial U_h} \cdot \dfrac{\partial \tilde{U}_h}{\partial M_h}$ = MSU of income to individual h.

As in (a), λ_h and ν_h are related by the equation $\nu_h = \lambda_h\, \tilde{p}_1^h$, where $\tilde{p}_1^h$ is the virtual price of X_1 for individual h. This means that zero covariance between ν_h and a_h follows from the equivalent income condition if all virtual prices are equated, that is

$$\text{cov}\ (\nu_h,\ a_h) = \text{cov}\ (\lambda_h \tilde{p}_1,\ a_h) = \tilde{p}_1\ \text{cov}(\lambda_h,\ a_h) = 0$$

if $\tilde{p}_1^h = \tilde{p}_1$, h =1,...,H, and cov $(\lambda_h,\ a_h) = 0$. It follows that

$$\mu = \sum_{h=1}^{H} \frac{\nu_h}{H} = \sum_{h=1}^{H} \frac{\lambda_h \tilde{p}_1}{H} = \tilde{p}_1 \bar{\lambda} = \tilde{p}_1\ \phi$$

where ϕ = MSU of government revenue,

so that μ is equated with the MSU of government revenue multiplied by the single virtual price of X_1. Hence, if the available information is also used to redistribute income, the optimality condition for the allocation of X_1 reduces to an equating of the virtual prices, $\tilde{p}_1^h = \tilde{p}_1$, h=1,...,H. This corresponds to the situation found under the same circumstances in (a).

Relaxing the linearity of the policy measures does not change the basic nature of the outcome. If one allows for a continuous joint distribution

$f(M, a)$ and a non-linear ex post allocation $\overline{X}_1 = \overline{X}_1(a)$, then the resulting optimality condition is

$$\int_M \frac{\partial V}{\partial U} \cdot \frac{\partial \tilde{U}}{\partial X_1} f(M, a)\, dM = \int_M \nu\, f(M, a)\, dM = \bar{\nu}\Big|_a = \mu$$

so that the mean MSU of raising the X_1 allocation is equated for all different values of a. When income is redistributed on the same informational basis, the optimality condition again reduces to an equalisation of virtual prices between all individuals. The model is not dependent on the particular two-dimensional structure used here, and could be extended to a larger set of observable and unobservable characteristics.

In conclusion, the results of direct quantity allocation can be divided between 'pure' and 'impure' cases, according to whether or not income is redistributed simultaneously using the same available information. 'Pure' cases are those with the best attainable income distribution, ensuring that there is no element of income readjustment involved in quantity allocation. The optimal allocation then requires an equating of the virtual prices of all individuals. Such an outcome is intuitively appealing, since virtual prices indicate people's subjective valuations of the good, which should be equated at an efficient allocation. 'Impure' cases occur when incomes are distributed in a fixed arbitrary manner, and policy is confined to manipulation of quantities. In these situations the optimal quantity allocation is influenced by the desire for income redistribution, so that a full equalisation of virtual prices is not implemented. The optimality conditions are instead defined in terms of virtual prices weighted by the individual's MSU of income. Theoretically speaking, there is no justification for the 'impure' cases, and a rational government would make full use of its available information. In practice such a global approach may not be forthcoming, in which event it is desirable to adjust for the income distribution when directly allocating quantities.

(ii) Indirect Allocation

In some cases it is not feasible to redistribute quantities by directly identifying the recipients or by observing some of their characteristics. This occurs, for example, if the necessary information is not available or if there is some administrative obstacle to direct allocation. Under these conditions allocation has to proceed by 'indirect' means which are not dependent in any way on the personal characteristics of the population. Three main approaches can be identified, based on prices, quantities or some form of generalised rationing constraint.

(a) Allocation by Price

The first possibility is to set the prices of the rationed goods so as to equate aggregate demand and supply. Such an outcome might emerge spontaneously from market forces, or could alternatively be imposed as a policy measure by deliberately setting the appropriate prices. The amount of each good received by individuals will depend chiefly on the prevailing income distribution.

Suppose that the first L goods are in short supply, such that their aggregate availability is given by $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_L$. The remaining N-L goods are allocated in the usual way, and are free to be purchased at a given set of prices, denoted by $\underline{q}$. Individuals receive fixed incomes M_h, h=1,...,H. Allocation of the first L goods by price requires setting their prices $p_i^\dagger$, i=1,...,L, at the levels which produce aggregate demands $\bar{X}_1,\ldots,\bar{X}_L$. The price vector $\underline{p}^\dagger$ is defined such that

$$\sum_{h=1}^{H} X_i^h(p_1^\dagger,\ldots,p_L^\dagger,\ \underline{q},\ M_h) = \bar{X}_i \qquad i=1,\ldots,L.$$

where X_i^h is the ordinary demand function of individual h for good i. The number of aggregate equations is equal to the number of $p_i^\dagger$ values to be set, so in general there will be a particular solution for $\underline{p}^\dagger$. The scope for any choice over $\underline{p}^\dagger$ is correspondingly limited.

This outcome will not usually coincide with the social optimum, as defined in (i) (a) above. An optimal allocation requires that $\lambda_h \tilde{\underline{p}}^h$ be equated for all individuals, where λ_h is the MSU of income to h and $\tilde{\underline{p}}^h$ is the vector of h's virtual prices. Those with a high λ_h value should ideally face a relatively low virtual price, and vice versa. This does not occur in the case considered here, where all individuals are paying the same price and individuals on low incomes will generally receive less of all goods. Allocation by price can only reach a social optimum when there is a first-best income distribution, such that the λ_h values are equated for all individuals. In that case there is no need to set different virtual prices for all individuals, and a uniform price is desirable in social welfare terms. Although price allocation is always Pareto efficient (equating the marginal rates of substitution for all individuals), it does not normally provide the best means of allocating goods in short supply.

(b) Quantity Rationing

An alternative approach is for the government to specify the quantities of each scarce good that individuals receive. Unlike direct allocation, the government is not able to base the distribution of goods on individual characteristics. This usually means that the first L goods are shared out evenly, giving an identical allocation to each individual. There is nothing to be gained from charging a price for these allocations, so it is assumed that the rationed goods are simply given out in fixed amounts, while the remaining goods are allocated in the usual monetary manner. The resulting allocation is such that

$$\bar{X}_i^h = \frac{\bar{X}_i}{H} \qquad \begin{array}{l} i = 1,\ldots,M \\ h = 1,\ldots,H \end{array}$$

where $\bar{X}_i^h$ is the fixed amount of the i^{th} good received by h, and H is the total number of individuals. For each individual there is a set of virtual prices, $\tilde{p}_i^h$, i=1,...,M, at which the quantity allocation would be demanded voluntarily. These satisfy the relation

$$\bar{X}_i^h = X_i^h\,(\tilde{p}_1^h,\ \tilde{p}_2^h,\ldots,\ \tilde{p}_M^h,\ \underline{q},\ M)$$

where X_i^h is the ordinary demand function of h for i; the absence of any prices on the rationed goods means that there is no income cost incurred, and net income remains equal to M_h. In general the level of virtual prices will decline with income, so that those on low incomes will have relatively low $\tilde{p}$ values. As in (a), this can be compared with the optimal direct allocation, which requires an equating of $\lambda_h \tilde{p}_i^h$ for all h. There is no cause to believe that the outcomes will coincide, but since λ_h is usually inversely related to M_h, the general pattern of virtual prices will be similar. To distribute scarce goods evenly therefore has a redistributive effect which may be seen as desirable in social welfare terms. The outcome will be superior to the price allocation of (a) as long as low income individuals have a higher MSU of consuming the good in question than high income individuals (as is quite likely). While an even share out is not in general Pareto efficient, it can easily produce a result socially preferred to allocation by price.

(c) General Constraints

Apart from uniform pricing or quantity measures, indirect allocation can take a number of other forms. The only absolute requirement is that, by definition, the method of allocation exhausts the available supply of the goods in question; otherwise there is freedom to choose any particular form of constraint to be imposed on individuals. The resulting constrained demands can be analysed in terms of the 'generalised' rationing theory set out in Section (2). On the whole a more flexible approach to allocation will produce a better outcome than the price or quantity allocation considered in (a) and (b) above.

A constraint imposed on the rationed goods could take any functional form, and the particular approach adopted could be made subject to government choice. In practice, however, a feasible scheme is likely to have a fairly simple structure. The prime example is a uniform linear constraint imposed on goods which are in short supply. This corresponds to a voucher scheme, whre all individuals are allocated a given amount of purchasing power which can be distributed freely among the rationed goods at fixed rates of exchange. Most of the discussion below assumes linearity of the constraint, although other approaches are also possible.

Suppose that the economy is again facing a shortage of the first M goods, with aggregate availabilities $\bar{X}_1, \ldots, \bar{X}_M$. The government wishes to impose a linear constraint on the consumption of these goods, such that the total demand for them is equated with their availability. Each individual therefore faces a constraint of the form

$$\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_M X_M = \alpha$$

where α, $\beta_1,\ldots,\beta_M$ are constants. It is assumed that no monetary price is charged for the first M goods, so they are allocated entirely on the basis of this constraint. All individuals receive a uniform allocation of 'vouchers', α, which can be 'spent' on the first M goods at 'prices', β_i, i=1,...,M. Effectively there are two separate allocation schemes operating in the same manner, one for the first M goods and one for the remaining N-M. Each can be viewed as a constraint on the other, and the only real distinction between them is that there is a uniform allocation of vouchers, α, whereas income is unevenly distributed with values M_h, h=1,..., H. Demand for the first M goods can be depicted in terms of generalised rationing theory, with virtual prices $\tilde{p}_1,\ldots,\tilde{p}_M$ defined by

$$\sum_{i=1}^{M} \beta_i X_i^{ch} (\underline{\tilde{p}}^h, \underline{q}, U_h) = \alpha$$

$$\frac{p_i-\tilde{p}_i^h}{p_M-\tilde{p}_M^h} = \frac{\beta_i}{\beta_M} \qquad \begin{array}{l} i = 1,\ldots,M-1 \\ h = 1,\ldots,H \end{array}$$

For a given set of β_i values, the resulting virtual prices will usually be different for all individuals.

The government wants to set the parameters α and $\beta_1,\ldots,\beta_M$ so that total demand for the first M goods is equal to their availability. This implies that

$$\sum_{h=1}^{H} X_i^h (\underline{\tilde{p}}^h(\underline{q}, \beta_1^\dagger,\ldots, \beta_M^\dagger, \alpha^\dagger, M_h), \underline{q}, M) = \bar{X}_i \qquad i=1,\ldots,M$$

where $\underline{\tilde{p}}^h$ is the vector of virtual prices of the first M goods for individual h. It can be assumed that there is no 'money illusion' with respect to voucher prices, allowing $\alpha^\dagger$ to be treated as an arbitrarily defined numeraire. The remaining parameters $\beta_1^\dagger,\ldots,\beta_M^\dagger$ must satisfy the M equations above; as with prices in (a) the number of equations and parameters is equal, so there is not in general any scope for choosing the $\beta^\dagger$ values. Once again the resulting virtual prices will not correspond to the optimal values in Section (1), part (a) . Since $\alpha^\dagger$ is equal for all individuals, however, those on a low income are able to obtain a relatively large share of total consumption of the first M goods. This should reduce the average level of their virtual prices, and thus be distributively benevolent, provided that low income recipients have a high MSU of income. As with quantity rationing, the outcome must be preferable to price allocation if the less well off have a higher MSU of consuming the rationed goods. A voucher system can also be expected to be superior to the quantity distribution in (b). In particular it leads to a levelling of purchasing power over the first M goods, coupled with efficient allocation. The consumer's choice within the linear constraint can be represented directly, in a manner analogous to ordinary consumer theory. The optimal $X_1^h,\ldots,X_M^h$ values are chosen to maximise

$$\tilde{U}_h(\underline{q},\ M_h,\ X_1^h,\ldots,X_M^h) \quad \text{s.t.} \quad \sum_{i=1}^{M} \beta_i X_i^h = \alpha$$

where $\tilde{U}_h$ is the rationed indirect utility function corresponding to quantity rations on the first M goods. Choice of $X_1^h,\ldots,X_M^h$ within the linear constraint yields the usual form of optimality condition, such that

$$\frac{\partial \tilde{U}_h}{\partial X_i^h} \Bigg/ \frac{\partial \tilde{U}_h}{\partial X_j^h} = \frac{dX_j^h}{dX_i^h} = \frac{\beta_i}{\beta_j} = \frac{\tilde{p}_i}{\tilde{p}_j} \qquad i,j = 1,\ldots,M$$

The marginal rate of substitution between any two goods within the constraint is equated with the 'voucher price' ratio, ensuring efficiency in allocation. Such efficiency is not attained in the quantity rationing of (b), and since the equal share out in (b) is a feasible position within the linear constraint used here, it follows that the voucher scheme is preferable to quantity rationing. A general rationing constraint can therefore combine the virtues of price and quantity allocation, allowing a more efficient allocation through the element of choice and a more equitable one through the evening out of purchasing power.

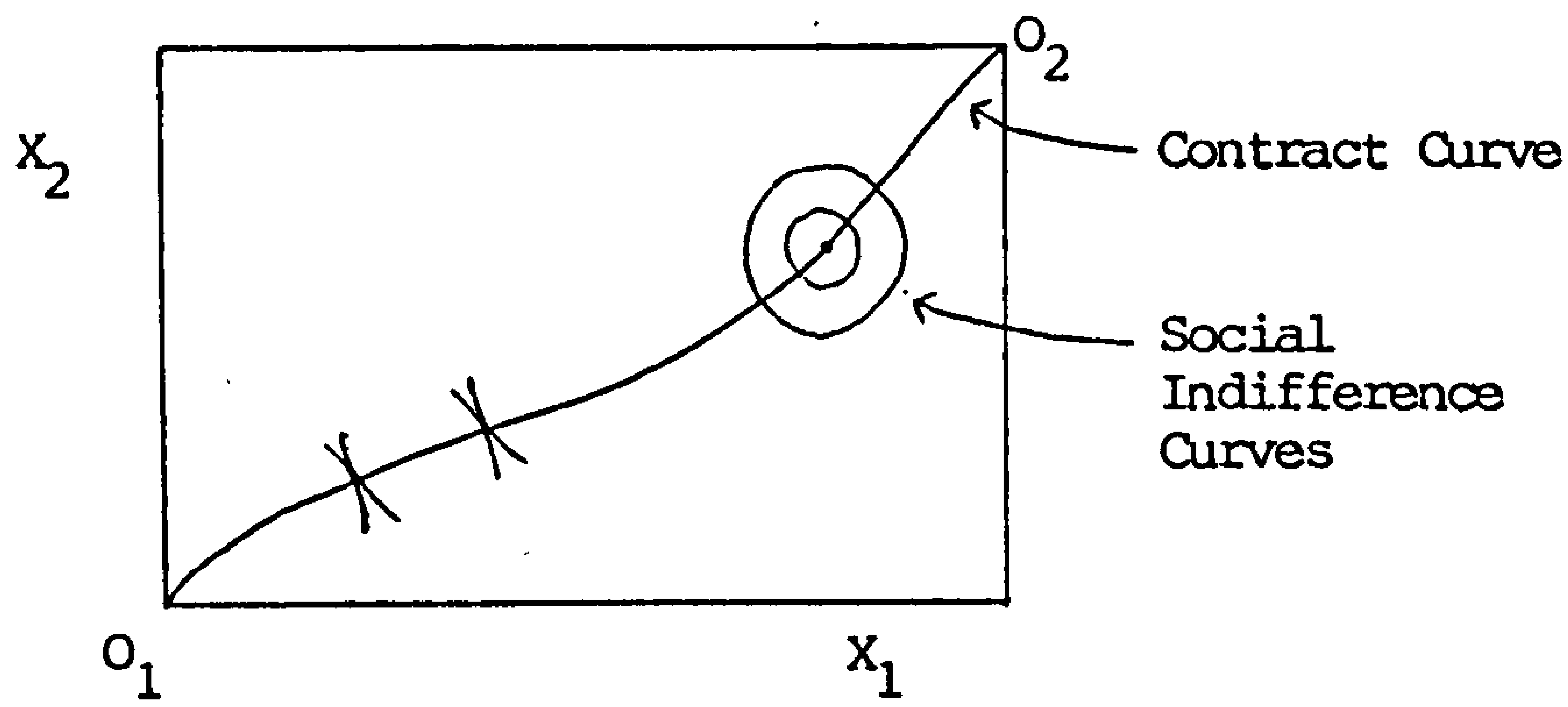
The position can be illustrated diagramatically for a two-good, two-consumer case. Let the first two goods be subject to an aggregate availability $\bar{X}_1$, $\bar{X}_2$, so that the allocation between the two individuals must satisfy

$$x_i^1 + x_i^2 = \bar{X}_i \qquad i=1,2$$

All other goods are sold in the normal manner, allowing preferences to be represented by the rationed indirect utility functions

$$\tilde{U}_h = \tilde{U}(\underline{q}, M_h, x_1^h, x_2^h) \qquad h=1,2$$

Any division of X_1, X_2 between the individuals can be represented as a point within an Edgeworth box of dimensions $(\bar{X}_1, \bar{X}_2)$. Pareto efficient allocations are given by the tangency points for the $\tilde{U}$ functions defining a contract curve in the usual way. The position is as below:



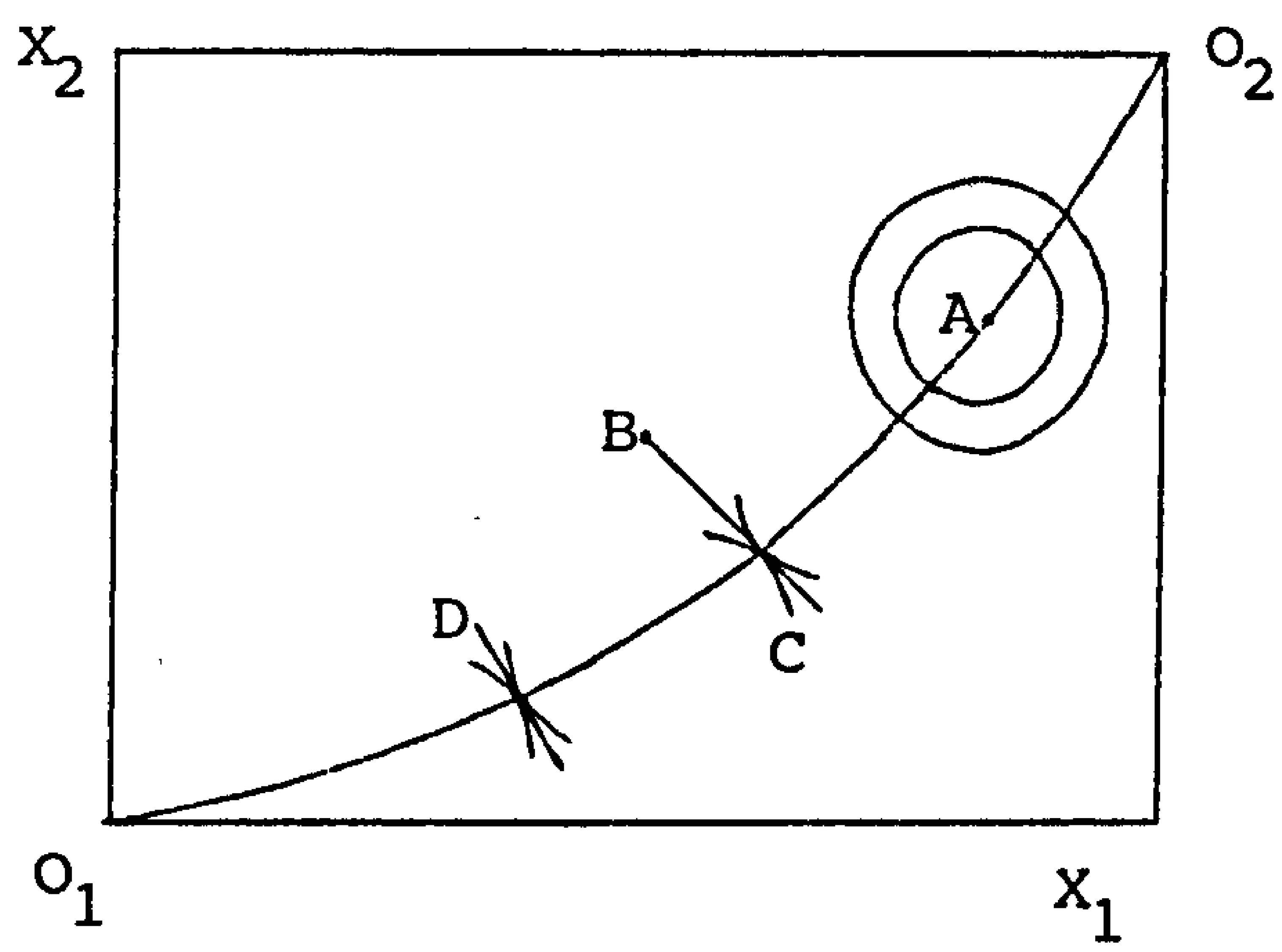
The diagram is set up in the conventional way, measuring the allocation received by individuals one and two from the bottom left-hand corner and top right-hand corner respectively. (In this case a fixed availability of the two goods is taken as literally true, so the Edgeworth box is perhaps more appropriate here than in its more general usage). On the contract curve the tangency of the $\tilde{U}$ functions guarantees that neither individual can be made better off without making the other worse off. The social valuation of different allocations is given by the social indifference curves, based on a welfare function $V=V(\tilde{U}_1, \tilde{U}_2)$. These curves satisfy the relationships

$$dV = \frac{\partial V}{\partial \tilde{U}_1}\left[\frac{\partial \tilde{U}_1}{\partial X_1} dX_1 + \frac{\partial \tilde{U}_1}{\partial X_2} dX_2\right] + \frac{\partial V}{\partial \tilde{U}_2}\left[\frac{\partial \tilde{U}_2}{\partial X_1}(-dX_1) + \frac{\partial \tilde{U}_2}{\partial X_2}(-dX_2)\right] = 0$$

$$\Rightarrow -\frac{dX_1}{dX_2} = \frac{\left[\frac{\partial V}{\partial U_1} \cdot \frac{\partial \tilde{U}_1}{\partial X_2} - \frac{\partial V}{\partial U_2} \cdot \frac{\partial \tilde{U}_2}{\partial X_2}\right]}{\left[\frac{\partial V}{\partial \tilde{U}_1} \cdot \frac{\partial \tilde{U}_1}{\partial X_1} - \frac{\partial V}{\partial U_2} \cdot \frac{\partial \tilde{U}_2}{\partial X_2}\right]}$$

where X_1, X_2 values are measured from the 0_1 origin. The resulting indifference curves are liable to be 'circular' with a negative slope in the central regions (similar distribution pattern for both goods) and positive slope towards the outer edges of the box (opposite distribution patterns for the two goods). If it exists, a social optimum will occur in the interior of the indifference map, at a point on the contract curve. This is the outcome of the direct quantity allocation of Section (1), part (a); on the diagram above it is drawn closer to O_2 than O_1, implying that individual one is in this case likely to have the lower income.

Using the Edgeworth Box to illustrate the alternative allocation methods yields the following:



Point A is the social optimum, located on the contract curve, and attainable by the direct allocation of Section (i). With symmetric social preferences a position such as A suggests that individual one is disadvantaged in terms of income, receiving a higher allocation of x_1 and x_2 at the social optimum in partial compensation.

An even distribution of x_1 and x_2 leads to the point B exactly in the centre of the Edgeworth Box. This is not in general on the contract surve, implying allocative inefficiency.

A voucher or coupon allocation scheme with uniform purchasing power produces a symmetrical budget line passing through B, with slope $-\beta_1/\beta_2$. If the 'prices' β_1 and β_2 are set to equate supply and demand, then the outcome is a point such as C on the contract curve. Point C is in general superior to B (unless they coincide), since B is available in each individual's budget set but remains unchosen. Hence a voucher scheme with equal purchasing power and efficient pricing is at least as good as an even quantity allocation.

Allocation by price gives a point on the contract curve, towards the bottom left-hand corner of the Edgeworth Box (on the assumption that individual two has the higher income and purchasing power). Although Pareto efficient, this outcome is inferior to C as long as the social optimum lies on the segment of the contract curve above C, favouring individual one. Such a position is likely to arise under normal circumstances, and, indeed, probably provides the main inducement to implement a rationing scheme. In comparison with point B the greater efficiency of price allocation may be offset by its bias towards higher income levels. The social ranking thus depends on the configuration of income levels, as well as on individual and social preferences.

Non-linear constraints are a theoretical, if not necessarily practical, possibility. Relative to linear constraints, their rationale depends on the social value of Pareto inefficient points, where no tangency between indifference curves or the budget line occurs. It is possible that a Pareto inefficient point achieved by non-linear constraints is superior to the outcome of a linear constraint, depending on individual and social preferences. The same goes for the comparisons with price allocation or quantity rationing, where non-linear constraints are potentially but not inevitably superior.

The full range of possible allocation schemes is illustrated by the following numerical example:

Example

Suppose that there are three goods, i=1,2,3, and two individuals, with preferences

$$U_A = (X_1^A)^{\frac{1}{6}} \; (X_2^A)^{\frac{1}{6}} \; (X_3^A)^{\frac{1}{3}} \quad ; \quad U_B = (X_1^B)^{\frac{1}{6}} \; (X_2^B-10)^{\frac{1}{6}} \; (X_3^B)^{\frac{1}{3}}$$

The first two goods are in short supply, such that

$$\bar{X}_1 = 60 \quad ; \quad \bar{X}_2 = 30$$

The remaining good ('all other consumption') is allocated on the basis of individual incomes, subject to a given price level p_3. Since all income must be spent on the third good (except under price allocation), it follows that

$$X_3^A = \frac{M_A}{p_3} \quad ; \quad X_3^B = \frac{M_B}{p_3}$$

Setting arbitrary numerical values $p_3 = 1$, $M_A = 125$, $M_B = 216$, one can write $X_3^A = 125$, $X_3^B = 216$. Given that no price is to be charged for the first two goods, the relevant utility functions are

$$\tilde{U}_A = 5(X_1^A)^{\frac{1}{6}} (X_2^A)^{\frac{1}{6}} \quad ; \quad \tilde{U}_B = 6(X_1^B)^{\frac{1}{6}} (X_2^B-10)^{\frac{1}{6}}$$

Social welfare, V, takes the form

$$V = U_A U_B$$

so that some value is placed on equality.

The outcomes of various possible allocation methods are as below:

<u>Direct Allocation (Social Optimum)</u> - This follows from maximising V subject to $\bar{X}_1 = X_1^A + X_1^B = 60$, $\bar{X}_2 = X_2^A + X_2^B = 30$

The solution is $X_1^A = 30$, $X_2^A = 10$, $X_1^B = 30$, $X_2^B = 20$

Hence $\underline{X}^A = (30, 10, 125)$, $\underline{X}^B = (30, 20, 216)$

$$U_A = 12.94 \quad , \quad U_B = 15.52 \quad , \quad V = 200.83$$

<u>Allocation by Price</u> - Prices for the first two goods are set so as to equate the demand for them with their availability. Using the individual demand functions, this means that

$$\frac{M_A}{4p_1} + \frac{M_B - 10p_2}{4p_1} = 60 \Rightarrow \quad 341 - 240\ p_1 - 10p_2 = 0$$

$$\frac{M_A}{4p_2} + 10 + \frac{M_B - 10p_2}{4p_2} = 30 \Rightarrow \quad 341 - 90\ p_2 = 0$$

Solving for p_1, p_2 gives $p_1 = 1.26$, $p_2 = 3.79$, implying a set of demands $x_1^A = 24.74$, $x_2^A = 8.25$, $x_1^B = 35.26$, $x_2^B = 21.75$

Hence $\underline{x}^A = (24.74, 8.25, 62.50)$, $\underline{x}^B = (35.26, 21.75, 89.06)$

$$U_A = 6.61 \quad , \quad U_B = 12.19 \quad , \quad V = 80.65$$

In this case the need to use income in purchasing X_1 and X_2 reduces the amount of X_3 received by both individuals.

Quantity Rationing - The first two goods are divided evenly between the individuals, such that

$$x_1^A = x_1^B = 30 \quad ; \quad x_2^A = x_2^B = 15$$

Hence $\underline{x}^A = (30, 15, 125)$, $x^B = (30, 15, 216)$

$$U_A = 13.84 \quad , \quad U_B = 13.83 \quad , \quad V = 191.43$$

General Linear Constraint (Voucher Scheme) - Individuals receive an identical endowment of purchasing power for goods 1 and 2 at an arbitrary face value of 100. The 'voucher prices' β_1, β_2 must satisfy

$$\frac{100}{2\beta_1} + \frac{100 - 10\beta_2}{2\beta_1} = 60 \Rightarrow \quad 200 - 120\ \beta_1 - 10\ \beta_2 = 0$$

$$\frac{100}{2\beta_2} + 10 + \frac{100 - 10\beta_2}{2\beta_2} = 30 \Rightarrow 200 - 50\ \beta_2 = 0$$

Solving for β_1, β_2 gives $\beta_1 = 1.33$, $\beta_2 = 4$, implying a set of demands $X_1^A = 37.5$, $X_2^A = 12.5$, $X_1^B = 22.5$, $X_2^B = 7.5$

Hence $\underline{X}^A = (37.5, 12.5, 125)$, $\underline{X}^B = (22.5, 17.5, 216)$

$U_A = 13.94$, $U_B = 14.10$, $V = 196.56$

Non-linear Constraint - Suppose that individuals face a non-linear constraint of the form

$$aX_1^2 + bX_2^2 = 100$$

Individual demands are such that

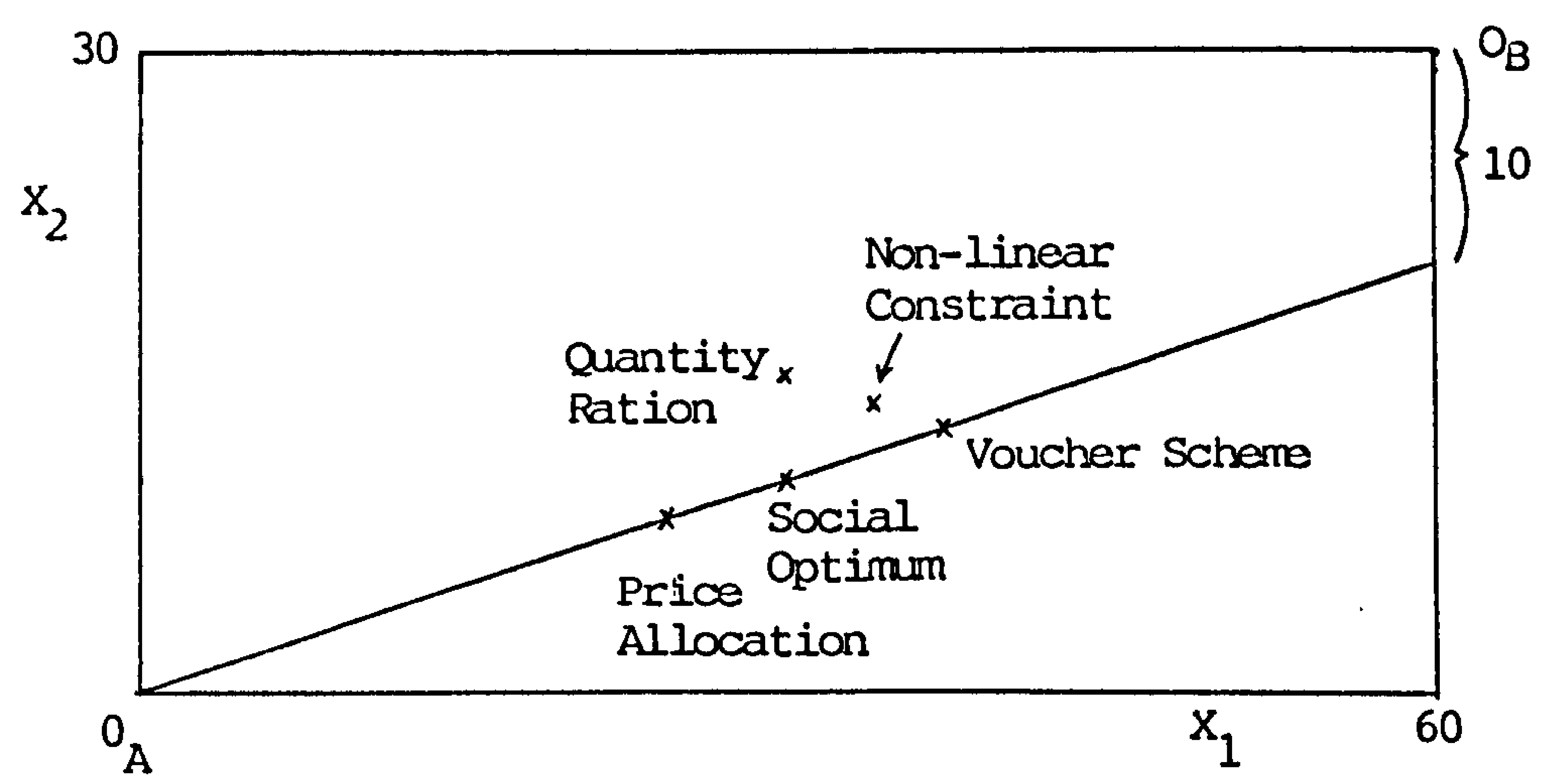
$$\frac{X_1^A}{X_2^A} = \frac{2bX_2^A}{2aX_1^A} \quad ; \quad \frac{X_1^B}{X_2^B-10} = \frac{2bX_2^B}{2aX_1^B}$$

Setting a and b to satisfy the equations above simultaneously with the resource constraints $X_1^A + X_2^B = 60$, $X_2^A + X_2^B = 30$ yields a solution $X_1^A = 34.32$, $X_2^A = 13.63$, $X_1^B = 25.68$, $X_2^B = 16.36$

Hence $\underline{X}^A = (34.32, 13.63, 125)$, $\underline{X}^B = (25.68, 16.36, 216)$

$U_A = 13.93$, $U_B = 14.03$, $V = 195.42$

The position diagrammatically is as below:



The contract curve is here given by the line $X_1^A = 3X_1^B$, on which the individual indifference curves have the same slope. The price, voucher and direct allocations all produce efficient points on the contact curve, but with widely differing levels of social welfare. By definition the optimal direct allocation must be the best of the three, and the voucher scheme dominates price allocation. An even share out of quantities leads to a point off the contract curve, but its more egalitarian impact makes it socially preferred to the more efficient distribution by price. The particular non-linear constraint used here fails to improve on the linear voucher scheme, leaving the latter as the best of the indirect allocation methods considered.

In general there is no definite ranking of indirect allocation schemes, the comparison depending on the particular case considered. Suitably chosen linear or non-linear rationing schemes should, however, be able to improve on allocation by price or quantity rations. As a feasible policy tool it would be hard to improve on a linear constraint allowing individuals to choose their own consumption pattern within the budget set offered. This can be seen as merely confirming the efficiency of consumer choice under 'monetary' allocation systems, although the efficiency counts for little when the income distribution is highly unequal. In the absence of a thorough income redistribution it is better to set up subsidiary allocation methods for scarce goods, rather than relying on the conventional method of a universal monetary distribution system.

It is possible to see the imposition of non-monetary constraints as a roundabout way of achieving price discrimination. Under perfect price discrimination (charging a separate price to each individual for each good) the first best can always be attained. A fully flexible choice of personalised prices is not achievable by the broad constraints assumed here, but some influence can nevertheless be exerted on the virtual prices faced by individuals. For a given income distribution, one can therefore use constraints as a means of shifting virtual prices towards the ideal set of personalised prices. The supposed benefits of uniform pricing are nullified when income is unevenly distributed, so there is little reason to preserve it.

When direct allocation is weighed against indirect, there are two particular situations which can be distinguished, according to whether or not the income distribution remains constant:

- For a fixed, arbitrary income distribution the best outcome achievable is the optimal direct allocation set out in part (i) above. Indirect allocation methods cannot improve on this unless they influence the underlying income distribution, an effect assumed not to occur here. If the optimal direct allocation is not attained (because of inability to observe certain characteristics, or for some other reason), then there arises some scope for indirect allocation. It remains true that the government should use any directly observed information in its possession, but this still leaves room for indirect allocation. When partial information is present the best approach is liable to be a mixture of the two methods, with a general functional form of constraint incorporating the available information on individuals; for example, one could have a voucher scheme where the allocation of purchasing power depends on observed individual characteristics.

- When income is freely redistributed on the basis of the available information, the need for direct quantity allocation is removed (except under the special conditions of 'specific egalitarianism'). Full information would lead to a genuine first-best social optimum, with no further need for redistribution. Partial information produces only a suboptimal income distribution, leaving a rationale for indirect allocation methods. This could possibly be restricted to a few scarce commodities, in the manner assumed above. It is also true that indirect allocation methods for a subset of 'scarce' goods can be used to influence the distribution of the remaining goods. Charging monetary prices for rations (possibly negative) could help to alleviate income inequalities, which may be of value if social preferences are egalitarian. Stronger effects are achievable when indirect taxes are present : the situation then is that of Guesnerie and Roberts (1984), where quantity controls influence tax revenue through 'forced' consumption of taxed commodities.

Unless the economy is at a first-best social optimum, there is always a case for allocation methods which diverge from the usual monetary allocation. This observation applies generally, and need not be limited to situations where a few goods are identified as being in short supply; the same conclusions would arise for any general type of production constraint.

(4) Conclusion

The analysis above does not seek to diminish the attractiveness of monetary allocation. It remains true that a palatable income distribution, efficient pricing and no other constraints would be the best way to allocate goods. Unfortunately, the concentration of all purchasing power in the single income parameter (as opposed to, say, differential pricing) implies that the allocation system is vulnerable to any inadequacies in the income distribution. Privilege in consuming one good means privilege in consuming all goods. In this way the very sensitivity of monetary allocation (which makes it a powerful agent for redistribution) acts against it when the income basis is deficient. The best remedy is to redistribute incomes, but failing this the next line of approach is to alter the allocation system.

Adding more constraints to allocation is hardly a panacea, as is demonstrated by the rarity of rationing and its unpopularity when it does appear. Further constraints could not remove any fundamental disincentives preventing redistribution (given that they are defined in quantity rather than income terms), and might not produce benefits sufficient to justify themselves as practical measures. The experience of the welfare state suggests that it is not easy to counteract the fundamental inequalities in society (see, for example, Le Grand (1982)). One should therefore be cautious in advocating non-monetary allocation, and should not give a false impression that allocation through incomes and pricing is a poor distributive tool; problems should be blamed on the way the tool is used rather than on the tool itself. This leaves a pragmatic conclusion : non-monetary allocation might be useful to us under the prevailing conditions, and thus deserves more attention than it has received in the past.

CHAPTER 4 : OPTIMAL WEALTH TAXATION

(1) Introduction

Of the many dimensions of inequality between individuals, theoretical discussion has concentrated mainly on the distribution of income. This is undoubtedly an important issue, but there is also a good case for looking at other aspects of inequality, the most conspicuous being the ownership of wealth. Statistics for the U.K. and elsewhere show that wealth is distributed in a highly unequal fashion, considerably more so than income (see, for example, Lydall and Tipping (1961), Atkinson and Harrison (1978), Harrison (1979), and Atkinson (1983), Chapter 7). The figures are sensitive to the particular definition of wealth being used, and still greater dispersion is observed in the ownership of certain assets, such as stocks and shares. A desire for equality, coupled with observation of the present wealth structure, would suggest the need to consider wealth taxation, both from a theoretical and practical perspective. The complexity of wealth holding means that a comprehensive theoretical discussion is virtually impossible, and the treatment has perforce to be selective. The rest of this introduction outlines some of the basic problems involved in wealth taxation, together with the particular approach to be used here.

A central difficulty facing the theoretical modelling of wealth is its close relationship with the economy's institutional structure. In a capitalist economy wealth is a major source of power and influence, both in specifically economic affairs and in society generally. A radical redistribution of wealth could therefore expect to meet strong political resistance, to the extent that its political feasibility is in doubt. This seems to be borne out by the extensive opposition faced by even minor increases in taxes facing the rich, and the rapidity with which appropriate loopholes are found. In the Marxian view a complete restructuring of capital ownership would require a change of economic system, entailing the end of the capitalist mode of production. This may well be true (given the absence of counter-examples), and it is certainly likely that any economy with a very egalitarian wealth ownership would be substantially different from those observed at present. The organisation of production in particular cannot be expected to be immune to changes in the wealth distribution. Such considerations imply that large scale wealth redistribution would be a difficult undertaking, raising issues which go beyond the normal subject boundaries of economics. When using simplified theoretical models it should not be forgotten that wealth ownership has substantial institutional and social aspects, which are not represented in a purely utilitarian and individualistic approach.

Issues frequently raised in discussing wealth tax proposals are the administrative difficulties in implementing the tax. These centre on the many forms in which wealth can be held, and the problems of identifying and quantifying them. A comprehensive definition of wealth would involve numerous separate items : among the possibilities are property, land, company securities, cash, bank deposits, life insurance policies, future pension rights and (as is sometimes argued) human capital, in the form of expected future earnings. Several of these assets are subject to fluctuating market prices, and are difficult to value accurately. Those items dependent on future expectations (such as pension rights or human capital) are the most problematical, and can only be treated inexactly using probabilistic assessments highly sensitive to the specific assumptions made. Consequently a wide ranging definition of wealth can only provide an approximate tax base, and is likely to mean significant administrative costs in obtaining the required information. A further administrative issue is the prevention of tax avoidance and evasion, so as to guarantee that the nominal tax rates are actually being paid. This will be an especially important concern with capital and wealth taxes, given that some individuals stand to lose a great deal from high tax rates. Illegal non-payment (tax evasion) could potentially be removed by stricter enforcement of the tax laws, presumably involving higher administrative expenditure. Where non-payment is legal (tax avoidance), there is little that can be done to restore the intended tax rates unless the definition of the tax is adjusted to remove the loopholes. In practice the movement has often been the other way round, with influential pressure groups securing legislation to introduce loopholes which lessen their tax burden. One can foresee that even if a substantial wealth tax were to be introduced, it would face continuing difficulties in its operation.

Assuming that the political and administrative obstacles can be surmounted, there arises the question of the design of a wealth tax. One possibility is an annual tax on wealth, according to some given definition, similar to the approach used for income taxes. Such taxes have been applied on a limited scale in some European countries, although never on a national basis in the U.K. They have occasionally been advocated (Flemming and Little (1974)), but local authority 'rates' remain the only significant annual property tax. The other main approach is to tax the transfer of wealth, rather than wealth holding, by imposing a tax on bequests or inheritances. This has been the main method applied in the U.K., although the taxes have always been easy to avoid (increasingly so in recent years; see Sandford (1983)), and have not brought about any dramatic changes in the wealth distribution. Taxation of wealth holding and transfer are not incompatible, and a full policy optimum is liable to include both to some degree. Inheritance issues are considered empirically by Harbury and Hitchens (1979) and theoretically by Meade (1964, 1973) and Stiglitz (1969). The discussion in this chapter concentrates instead on taxes within a given generation, that is, on a wealth tax in the literal sense. This does not imply that bequest or inheritance taxes are less effective redistributive tools, and it is, in fact, possible to argue the opposite case (Atkinson (1972)).

Even when attention is restricted to wealth taxes, there remain several ways in which the tax can be defined. These issues will be considered in the following sections.

Within an optimal policy model, a wealth tax does not incur the same disincentives as an income tax, but other forms of individual response may be present. Two important effects are the shifting of wealth between different assets and the impact of the tax on saving behaviour. The former is essentially the tax avoidance issue, whereby the imposition of a wealth tax induces wealth holders to transfer their wealth into untaxed assets or to move it abroad. The extent to which this occurs depends on the costs involved, and whether or not the introduction of the tax is fully anticipated. This aspect is considered in Section (4), and in practice is closely bound up with administrative issues, as well as the purely economic effects of the tax. A saving disincentive may arise because an annual wealth tax reduces the net returns to wealth, and consequently disrupts the pattern of wealth accumulation. One way to consider this is within the context of economic growth models, as in Atkinson and Sandmo (1980). Attention here is focused more on the redistributive role of wealth taxation, where social welfare depends on the current consumption levels of individuals, which in turn are influenced by wealth taxation. Nevertheless, it is important to be aware of the numerous links between wealth ownership and production, which could lead to tax effects different from those represented in models of interpersonal redistribution.

The discussion of this chapter is divided as follows. Section (2) considers taxes based solely on the current observed wealth of individuals, without taking account of the taxpayer's total lifetime wealth. Section (3) allows taxes to be imposed on lifetime or 'true' wealth, which can be accomplished in a life-cycle model by letting tax rates depend on the individual's age as well as the current wealth level. Section (4) looks at the issues which must be faced when the introduction of taxation is anticipated by taxpayers.

(2) Optimal Taxation of Current Wealth

The most basic annual wealth tax depends on estimates of each individual's current wealth at the date when the tax is imposed. This neglects the possibility of systematic variation in wealth holding over the life cycle, so that current wealth may not be a reliable indicator of an individual's true wealth (Atkinson (1971 a & b). For example, at a given level of current wealth the taxpayer could be a relatively poor person at an age when wealth is near the maximum or a relatively rich one at an age when it is near the minimum. To separate these individuals it may be preferable to base taxes instead on 'true' lifetime wealth, by including information on the person's age in the tax system. The policy outcome depends significantly on which approach is used; taxes on current wealth are discussed in this section, and taxes on true wealth in Section (3).

The policy framework is the same as previously, with the government aiming to maximise social welfare by means of redistributive taxation. Individuals ae assumed to organise their saving and consumption in a rational manner over the life cycle, following the optimal saving of Ramsey (1928). It can be argued that most individuals do not save in such a systematic fashion, and that there would be little response to marginal changes in tax rates. This could well be the case, but the life-cycle model has the virtue of emphasising any disincentives present, and thus errs on the pessimistic side. Generally, the less is the individual behavioural response the greater are the prospects for redistribution, so any inaccuracy in this respect does not diminish the value of wealth taxation. Savings are sometimes modelled as being determined on class lines, arising mainly out of profit as opposed to wage income, or alternatively as being institutional in nature. These views are quite compatible with a role for personal wealth taxation, and, if anything, suggest a lesser degree of difficulty with individual disincentives (leaving aside the possible institutional aspects). The theoretical approach taken below can be seen as exaggerating rather than underestimating the saving disincentive from taxing personal wealth.

It is assumed that the population differs only in initial wealth holdings, K_o, distributed as $F(K_o)$, and returns to wealth are the only income source. Activity is restricted to consuming part of interest income and saving the rest, although it would be straightforward to include employment in the model. Consumption of income yields utility $U(C)$, where C is consumption and U satisfies $U'(C) > 0$, $U''(C) < 0$ (so that utility is derived only from the income return to wealth, not from wealth itself). Time is continuous, and individuals have a known life span, with a certain value being placed on wealth left over at the terminal date, T. The individual objective is to maximise the discounted intertemporal sum of utilities plus the value of terminal wealth

$$U^* \equiv \int_{t=0}^{T} e^{-\rho t} U(C_t)\, dt + e^{-\rho T} Z(K_T)$$

where ρ is the discount rate on future utility, Z is the current valuation of terminal wealth and K_t denotes the capital held at date t, for $t \epsilon [0,T]$. Income returns to wealth are linear, equal to rK_t, but a non-linear formulation could also be used if preferred. Wealth taxation is described by the general function $B(K_t)$, leaving a net return to wealth of $rK_t - B(K_t)$; the tax can also be interpreted as levied on the income derived from wealth, and a clear distinction between wealth and income taxes appears only when alternative income sources are introduced. By definition all income at a given date must either be consumed or saved, implying the condition

$$rK_t - B(K_t) = \dot{K}_t + C_t$$

where the dot denotes differentiation with respect to time, so that $\dot{K}_t$ is saving at time t. The individuals full optimal saving problem is to maximise U^* subject to

$$\dot{K}_t = rK_t - B(K_t) - C_t \qquad \forall_t$$
$$K_o = \bar{K}_o$$

for some fixed level of initial wealth, $\bar{K}_o$. A solution can be obtained by optimal control methods, with K as a state variable and C as a control. The Hamiltonian and first-order conditions are:

$$H = e^{-\rho t}\, U(C) + e^{-\rho t}\, \lambda[rK-B(K)-C]$$

$$H_C = e^{-\rho t}\, U'(C) - e^{-\rho t}\, \lambda = 0 \qquad \Rightarrow \quad \lambda = U'(C)$$

$$H_K = e^{-\rho t}\lambda\, [r-B'(K)] = -e^{-\rho t}\, \dot{\lambda} + \rho\, e^{-\rho t}\lambda$$

with transversality condition $\lambda_T = U'(C_T) = Z'(K_T)$

Substituting for λ in the condition for K implies that

$$\dot{C}_t = \frac{-U'(C_t)\ (r-\rho-B'(K_t))}{U''(C_t)} \qquad \forall t$$

which is a standard optimality condition for a saving problem. A relationship of this type applies to all individuals, with the appropriate K_o superscript.

The population is constant and in a steady state, so that the cross-sectional distribution $f(K_o)$ is invariant over time and forms the basis of utilitarian social welfare. In reality the position is almost certainly much more complicated, with the presence of population growth, inheritance behaviour and general uncertainty about life spans. Policy design including all these features could only be undertaken over a long-term planning horizon, and would be difficult to model theoretically. Although a steady-state model must in literal terms be inaccurate, it can be viewed as an approximation to a society with a relatively stable population and pattern of wealth ownership. Assuming a steady state means that there is no redistribution of true, rather than current, wealth; individuals are unidimensional and differentiated only by their true wealth, K_o, which has a distribution invariant over time. The current wealth tax $B(K_t)$ does not distinguish the age of the wealth holder, and consequently cannot bring about any changes in the underlying distribution $f(K_o)$. Inheritance behaviour would in practice mean that $f(K_o)$ is linked in some way with terminal wealth, leading inevitably to variations over time in $f(K_o)$. This is absent from the present model, so there is no explanation of wealth transmission between generations.

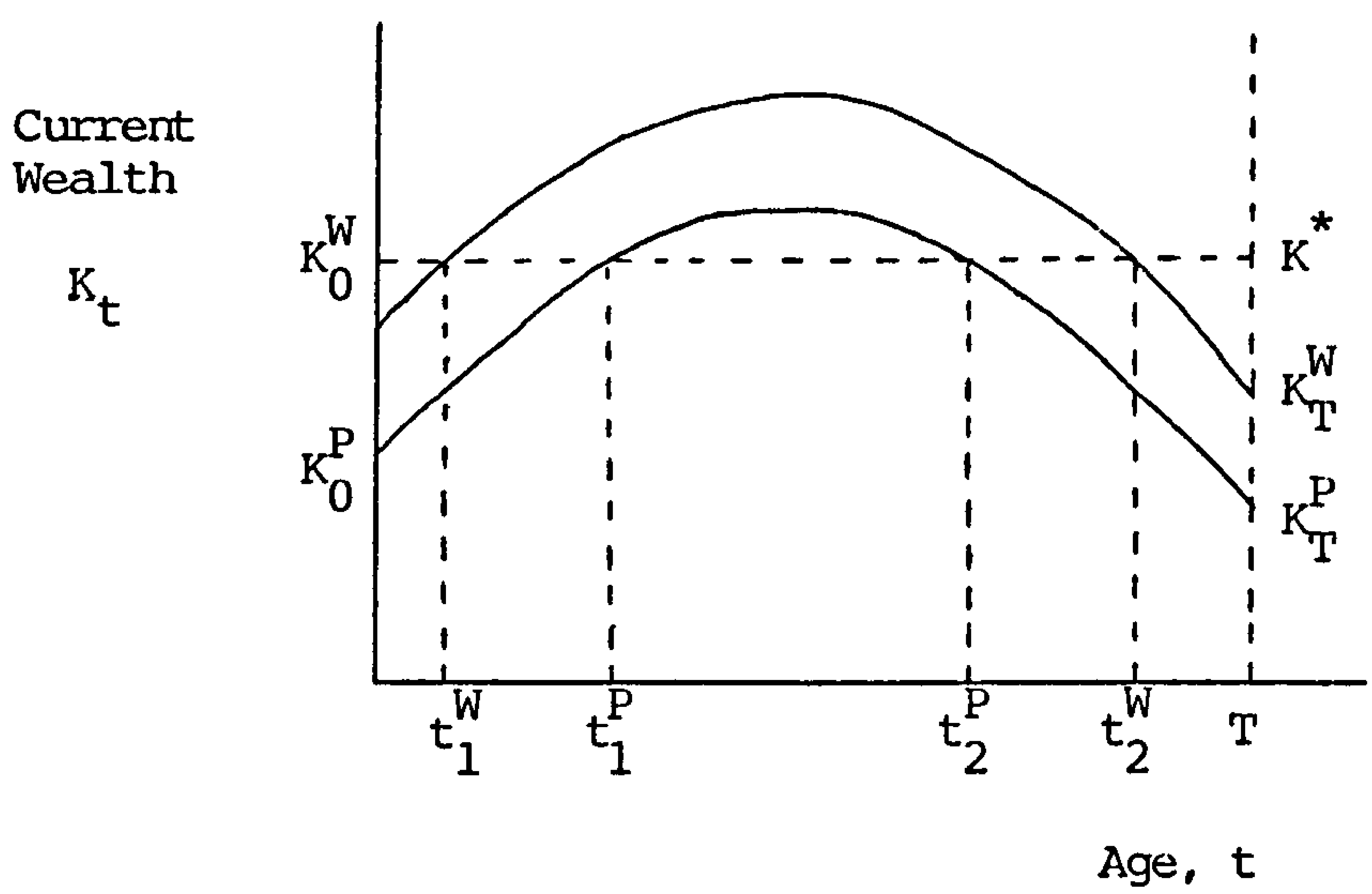
Under the assumptions above, the government's policy problem can be viewed in terms of a single generation born at the same date and living through the common life span of 0 to T. If consumption and current wealth are indexed by the initial wealth holding, K_o, then social preferences can be written as

$$V = \int_{\underline{K}_o}^{\overline{K}_o} \left[\int_{t=o}^{T} e^{-\rho t} U(C_t^{K_o})\, dt + e^{-\rho T} Z(K_T) \right] f(K_o)\, dK_o$$

assuming a standard utilitarian form. For the budget to balance at each date it must be the case that

$$R = \int_{\underline{K}_o}^{\overline{K}_o} \int_{t=o}^{T} B(K_t^{K_o})\, dt\, f(K_o)\, dK_o$$

where R is the revenue requirement in each period. The government's problem is to choose B(K) to maximise V, subject to the revenue constraint and the individual optimal saving conditions above. This is potentially highly complex, and a complete analysis is not given here. Some idea of the issues involved can be gained by considering the likely time paths of current wealth holdings. These will in general adhere to a 'hump saving' pattern (Harrod (1948)), as below



The diagram shows the probable life-cycle saving behaviour of two individuals, designated 'wealthy' and 'poor' according to their initial wealth holdings, K_o^W and K_o^P. With the same preferences current wealth holdings will usually be an increasing function of K_o at all dates, so that the two time paths do not cross. It can be seen from the diagram that certain current wealth levels, such as K^*, are held by both individuals at two separate stages of the life cycle; the poorer individual holds K_* at ages closer to the centre of the life span, so that $t_1^W < t_1^P$ and $t_2^P < t_2^W$. A tax imposed on current wealth K^* therefore influences the saving behaviour of both individuals at two distinct ages, and the same is true for other current wealth levels. More extreme values of current wealth may not be held at all by certain individuals, or held only once. Given that the optimal saving paths are functions of the wealth tax rate, the choice of $B(K)$ is clearly going to have an intricate effect on saving patterns, and the nature of the optimal $B(K)$ need not be straightforward. As often occurs with non-linear taxation general conclusions about the shape of the optimal tax schedule are not available.

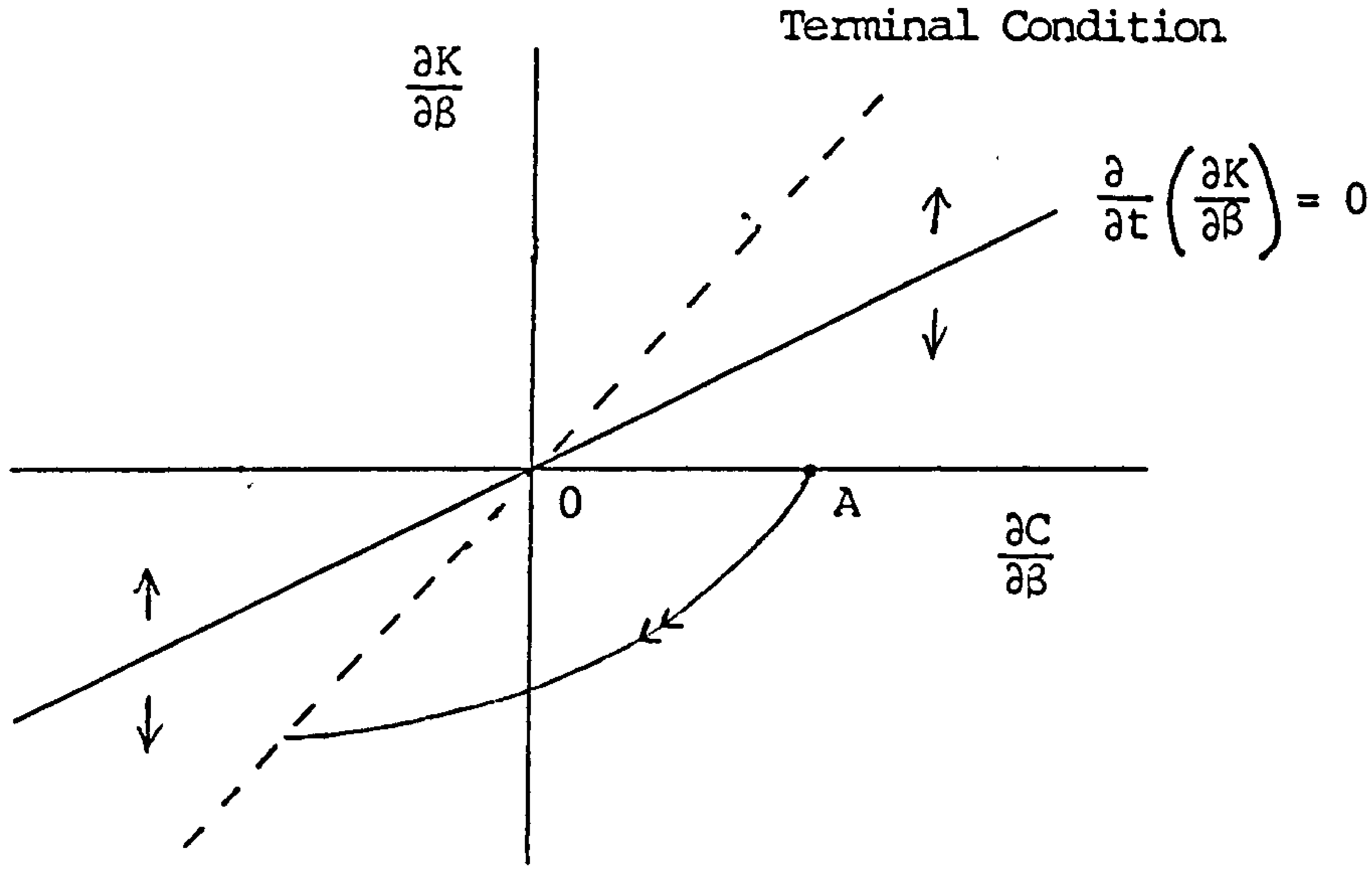
Nevertheless, it is possible to make some observations about the signs of marginal tax rates. In particular, it cannot be guaranteed that the optimal tax schedule $B^*(K)$ has non-negative marginal tax rates at all levels of current wealth (as might be expected). To show this, suppose that there exists an optimum at which $B^{*\prime}(K) \geqslant 0$ at all K, including the highest observed value K^{max}. It is known that K^{max} is attained only by the wealthiest group of individuals at their maximum current wealth. Suppose that the next wealthiest group has a discretely lower K_o value, so that a change in the marginal tax rate at K^{max} affects only those with true wealth $\bar{K}_o$. This is a reasonable assumption, since at the upper extremes of the wealth distribution we are more likely to be talking about discrete individuals than continuous groups. Let $B^{*\prime}(K^{max})$, denoted by β, be reduced marginally to $\beta - d\beta$, and consider the effect on the optimal saving plan of the wealthiest individual. Differentiating the savings constraint with respect to β gives

$$\frac{\partial \dot{K}}{\partial \beta} = \frac{\partial}{\partial t}\left(\frac{\partial K}{\partial \beta}\right) = (r - B'(K))\,\frac{\partial K}{\partial \beta} - \frac{\partial C}{\partial \beta}$$

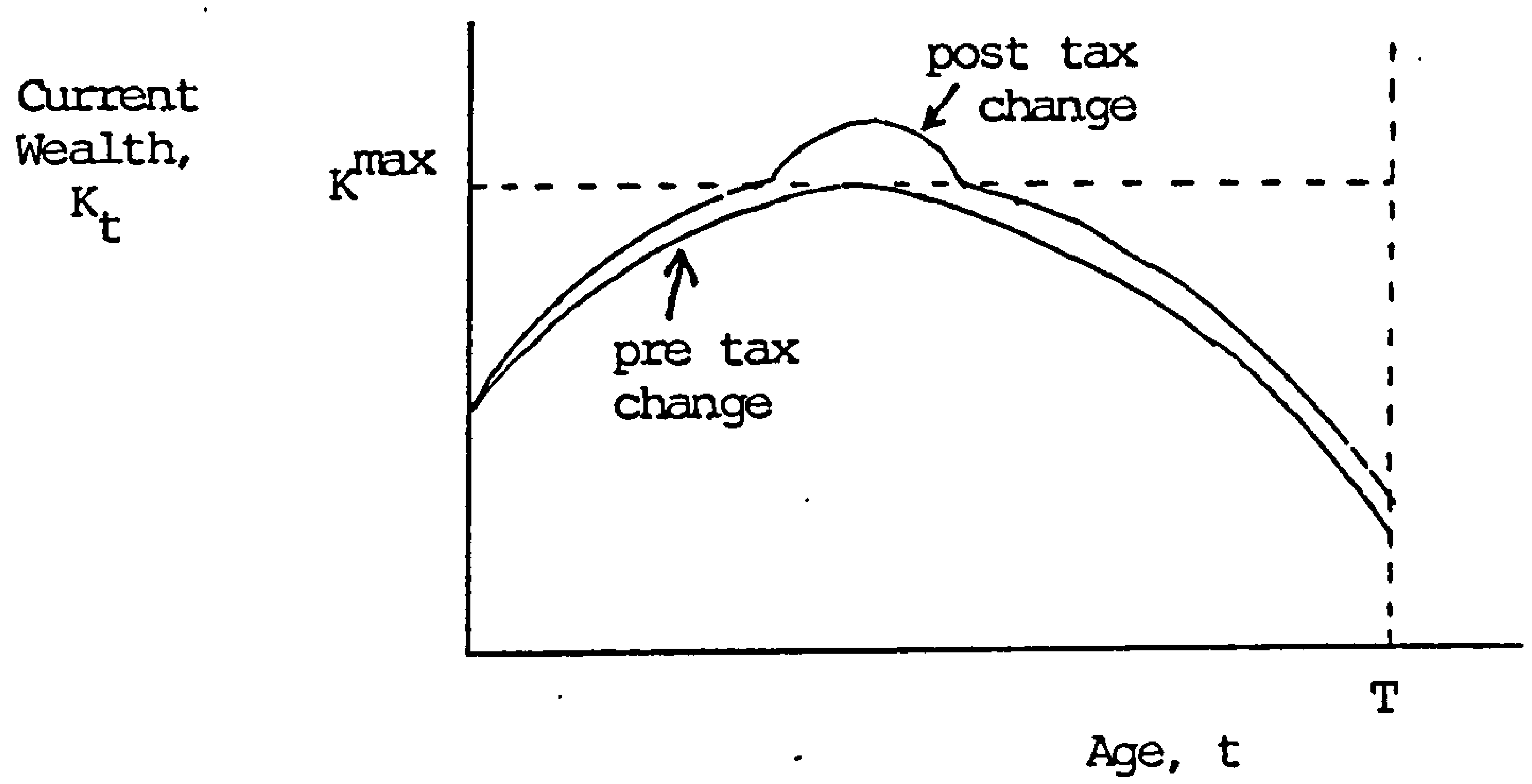
and differentiating the terminal condition $U'(C_T) = Z'(K_T)$ yields

$$U''(C_T)\,\frac{\partial C_T}{\partial \beta} = Z''(K_T)\,\frac{\partial K_T}{\partial \beta}$$

These relations can be depicted on a phase diagram relating $\partial K/\partial\beta$ to $\partial C/\partial\beta$, as below



The equation $\frac{\partial}{\partial t}\left(\frac{\partial K}{\partial \beta}\right) = 0$ is a straight line with slope $1/_{r-B'(K)}$, above which $\partial K/\partial\beta$ is rising and below which $\partial K/\partial\beta$ is falling. It rotates about the origin as K varies over time, but is positively sloping as long as taxation does not overcome the interest return to wealth. The terminal condition is also linear, with slope $U''(C_T)/Z''(K_T) > 0$, and may or may not be steeper than $\partial/\partial t\ (\partial K/\partial\beta) = 0$ (as makes no difference in this case). Since a change in β has no influence on K_o, the time path of $\partial K/\partial\beta$ against $\partial C/\partial\beta$ on the phase diagram must start on the $\partial C/\partial\beta$ axis. Under plausible assumptions $\partial C_o/\partial\beta > 0$ also holds true. To see this consider the fact that $\dot{C} < - U'/U''(r-\rho-\beta+ d\beta)$. If $- U''/U'$ falls with higher C (decreasing absolute risk aversion), then lower consumption at K^{max} is required to restore the optimal saving condition. For $K < K^{max}$ there is no change in tax rates, and the time path of consumption is governed by the same equation as previously. Lower consumption at K^{max} is achievable only if $\partial C_o/\partial\beta > 0$, reducing the values of $\dot{C}$ and C at each given level of K; saving then proceeds more quickly and K^{max} is attained in a slightly shorter period. On the phase diagram, the time path of $\partial K/\partial\beta$ against $\partial C/\partial\beta$ starts at a point like A and finishes on the terminal condition line, never going above the $^{\partial K}/_{\partial \beta}$ axis. It follows that a reduction of the marginal tax rate at the top end of the scale leads to an increase in wealth holdings at all points of the life cycle. The effect on the time path of K_t is as below



with the individual induced to enter the new tax range beyond the previous K^{max} (although, unlike the above diagram, the effect is infinitesimal). Because current wealth is higher at all ages and $B^{*\prime}(K) \geqslant 0$, it must be the case that revenue receipts from the wealthiest individual are increased by the tax change. This applies even if $B^{*\prime}(K^{max})$ was originally zero, as the decreased tax paid at the single point K^{max} is outweighed by the increased tax receipts elsewhere in the life cycle. The wealthiest individual must be better off after the tax change (since the initial savings path could still have been chosen) and nobody else is affected (given the discrete separation of $\bar{K}_o$), so there exists a perturbation in the tax schedule which both adds to government revenue and increases social welfare. It cannot therefore be true that $B^*(K)$ is optimal, and the true optimum has a negative marginal tax rate at some point of the tax schedule. Such an outcome does not match the conventional view of redistributive wealth taxes, and suggests that this type of model could lead to complex and counter-intuitive tax optima.

The result is stronger than those typically found under optimal non-linear income taxes (Seade (1977)), since marginal tax rates are here known to be negative at some point, rather than zero.

Evidently current wealth taxation is not an ideal redistributive tool in an economy with overlapping life-cycle savings patterns. The problem is basically that the tax imposed on any particular wealth level must be paid by all individuals holding that wealth, regardless of the stage they have reached in the life cycle. The tax cannot therefore distinguish the true wealth levels of taxpayers, which hampers its ability to implement redistribution. Such difficulties are reduced where there is little overlap between individual life-cycle patterns, so that current wealth holding gives a better indication of true wealth (although a large overlap implies that there is less need for redistribution in the first place). In the extreme case of no overlap between wealth holdings at any stages in the life cycle it would be possible to identify true wealth from current wealth holdings, allowing a full redistribution of true wealth. This can also occur if no life-cycle saving behaviour is taking place, with capital holdings constant, and current wealth accurately indicating true wealth. In populations with overlapping wealth patterns, however, the shortcomings of a tax on current wealth encourage the inclusion of information on true wealth in the tax system. This possibility is considered in the next section.

(3) Optimal Taxation of True Wealth

When capital holdings vary over the life cycle, observation of current wealth is insufficient to indicate 'true' wealth, interpreted as total lifetime consumption plus any terminal wealth. At the most, if the individual's preferences are known, observation of current wealth gives a range of possible true wealth values within which an individual lies. If true wealth is to be identified, then information must be 'dated' by the individual's age, regardless of the particular point in the life cycle observed. Information on true wealth (assuming known preferences) can be derived from observing the initial or terminal wealth level, the wealth holding at any age in between, or the current value of the total lifetime wealth at any age. True wealth as defined here is essentially a two-dimensional entity, involving a current wealth value (either the current holding or a current valuation of lifetime wealth) and the age of the individual by whom it is held. Taxes in the previous section were based only on the current wealth dimension, neglecting the age dimension; it is equally possible to have the opposite, with a 'wealth' tax based only on the age dimension. But for wealth taxation to be implemented in the true sense of the term, the aim should be to tax true wealth, allowing both for current wealth and age. The discussion below considers the issues involved when redistributive taxes are based on true wealth.

Suppose that individuals have identical preferences, but differ in their true wealth endowments, denoted by their wealth at age zero, K_o. Life spans are assumed to be infinite, and the population is in steady state, such that its cross-section is equivalent to the life cycle of a single individual. At each date a new generation of fixed size, H, is born, possessing a given exogeneous distribution of wealth endowments, $f(K_o)$. As in Section (2), individuals pursue optimal saving behaviour, subject to an interest return, r, and a common utility discount rate, ρ. The government can observe K_o, and is therefore able to tax true wealth.

Basic Case : the Desirability of a Redistributive Capital Levy

The government wishes to maximise a social welfare objective of the form

$$V = \int_{\underline{K}_o}^{\overline{K}_o} \int_{t=0}^{\infty} e^{-\theta t} U(C_t^{K_o})\, dt\, f(K_o)\, dK_o$$

where θ is the discount rate applied to utilities. If $\theta = 0$, then V can be seen as taking a cross-sectional form, whereas $\theta > 0$ suggests that the government is discounting the future utilities of each generation. Observability of K_o means that a tax schedule $b(K_o)$ can be imposed.

Assume initially that $\rho = \theta$, so that the pattern of individual savings exactly mirrors the government's. With no conflict over saving behaviour it remains only for the government to impose redistributive taxes on initial wealth, subject to a budget constraint

$$\int_{\underline{K}_o}^{\overline{K}_o} b(K_o)\, K_o\, f(K_o)\, dK_o = 0$$

Given that $U''(C) < 0$, it is always optimal to equalise initial wealth holdings, so that

$$b^*(K_o) = \frac{K_o - \tilde{K}_o}{K_o}$$

where $\tilde{K}_o$ is mean initial wealth. Taxation of this sort has no irregular features, and is always progressive in average terms, with $\partial b / \partial K_o = \tilde{K}_o / K_o^2 > 0$ No further taxation is needed at later dates, and the outcome is a first best.

When $\rho \neq \theta$ there is disagreement between individuals and the government over optimal saving behaviour. The most likely case is where $\rho > \theta$, so that individuals discount utility at a higher rate than the government : this happens, for instance, when a cross-section social welfare objective is being used. The divergence in saving attitudes can be removed if an appropriate flat rate wealth tax (or, equivalently, an interest rate tax) is used to influence individual saving. If a constant tax of b is imposed, then optimal saving satisfies the relation

$$\dot{C} = \frac{-U'}{U''} (r-\rho-b)$$

The government, on the other hand, would like to adhere to the time path

$$\dot{C} = \frac{-U'}{U''} (r-\theta)$$

Comparing the two equations, setting a constant tax of $b = \theta - \rho$ is enough to remove the discrepancy between them. When $\theta > \rho$ a positive tax is imposed in order to slow down individual savings, whereas $\theta < \rho$ (the more likely case) leads to a subsidy to encourage wealth accumulation. As previously, a first date set of wealth transfers can be implemented to equalise true wealth holdings. It remains only to combine these policies within a single revenue constraint. For each generation the current value (at age zero) of the receipts from the interest rate tax is

$$-R_o(K_o^\dagger) = (\theta-\rho) \int_{t=0}^{\infty} e^{-rt} K_t^\dagger(K_o^\dagger)\, dt$$

where $K_t^\dagger$ is the aggregate capital holding on the optimal saving path and R_o is the associated revenue requirement. R_o rises in magnitude with aggregate (post redistribution) initial wealth $K_o^\dagger$, and so $K_o^\dagger$ is determined by the relation

$$\int_{\underline{K}_o}^{\overline{K}_o} K_o\, f(K_o)\, dK_o = K_o^\dagger + R(K_o^\dagger)$$

where the left-hand side is the exogenous aggregate wealth endowment. The policy optimum therefore takes the form:

At t=0 Redistributive wealth transfers	At all other t Interest rate taxation
$b(K_o) = \dfrac{K_o - \tilde{K}_o + {}^{R_o}/_H}{K_o}$	$b = \theta - \rho$

with R_o satisfying the relations above. Taxes are divided into two distinct elements, a set of wealth transfers imposed at the outset, and an invariant interest rate tax guiding subsequent saving behaviour.

In practical terms the redistributive element of the optimum is best described as a capital levy, involving a once and for all rearrangement of a particular generation's assets. An annual redistributive wealth tax is not needed, and the role of annual taxation is restricted to the encouragement or discouragement of saving. Strictly speaking, a fully anticipated levy can be imposed at any stage of the life cycle, since it is based on invariant lifetime wealth. The choice of implementation at age zero is therefore arbitrary, although it seems as reasonable an assumption as any. At later dates the initially wealthy might be tempted to over-consume at early ages to prevent payment of their tax bill; if this did happen, it might lead to bargaining equivalent to that in Section (4). The scope for strategic suboptimal saving is minimised by early wealth redistribution, as in the model. Otherwise the instantaneous nature of a capital levy removes any saving disincentives, leading to a first-best outcome.

Exception: Increasing Returns to Wealth

A single-date redistribution of assets can be regarded as the basic wealth tax optimum whenever true wealth is identified. The one major exception is when increasing returns to wealth are present, so that a larger stock of wealth generates a higher income per unit of capital. In these circumstances an equalisation of true individual wealth minimises total income, creating a potential conflict between the redistributive desire to equalise consumption and the need to achieve the highest possible aggregate income. The difficulty can be avoided by adopting a centralised solution, in which all wealth is held by the government, and saving decisions are made collectively rather than by individuals. Instead of equalising inital wealth, the government confiscates the endowments into a central fund and subsequently pays out uniform incomes at the level commensurate with optimal aggregate saving. Where individual saving opportunities remain open, the government can ensure voluntary compliance with its plans by offering to subsidise private returns to wealth up to the social level. If, for instance, M(K) denotes returns to wealth (with $M'>0$, $M''>0$), then a guaranteed return to private wealth of (1/N)M(NK) ensures that individual saving plans coincide with the government's. Since incomes conform to the optimal consumption pattern, it follows that individuals choose to consume their current incomes; no private savings occurs, and there is no need actually to pay out the (hypothetical) subsidy. Wealth holding in these circumstances is a 'natural monopoly', requiring a concentration of wealth in order for a first best to be attained. Although the implications for wealth ownership are very different from the standard case, the impact on consumption patterns is similar, and in the present framework individuals are no worse off for the loss of their private wealth holdings. A real policy conflict arises only if a decentralised solution is felt to have some intrinsic value, to be set against the associated loss of aggregate income.

Regardless of whether or not the model produces a centralised optimum, one can always avoid saving disincentives by taxing lifetime wealth. Compared with Section (2) the only additional information requirement is to know an individual's age, something which is well defined and readily observed. It may not be easy in practice to estimate lifetime wealth, but the main difficulties are common to a current or lifetime definition, so that current wealth based measures have no real advantage on these grounds. Saving disincentives need not be a major obstacle to wealth redistribution, provided that policy concentrates on lifetime wealth.

(4) Anticipation of Tax Implementation

Given the outcome of Section (3), the main obstacle to wealth redistribution is likely to be tax avoidance. This section considers the problems posed when tax implementation is anticipated by wealth holders.

In practice a wealth tax cannot easily be introduced without wealth holders having foreknowledge of the policy measures. If an extensive wealth redistribution results from an open democratic decision, then the wealthy have both the incentive and the opportunity to find out about it and take appropriate action. At the very least they can consume much of their wealth before taxes are enforced (Aumann and Kurz (1977)), and they can probably do better than this by finding a means of tax avoidance; for instance, by converting wealth into alternative untaxed forms or by transferring it abroad to somewhere with lower tax rates. When such responses occur, it is in the government's interest to make specific allowance for them in formulating policy.

To model this theoretically one has to make a distinction between the date at which policy decisions are made and the date at which they are implemented. The steady-state framework in the previous sections fails to do so, leaving no time period when the tax regime is not in force and avoidance can occur. The discussion below instead uses a simple two date model, with a delay between the formulation and introduction of policy: at date one a public decision is made about policy measures, which are then fully anticipated at date two. It is convenient to distinguish between general anticipation by all individuals and a case where anticipation concerns only a wealthy subset of the population.

Generally Anticipated Tax

In this case all individuals are aware of the first-date policy measures, and make an appropriate response at the second date. Since the whole population is involved, it is assumed that individual responses are unco-ordinated and that the model is non-co-operative in nature.

All individuals are identical apart from their initial wealth, and make a single consumption/saving decision at date one. If K is wealth at date one and C_1, C_2 are consumption at dates one and two respectively, then the budget constraint is $C_2 = (1+r)(K-C_1)$. The individual chooses C_1 to maximise

$$U^* = U(C_1) + \frac{U((1+r)(K-C_1))}{1+\rho}$$

yielding an optimality condition

$$U'(C_1) = \frac{1+r}{1+\rho} U'(C_2)$$

Marginal utilities at each date are equated, with a suitable allowance for the interest and discount rates.

If the government's taxation ignores anticipation effects, it aims to maximise social welfare at date two; with utilitarian social preferences this simply means an equalisation of whatever wealth is held at that time. Facing such taxes the i^{th} individual can expect a second date consumption of

$$C_2^i = \frac{1}{N} [(1+r)(K^i - C_1^i) + \sum_{j \neq i} \tilde{C}_2^j]$$

where $\tilde{C}_2^j$ denotes the anticipated date two wealth (before taxation) of other individuals and N is the size of the population. Optimal saving then satisfies

$$U'(C_1^i) = \frac{1}{N} \cdot \frac{(1+r)}{(1+\rho)} U'(C_2^i)$$

with C_2^i given by the equation above. First date consumption therefore depends on expectations of the behaviour of others, which may or may not be accurate. The larger are the anticipated wealth holdings of others, the smaller is first period saving. In general, out of any savings the individual can consume only a fraction ${}^{1+r}/_N$ in period two, so if N is large the marginal returns to saving are close to zero. This is partly counteracted by the rise in $U'(C_2)$ as saving falls and by possible low anticipated values of other people's saving, but these effects are likely to be outweighed by the impact of the redistributive taxation. Equalisation of date two consumption can therefore be expected to induce a collapse of saving behaviour and considerable reduction of date two consumption. The outcome can easily be worse in social welfare terms than the no policy situation, and there is little justification for undertaking redistribution in this way.

If the government is to maximise social welfare, it must make allowance for individual anticipation of policy responses. Suppose that in the two period model above, the population is divided into two groups, a 'wealthy' group comprising a proportion α^W of the population and owning initial wealth K^W, and a 'poor' group comprising $\alpha^P = 1 - \alpha^W$ of the population and each owning wealth K^P. The government wishes to set fully anticipated taxes b^W and b^P which maximise social welfare over both periods and satisfy a revenue constraint

$$R = \alpha^W b^W (1+r)(K^W - C_1^W) + (1-\alpha^W) b^P (1+r)(K^P - C_1^P)$$

where taxes are imposed on net wealth at the start of the second period. Provided that the government is aware of individual anticipation effects, it is able to impose them as constraints on the optimal policy decision. Individuals adhere to the condition

$$U'(C_1) = \frac{(1+r)(1-b)}{1+\rho} \; U'((1+r)(1-b)(K-C_1))$$

for both wealthy and poor groups. The outcome of this adjustment can be summarised by the functions $C_1^{*j}(b^j)$, j=w,p, giving the response of first date consumption to the rate of wealth taxation. It is probable that $\partial C_1^{*j} / \partial b^j > 0$, so that rising taxes reduce saving and increase first period consumption, although the opposite effect is also conceivable (through a rising b causing a sufficiently large increase in the marginal utility of consumption in the second period). Either way round the movements of C_1 lead to a disincentive constraint which has to be allowed for in policy formulation. Assuming utilitarian social preferences, the optimal tax problem is as below:

$$L = \sum_{j=w,p} \left[\alpha_j\left(U(C_1^{*j}) + \frac{U}{1+\rho}((1+r)(1-b^j)(K^j-C_1^{*j})\right)\right]$$

$$+ \mu\left[- R + (1+r) \sum_{j=w,p} \alpha^j b^j (K^j - C_1^{*j})\right]$$

$$\frac{\partial L}{\partial b^j} = - \alpha_j(1+r)(K^j-C_1^{*j}) \frac{U'(C_2^j)}{1+\rho} + \mu\alpha_j(1+r)(K^j-C_1^{j*})$$

$$- \mu(1+r)\alpha^j b^j \frac{\partial C_1^{*j}}{\partial b^j} = 0$$

$$\Rightarrow \mu = \frac{U'(C_2^j)}{(1+\rho)\left[1-\frac{b^j}{(K^j-C_1^{*j})} \cdot \frac{\partial C_1^{*j}}{\partial b^j}\right]} \qquad j=w,p$$

Without the adjustment effects on C^*, the optimum simply equates the MSU of government revenue, μ, with the discounted second date marginal utilities of consumption for both individuals; second date consumption is equalised if preferences are identical. The presence of a non-zero $\partial C_1^*/\partial b$ clearly disrupts the preferred social optimum. The most probable outcome is that $\partial C_1^{*j}/\partial b > 0$, j=w,p, and that the wealthy are paying positive taxes, so that $b^W > 0$, while the poor are either being subsidised, such that $b^P < 0$, or are paying a substantially lower tax rate than the wealthy. This yields the relations $U'(C_2^W) < U'(C_2^p)$ and $C_2^W > C_2^p$, implying that the wealthy are able to prevent an equalisation of date two wealth by transferring their consumption to the first period. On the other hand, if the reversed saving disincentive $\partial C_1^{*W}/\partial b < 0$ holds true, then the wealthy can possibly have lower second date consumption than the poor. Regardless of the precise form of the policy optimum, it is apparent that the desired extent of redistributio is not achieved, despite the full observability of all relevant characteristics.

The only disincentive allowed in this case has been the adjustment of first period consumption, or, in other words, the individual saving decision. Indeed, the two period model is sometimes used as a simple version of the life-cycle savings framework of the previous two sections. Although both are concerned with savings disincentives, however, the approach taken here is significantly different from the previous sections. In the format of Section (3), the government can tax wealth in all periods, achieving a first best by equalising initial wealth. The crucial feature of the present model is the existence of the pre-tax period, which allows individual adjustments to take place, and not the presence of individual saving. Anticipation effects can take alternative forms, and these may in practice be more effective (see below). Policy anticipation in a sense revives the saving disincentives found lacking in Section (3), but the underlying behavioural response is really a tax avoidance measure rather than a true saving decision.

Any model representing anticipation of policy must essentially have a two-period structure, with the date of tax implementation dividing the model into pre- and post-tax periods. One feature neglected in the above two-period model is the fact that the post-tax period is likely to be longer in duration than the pre-tax period for most individuals. This would give greater weighting to the post-tax period (albeit reduced by any future utility discounting), and would lessen the value of consuming a large part of personal wealth before the taxes are imposed. The government's position may therefore be stronger in practice than in the model considered here.

A higher degree of tax avoidance is attainable if strategic bargaining behaviour takes place (Aumann and Kurz (1977)), or if wealth can be converted into untaxed forms. These two possibilities are considered below.

Concerted Tax Avoidance : a Bargaining Model

The crux of wealth redistribution is the treatment of the highly unequal wealth ownership found at the top end of the scale, and those taxes imposed in practice have generally been limited to individuals above a certain level of wealth. The issues at stake are largely between the government and the very wealthy, and the latter have much greater incentives to avoid taxes than the average person. In the model considered below, anticipation applies only to a relatively small subset of wealthy individuals, and the remainder of the population pay taxes passively without any behavioural adjustment. This is not an especially implausible picture of reality, considering that the wealthy generally have greater access both to information and to the facilities for tax avoidance.

Let the model take a similar form to that above, with two groups in the population and early consumption as the sole means of avoiding the second date taxation; other forms of tax avoidance could easily be included, but would not change the fundamental nature of the outcome. Policy anticipation applies only to the wealthy group, who form a small and homogeneous part of the population. If this group can be unified in its response to redistributive taxes, then it can attain a better position than it would by responding individually to taxation policy. It is assumed here that such harmonious action occurs, whether or not by conscious agreement. The wealthy are thus confronting the government in a game theoretical situation, and strategies are defined by date one consumption and the wealth tax rate respectively. Because the poor do not anticipate the tax (or do not act on their anticipation), their behaviour is predictable to the other agents, and plays only an incidental role in the game. The game structure can be described by

	Strategy	Payoff
Government	b	$V(U_1^W, U_2^W, U_1^P, U_2^P)$
Wealthy	C_1^W	$U_1^W + \frac{U_2^W}{1+\rho}$

where b is the tax rate on the wealthy in the second period. Choice of b and C_1^W automatically determines C_2^W, b^P, C_1^P and C_2^P, via the government revenue constraint, the budget constraint of the wealthy and the known saving behaviour of the poor. Within this framework it is in the common interest of the government and the wealthy to find a co-operative solution through bargaining, rather than to accept the non-co-operative position. This does not necessarily mean that organised bargaining takes place, and co-operation can take the form of a tacit understanding of the responses of each side to the strategies pursued by the other. A Nash bargaining solution is obtainable in the usual way, given a pair of optimal threats and a set of feasible payoff pairs. For the wealthy the most potent threat is that of consuming all their wealth at date one, so that $C_1^W = K^W$; for the government it is to set b to unity. The government's threat presents a difficulty in that it may not always be a rational course of action in the event of a failure in bargaining. In social welfare terms it may be preferable to set $b < 1$ in such a way as to equalise second date incomes (under separability of social welfare over time), implying that in a 'perfect' equilibrium the government is restricted to a weaker threat than $b = 1$. The set of feasible payoff pairs is defined by the assumed payoff functions, combined with the bounds imposed on the strategic variables. It is compact but not necessarily convex, meaning that the existence of a solution can be guaranteed, but not uniqueness.

To illustrate the nature of the bargaining outcome, an algebraic example is worked out below. Preferences are chosen for ease of calculation, rather than plausibility.

Example

Let N = size of population

αN = number of 'wealthy' individuals

$(1-\alpha)N$ = number of 'poor' individuals

K^W, K^P = initial capital holdings of wealthy and poor individuals

Utility at each date is synonymous with consumption, so that total utility is

$$U^{j*} = C_1^j + \frac{C_2^j}{1+\rho} \qquad j=W,P$$

Social preferences are maximin for the intertemporal sum of utilities, that is

$$V = \min_{W,P}\left(C_1^j + \frac{C_2^j}{1+\rho}\right)$$

This is more egalitarian than utilitarianism, and avoids any irrationality of the government's threat of b=1.

The linearity of individual preferences means that all consumption is concentrated at the first or second date according to whether ρ is greater than or less than r. In the former case second date taxation is clearly powerless, so it is assumed that $r > \rho$. Individuals would ideally like to consume all their wealth at date two, and this is always true of the poor group (who do not anticipate taxes, and in any case benefit from them). The wealthy, however, face the threat of a punitive tax rate at date two and may choose to consume part of their wealth at date one.

The payoff functions for the wealthy and the government are such that

$$\pi^W = C_1^W + (K^W - C_1^W)\left(\frac{1+r}{1+\rho}\right)(1-b)$$

$$\pi^g = \frac{C_2^P}{1+\rho} = \left[K^P + \left(\frac{\alpha}{1-\alpha}\right)(K^W - C_1^W)b\right]\left(\frac{1+r}{1+\rho}\right)$$

where the government's revenue requirement is assumed to be zero.

The optimal threats for the wealthy and the government are $C_1^W = K^W$ and b=1 respectively, producing payoffs $\tilde{\pi}^W = K^W$, $\tilde{\pi}^g = K^P\left(\frac{1+r}{1+\rho}\right)$

The Nash bargaining solution is obtained by maximising

$$(\pi^W - K^W)(\pi^g - K^P(1+r/1+\rho))$$

subject to the above equations for π^W and π^g, and to the conditions $0 \leqslant C_1^W \leqslant K^W$, $0 \leqslant b \leqslant 1$. Substituting for π^W and π^g, the problem involves setting C_1^W and b to maximise

$$(K^W - C_1^W)^2\ (r-\rho-b(1+r))(1+r)b = (K^W - C_1^W)^2[(r-\rho)(1+r)b-(1+r)^2b^2]$$

Since the expression must be positive, it is immediately clear that $C_1^W = 0$. b is determined so as to maximise the square-bracketed expression, and setting the first derivative to zero yields

$$(r-\rho)(1+r)-2(1+r)^2b=0 \Rightarrow b = \frac{r-\rho}{2(1+r)}$$

The tax rate obtained thus rises with the interest rate and falls with the utility discount rate.

The various possible outcomes of this model can be summarised as below:

	b	C_1^W	Π^g	Π^W
No intervention	0	0	$\left(\frac{1+r}{1+\rho}\right) K^P$	$\left(\frac{1+r}{1+\rho}\right) K^W$
Non-co-operative solution	1	K^W	$\left(\frac{1+r}{1+\rho}\right) K^P$	K^W
Nash bargaining solution	$\frac{r-\rho}{2(1+r)}$	0	$\left(\frac{1+r}{1+\rho}\right)K^P + \frac{1}{2}\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{r-\rho}{1+\rho}\right)K^W$	$K^W + \frac{1}{2}\left(\frac{r-\rho}{1+\rho}\right)K^W$
Asymmetric soln: govt. dominant	$\frac{r-\rho}{1+r}$	0	$\left(\frac{1+r}{1+\rho}\right)K^P + \left(\frac{\alpha}{1-\alpha}\right)\left(\frac{r-\rho}{1+\rho}\right)K^W$	K^W
Asymmetric soln: wealthy dominant	0	0	$\left(\frac{1+r}{1+\rho}\right)K^P$	$\left(\frac{1+r}{1+\rho}\right)K^W$

With no policy intervention all individuals consume their wealth at the second date, whereas the non-co-operative solution forces the wealthy to consume at the first date to avoid the tax rate of $b=1$. In terms of consumption/utility, the total difference between these solutions is

$$\alpha N\left(\left(\frac{1+r}{1+\rho}\right) K^W - K^W\right) = \alpha N\left(\frac{r-\rho}{1+\rho}\right) K^W$$

which can be viewed as the aggregate utility loss from non-co-operative government intervention. These utility units can be recovered through co-operative behaviour, involving an agreement to defer all consumption to date two. At the Nash bargaining solution the extra utility is divided equally between the government and the wealthy, who then distribute it evenly between the poor and themselves respectively.

It is also possible that unequal bargaining skills lead to an unequal distribution of the co-operative surplus between the government and the wealthy. The two extreme cases correspond to the asymmetric solutions given above. Where the government is dominant, it can act as a 'leader' in the Stackelberg sense, deciding its own policy in the knowledge of the response of the wealthy. Taxation can therefore be increased to the level $b = {}^{r-\rho}/_{1+r}$, at which the wealthy are indifferent between consuming at the first and second dates, and the utility surplus goes to the government. This solution is equivalent to the optimum in the previous model, where there is no unity of response between individual taxpayers. At the opposite extreme the government is frightened off imposing any taxation by the threats of the wealthy to consume all their capital in the pre-tax period. The outcome then coincides with the no intervention position, and the entire utility surplus is appropriated by the wealthy. In a more general bargaining sense one can say that the tax rate must fall somewhere in the interval $0 \leq b \leq {}^{r-\rho}/_{1+r}$, less than or equal to the rate obtained when taxpayers exert no bargaining power.

The example above is not intended to be realistic, but it does illustrate the ability of the wealthy to secure a lower tax rate if they can threaten a unified response to redistributive taxation - in the extreme case they may be able to deter the introduction of any taxation. The government's position in more complex models would be stronger if preferences displayed diminishing marginal utility, so that the threats made by the wealthy represented a higher utility loss. The same would also be true if the second period had been given a greater weighting than the first. On the other hand alternative means of tax avoidance benefit the wealthy, and some of these are considered below.

Alternative Forms of Tax Avoidance

Tax avoidance in practice is often accomplished by transferring wealth into assets which are untaxed or face a lower rate of taxation than the standard definition of wealth. Unless the tax basis is truly comprehensive it is always possible to make a response of this kind, and wealth may also be transferable abroad to tax havens with little or no taxation. Whether these loopholes can be closed is largely an administrative issue, but if any are left open they will lead to a further constraint on the ability to redistribute wealth.

Comprehensive tax avoidance occurs if wealth can be moved costlessly outside the tax system, earning the same interest return and allowing consumption to continue as before; such conditions render redistributive taxation completely ineffective and discourage the implementation of any tax measures. In reality tax avoidance almost certainly involves some costs, either in a lump-sum form or as a lower interest return to transferred wealth. In the two period model above, this can be depicted by assuming that there are two assets available for holding wealth, one of which is taxed, bearing a net return of $(1+r_1)(1-b)$, and the other untaxed, with a return $1+r_2$, $r_1 > r_2$. Individuals always hold their wealth in the form which maximises the income return, and pay the tax only if $(1+r_1)(1-b) \geq 1+r_2$. Hence the maximum tax rate imposed is such that

$$(1+r_1)(1-b) = 1+r_2 \Rightarrow b = \frac{r_1-r_2}{1+r_1}$$

since any tax above this level provokes a move into the untaxed asset. The policy optimum in general will involve a wealth limit above which the maximum rate is imposed, and a set of lower tax rates (in some cases negative) for those below the limit. No wealth is ever held in the untaxed asset at the optimum, although more complex models could yield optima in which this is true (for example, where individuals face differing costs of transferring wealth between assets). Consumption and saving behaviour follow a similar pattern to the previous model, and pre-tax consumption continues to be an alternative form of avoidance. The outcome in social welfare terms must be worse than in the single asset case, and the government has an incentive to try to include the second asset in the tax base. It will choose not to do so only if under a direct constraint (as, for example, when wealth is held abroad) or if the administrative costs outweigh the benefits.

As before the wealthy may be able to exert their collective bargaining power. The ability to transfer assets into untaxed forms provides them with an alternative threat to pre-tax consumption, and leaves them at least as well off as in the case considered above. The government may be able to counteract wealth transfers by threatening an extension of the tax base, although the effectiveness of this depends on the costs involved. Analysis is otherwise similar to the previous case, with the same possible range of outcomes.

In conclusion, the presence of policy anticipation provides a genuine constraint on wealth redistribution, even when wealth is fully observed. This contrasts with the steady-state policies of Section (3), where a single date first-best wealth redistribution was generally achievable. The importance of anticipation depends on a wide variety of different features, even in the simplified models used here. Anticipation is liable to exact a greater welfare cost :

(i) the greater is the initial wealth inequality

(ii) the faster marginal utility declines with consumption, that is, the lower is $U''(C)$

(iii) the longer is the pre-tax period relative to the post-tax period

(iv) the lower is the interest return to capital

(v) the higher is the discount rate on utility

(vi) the lower is the fixed or interest cost of converting wealth into untaxed forms

(vii) the greater is the extent of concerted tax avoidance by taxpayers

(viii) the more egalitarian are social preferences.

This list is not comprehensive, and in practice other issues are also relevant. The importance of disincentives depends on the particular case in question, but it does seem that in general some constraint on redistributive policy will be present. Whatever the other circumstances, it is always true that anticipated taxes cannot directly confiscate wealth, in that wealth holders can choose to avoid all tax payment by consuming their wealth during the pre-tax period. Some tax avoidance is almost certain to take place, and the tax's efficacy depends largely on the inconvenience to individuals of taking these measures. The best way to redistribute wealth is to avoid anticipation by surprising the taxpayers, a feat which may be difficult to achieve in practice.

(5) Conclusion

This chapter has considered redistributive wealth taxation in the context of individuals undertaking life-cycle saving. As was pointed out in the introduction, many other factors are relevant, and the design of wealth taxes involves major administrative, institutional and political aspects. The present discussion is limited to the theoretical feasibility of redistributing wealth, in terms of the ability to overcome individual disincentives, and may consequently overestimate the scope for redistribution in reality. Within this restricted frame of reference, three main issues emerge.

Firstly, for wealth redistribution to be efficient it is preferable to base taxes on an individual's full lifetime wealth, rather than the current wealth holding. Within a life-cycle framework this means including information on the individual's age, as well as observed wealth, allowing an accurate indication of the person's true wealth to be obtained. When taxes are imposed on observed current wealth alone, the analysis becomes complex in form, and, as was noted in Section (2), the resulting tax optimum may not correspond to the usual notion of redistributive taxation, involving negative marginal tax rates at some points. Saving disincentives always arise in taxing current wealth, and the extent of possible redistribution is consequently limited. Since information on age is generally available, there is little cause to formulate taxes in this way, although wealth tax proposals in practice do not always make the distinction between current and true wealth.

Secondly, when taxes are based on true wealth and are introduced without anticipation it is possible to avoid any individual saving disincentives. The models of Section (3) generally permit a first-best outcome, either by lump-sum wealth transfers (decreasing or constant returns to wealth) or by a centralised wealth holding (increasing returns to wealth). Saving disincentives need not be a significant problem, as long as policy measures have the appropriate informational basis.

The third issue is the role of tax anticipation, considered in Section (4). When the introduction of taxes is known beforehand, individuals have the chance to avoid payment by pre-tax consumption or transfer into untaxed assets. To maximise social welfare the government must include these responses as constraints in its policy formulation, and it may find itself in a bargaining situation if wealthy taxpayers can make a concerted response. The first best under these circumstances is not attainable, even if the government has full information on individual characteristics. Anticipation effects therefore pose the major obstacle to redistributing wealth in the theoretical approach used here.

In view of the comments above, it seems that the ideal redistributive wealth tax is an unanticipated single date reorganisation of lifetime assets, operating on the same lines as a capital levy. In reality, despite the social benefits to be gained from secrecy, it is likely that any redistributive tax measure will be known beforehand, allowing individuals to make suitable tax avoidance arrangements. These are the only real disincentives encountered in the models above, and must be regarded as the main theoretical barrier to redistributing personal wealth.

CHAPTER 5 : OPTIMAL UNEMPLOYMENT BENEFITS

(1) Introduction

For the majority of people, wages received from employment are by far the largest component of their total earnings in any given period. The loss of this income through unemployment consequently has a serious effect on the welfare of the individuals concerned, and presents society with the problem of whether or not to take palliative action. Available policy responses can broadly be classified as preventive and curative. Preventive measures fall in the sphere of macroeconomics, and require intervention in the economy to prevent the occurrence of unemployment. As in medicine, prevention is liable to be better than cure, and ideally the successful implementation of macroeconomic measures would greatly reduce unemployment as a problem. Unfortunately, total prevention has not in practice been accomplished, and recent years have seen a return of unemployment on a scale approaching that of the 1930's. In these circumstances there is an evident need for curative measures, of which unemployment benefits are the central element. The assumption made in the following discussion is that the employment situation is exogenously fixed, confining attention to the microeconomic aspects of the problem. This is not intended to imply that macroeconomic policies are ineffective, and, indeed, macroeconomic measures ought to be a high priority of any rational government.

Unemployment benefit schemes vary widely between countries, and can be quite complex in their structure. In Britain, for example, there has always been a distinction between unemployment benefit per se and means tested supplementary benefit, as well as various other payments related to an individual's particular circumstances. The analysis below does not attempt to model any particular benefit system, and assumes that a uniform payments schedule applies to all individuals. Similarly, the models also abstract from the intricacies of government finance, mostly assuming a uniform lump-sum tax levied on the employed (although this assumption can easily be relaxed). The aim is to concentrate on certain broad aspects of all benefit schemes, such as the level of benefit payments relative to employment earnings (the 'replacement ratio') and the optimal time pattern of benefit payments.

There are two main alternative ways of viewing unemployment benefits, as either insurance or a redistributive tool:

(a) Unemployment benefit as insurance

In this view an individual's career is seen as an intertemporal sequence of alternating employment and unemployment spells. Unemployment is a random event which cannot be foreseen with complete accuracy, but can be assigned a probability in a similar way to other events covered by insurance schemes. 'Actuarially fair' insurance implies that the expected premium payments for a given time period are equal to the expected insurance claims in that period, so that neither the insurer nor the insured makes a net return from the scheme. In theory this would be the outcome of privately supplied insurance provision under competitive conditions, with each insurer making zero profits. The first state insurance schemes sought to emulate private insurance, and thus were also based on notions of 'actuarial fairness'. For example, the initial U.K. scheme in 1911 was careful to avoid any suggestion that the unemployment-prone were to be subsidised, and emphasised the need for the scheme to be self-financing. These ideas were effectively abandoned in the inter-war period, when it became clear that the growing numbers of unemployed could never finance their own claims (although the terminology of self-financing still survives in the present system, with the distinction between 'unemployment benefit' and 'supplementary benefit'). It is difficult to imagine a practicable benefit scheme run solely on insurance principles, and, in any case, the desirability of such an approach is unclear when seen in social welfare terms.

(b) Unemployment benefit as redistribution

State unemployment insurance schemes differ from (hypothetical) private ones in that they can compel individuals to participate, even when their premium payments exceed their expected claims. This means that there is no need for schemes to break even over each individual's career separately, and that a single scheme can be implemented, with receipts and claims balancing over a cross-section of the population at any given time. Such a scheme introduces the possibility of redistribution between individuals, and a government basing its policy on optimality considerations would generally wish for some redistribution to take place. The outcome is not insurance in the usual sense applied to individuals, and ought to be regarded either as a redistributive policy based on unemployment-proneness or as a form of collective insurance for society as a whole.

It is sometimes argued that unemployment benefits are not designed for redistributive purposes, and that redistribution should be limited to specific ad hoc measures (this point is made, for example, in Disney (1980)). Viewed theoretically, unemployment benefits are lump-sum payments and therefore have no drawbacks on efficiency grounds - the only potential problem is an informational one, concerning whether or not unemployment benefits concentrate on one particular dimension of inequality to the exclusion of others. In the following analysis informational difficulties are ruled out by restricting observable individual differences to employment status. Although informational questions are certainly relevant to unemployment benefits in their redistributive role, they are more pertinent to the subject matter of Chapter 2 than to the current chapter.

Throughout the rest of this chapter, optimal unemployment benefits are derived on the basis of constrained maximisation of a social welfare function. Inevitably the procedure raises redistributive questions, and implies a rejection of the pure insurance view of unemployment benefits

The basic theoretical model is given in Section (2), and is in terms of conventional utility theory. Later sections elaborate on the basic model by introducing additional relevant issues.

(2) Basic Theoretical Model

The model described here is the simplest possible timeless case, based on a given cross-section of employed and unemployed individuals. Unemployment is taken to be exogenous and uninfluenced by government policy.

Let the working population comprise H individuals, with identical preferences known by the government. Of these, suppose that E are in employment and H-E are unemployed. All individuals have T hours of available working time, are paid an hourly wage, W, and receive no non-employment income.

The Government pays a lump-sum benefit, B, to the unemployed, financed by a uniform lump-sum payment, b, charged to those in employment. If R is the revenue requirement, then the government's revenue constraint is $R + (H-E)B = Eb$.

Individual preferences are represented by the standard neoclassical theory of labour supply, which treats leisure as an additional commodity in the utility function (alternative approaches are considered below). With a free choice of working hours, the utility of the employed is given by the indirect utility functions

$$U_i = U(\underline{p}, W, WT-b)$$

where $\underline{p}$ is the vector of consumer prices. (If desired, a rationed function $\tilde{U}(\underline{p}, WT-b, \tilde{T})$ could be used, where $\tilde{T} < T$ is the number of leisure hours to which the individual is restricted. This would not significantly affect the nature of the model). The unemployed are constrained to undertake no employment, and therefore must consume the full T hours per week as leisure. Applying standard rationing theory (as in Neary and Roberts (1980)), their utilities are expressible as

$$\tilde{U}_j = \tilde{U}(\underline{p}, B, T)$$

where $\tilde{U}$ is the rationed indirect utility function corresponding to the preferences in U.

An important consideration is the possibility of a disincentive effect, whereby the employed can declare themselves unemployed and claim benefits if it is in their interest to do so. This depends on the government's ability to distinguish between the 'bona fide' unemployed and those who are unemployed by choice - if the distinction is observable, then such behaviour cannot occur. In reality there seems to be no foolproof way to identify the voluntarily unemployed, and it is desirable to allow for this in the model. The problem is akin to that of moral hazard in insurance markets, where the insured party can influence events without the knowledge of the insurer. To eliminate moral hazard at the optimum, it is necessary to assume that all employed persons have a utility at least as great as that of the unemployed. This can be guaranteed in the present case by imposing the constraint $U \geqslant \tilde{U}$.

Social preferences are individualistic, described by the social welfare function $V = V(U_1,\ldots,U_H)$, which becomes $V^{\dagger} = \sum_h V(U_h)$ in the special case of utilitarianism (where $\partial V/\partial U > 0$ and $\partial^2 V/\partial U^2 < 0$ will generally hold true). Under the present assumption of a uniform employed population these functions are expressible as $V = V(U, \tilde{U})$ in the general case and $V^{\dagger} = EV(U) + (H-E)V(\tilde{U})$ under utilitarianism. The government's optimal policy problem is to choose B and b to maximise $V^{\dagger}$ subject to the revenue constraint and the moral hazard condition.

The appropriate Lagrangian for utilitarian preferences is

$$L = EV(U) + (H-E)V(\tilde{U}) + \phi(Eb-R-(H-E)B) + \psi(U-\tilde{U})$$

with optimality conditions

$$\frac{\partial L}{\partial B} = (H-E)\frac{\partial V}{\partial \tilde{U}} \cdot \frac{\partial \tilde{U}}{\partial B} - \phi\,(H-E) - \psi\frac{\partial \tilde{U}}{\partial B} \leqslant 0$$

$$B\left((H-E)\frac{\partial V}{\partial \tilde{U}} \cdot \frac{\partial \tilde{U}}{\partial B} - \phi\,(H-E) - \psi\frac{\partial \tilde{U}}{\partial B}\right) = 0$$

$$\frac{\partial L}{\partial b} = E\frac{\partial V}{\partial U} \cdot \frac{\partial U}{\partial b} + \phi E + \psi\frac{\partial U}{\partial b} \leqslant 0$$

$$b\left(E\frac{\partial V}{\partial U} \cdot \frac{\partial U}{\partial b} + \phi E + \psi\frac{\partial U}{\partial b}\right) = 0$$

$$\frac{\partial L}{\partial \psi} = U-\tilde{U} \geqslant 0 \qquad\qquad \psi(U-\tilde{U}) = 0$$

Firstly, note that if the moral hazard constraint is not imposed, the optimum requires that

$$\phi = \frac{\partial V}{\partial \tilde{U}} \cdot \frac{\partial \tilde{U}}{\partial B} = -\frac{\partial V}{\partial U} \cdot \frac{\partial U}{\partial b}$$

This is simply a first-best redistribution of income between the employed and unemployed, equating the mean MSU of income for both groups. Such an outcome would arise if the government could observe voluntary unemployment.

For most situations, however, the presence of a moral hazard constraint will significantly alter this outcome. In particular, when leisure is a normal good the optimum must occur at a corner solution satisfying $U=\tilde{U}$. To show this, suppose that at the optimum $U>\tilde{U}$. Then from the optimality condition, $\psi=0$. It is known that B, $b\neq 0$, since setting them to zero gives $-{}^{\partial V}/_{\partial U} \cdot {}^{\partial U}/_{\partial b} > {}^{\partial V}/_{\partial \tilde{U}} \cdot {}^{\partial \tilde{U}}/_{\partial B}$, which is a contradiction (given that $b=B=0$ and $U>\tilde{U}$). Hence, $\phi = {}^{\partial V}/_{\partial \tilde{U}} \cdot {}^{\partial \tilde{U}}/_{\partial B} = -{}^{\partial V}/_{\partial U} \cdot {}^{\partial U}/_{\partial b}$, as at the first best. From standard rationing theory, $\tilde{U}(\underline{p}, B, T) = U(\underline{p}, \tilde{W}, B+\tilde{W}T)$ where $\tilde{W}$ is the 'virtual wage', at which individuals would voluntarily choose T hours of leisure (so that $\tilde{W} < W$). The relation $\partial\tilde{U}/_{\partial B}(\underline{p}, B, T) = \partial U/_{\partial B}(\underline{p}, \tilde{W}, B+\tilde{W}T)$ also holds true. Thus, in the present case

$$\frac{-\partial V}{\partial U} \cdot \frac{\partial U}{\partial b}(\underline{p}, W, WT-b) = \frac{\partial V}{\partial \tilde{U}} \cdot \frac{\partial \tilde{U}}{\partial B}(\underline{p}, B, T) = \frac{\partial V}{\partial U} \cdot \frac{\partial U}{\partial B}(\underline{p}, \tilde{W}, B+\tilde{W}T)$$

and $$-\frac{\partial U}{\partial b}(\underline{p}, W, WT-b) \geqslant \frac{\partial U}{\partial B}(\underline{p}, \tilde{W}, B+\tilde{W}T)$$

given that $U > \tilde{U}$ and $\partial^2 V/_{\partial U^2} \leqslant 0$.

Inverting gives $\frac{\partial C}{\partial U}(\underline{p}, W, U) \leqslant \frac{\partial C}{\partial U}(p, \tilde{W}, \tilde{U})$

From the properties of cost functions, $\partial^2 C/_{\partial U \partial W} = \partial l/_{\partial U}$, where l is the compensated demand for leisure. If leisure is a normal good, it must be true that $\partial l/_{\partial U} > 0$; since $\tilde{W} < W$ and $\partial^2 C/\partial U^2 > 0$ (for diminishing marginal utility of income), the above condition on $\partial C/_{\partial U}$ can then be satisfied only if $U < \tilde{U}$. This contradicts the initial assumption, and the optimum occurs at the corner solution $U = \tilde{U}$.

The converse is also true, so that a corner solution occurs only if leisure is a normal good at the optimum. To show this, suppose that there is a corner solution where $U = \tilde{U}$ and $\partial l/_{\partial U} < 0$ apply simultaneously. From the optimality conditions it is known that $\psi > 0$ and hence $\partial U/_{\partial B} > -\partial U/_{\partial b}$, given that $U = \tilde{U}$. Inverting yields

$$\frac{\partial C}{\partial U}(\underline{p}, W, U) > \frac{\partial C}{\partial U}(\underline{p}, \tilde{W}, \tilde{U}) = \frac{\partial C}{\partial U}(\underline{p}, \tilde{W}, U)$$

for $U = \tilde{U}$.

This is a contradiction, since by assumption $\partial l/_{\partial U} = \partial^2 C/_{\partial U \partial W} < 0$, while it is known that $\tilde{W} < W$. Consequently, one can say that the corner solution $U = \tilde{U}$ occurs iff leisure is a normal good at the solution point (leisure need not necessarily be normal at other income, wage or price levels). The relationship is not guaranteed to hold for social preferences more complex than the utilitarian form used here.

As leisure is usually regarded as a normal good, the main case of interest is the corner solution equating utilities. When $U = \tilde{U}$ does apply at the optimum, the solution reduces to

$$\frac{-\partial U}{\partial b} = \left(\frac{E}{E \frac{\partial V}{\partial U} + \psi} \right) \cdot \phi \text{ and } \frac{\partial \tilde{U}}{\partial B} = \left(\frac{H-E}{(H-E) \frac{\partial V}{\partial \tilde{U}} - \psi} \right) \cdot \phi$$

where ϕ is the MSU of government revenue. Since $\partial V/_{\partial U} = \partial V/_{\partial \tilde{U}}$ (for $U=\tilde{U}$), it is evident that the MSU of income to the unemployed is greater than that to the employed. The government wishes to raise benefits above the optimal level, but is prevented from doing so by the moral hazard constraint. This optimum is similar in nature to the one obtained in the wage taxation model of Dasgupta and Hammond (1980). In their case the disincentive constraint arises from the reporting of individual characteristics; here it depends on the ability of individuals to declare themselves unemployed.

It remains conceivable that, contrary to the usual assumption of labour supply theory, leisure is not a normal good. For example, this might arise where the unemployed have such a surfeit of leisure that it makes a negligible (or even negative) contribution to utility. If inferiority of leisure does happen to hold in the region of the optimum, then it will be a first-best income redistribution, equating the MSU's of income to the employed and unemployed.

The two main features of interest are the benefit level and the effects of changes in the exogenous employment parameters. These are considered below.

Optimal Benefit Level

At a corner solution, total expenditure on goods can be written in terms of compensated demands as

$$W(T-1)-b = \sum_{i=1}^{N} p_i x_i(\underline{p}, W, U)$$

for the employed, and

$$B = \sum_{i=1}^{N} p_i x_i(\underline{p}, \tilde{W}, \tilde{U})$$

for the unemployed, where $U = \tilde{U}$ and $W < \tilde{W}$. It is known that

$$W \frac{\partial l}{\partial W} + \sum_{i=1}^{N} p_i \frac{\partial l}{\partial p_i} = W \frac{\partial l}{\partial W} + \sum_{i=1}^{N} p_i \frac{\partial x_i}{\partial W} = 0$$

by the zero degree homogeneity of the demand for leisure, and symmetry of cross-substitution effects. Hence

$$\sum_{i=1}^{N} p_i \frac{\partial X_i}{\partial W} = - W \frac{\partial l}{\partial W} > 0$$

given the negative substitution effect for leisure. Applying this to the expression above yields

$$W(T-1)-b = \sum_{i=1}^{N} p_i x_i(\underline{p}, W, U) > \sum_{i=1}^{N} p_i x_i(\underline{p}, \tilde{W}, U) = B$$

since $\tilde{W} < W$ and all else remains constant. If the replacement ratio is defined as $^{B}/_{W(T-1)-b}$, then this must be less than unity.

The same also applies when leisure is inferior. In this case an interior solution is obtained, such that $\tilde{U} < U$ and marginal utilities of income are equated. Considering the expressions for W(T-1)-b and B above, the effect of W on them is unchanged, since the substititution term $^{\partial l}/_{\partial W}$ remains negative. The impact of variation in U is such that

$$\frac{\partial C}{\partial U} = W \frac{\partial l}{\partial U} + \sum_{i=1}^{N} p_i \frac{\partial x_i}{\partial U} > 0$$

With inferiority of leisure, however, it is known that $\partial l/_{\partial U} < 0$, and hence

$$\sum_{i=1}^{N} p_i \frac{\partial x_i}{\partial U} > 0$$

Bearing in mind that $\tilde{W} < W$ still holds, and that $\tilde{U} < U$, one can then write

$$W(T-l)-b = \sum_{i=1}^{N} p_i x_i(\underline{p}, W, U) > \sum_{i=1}^{N} p_i x_i(\underline{p}, \tilde{W}, \tilde{U}) = B$$

mirroring the relation that holds for the corner solution case.

It follows that in the present model the optimal replacement ratio is always less than unity. One can also note that it must in general be greater than zero, if the optimality conditions are to hold. Otherwise the optimal value can potentially range anywhere between zero and unity, depending on the functional form of preferences.

Some of the factors influencing the optimal replacement ratio can be illustrated by the example of the CES utility function.

<u>Example</u> Let preferences take the CES form, such that

$$U = [\delta x^{-\theta} + (1-\delta)l^{-\theta}]^{-\frac{1}{\theta}} \qquad 0 < \delta < 1, \qquad \theta \geq -1$$

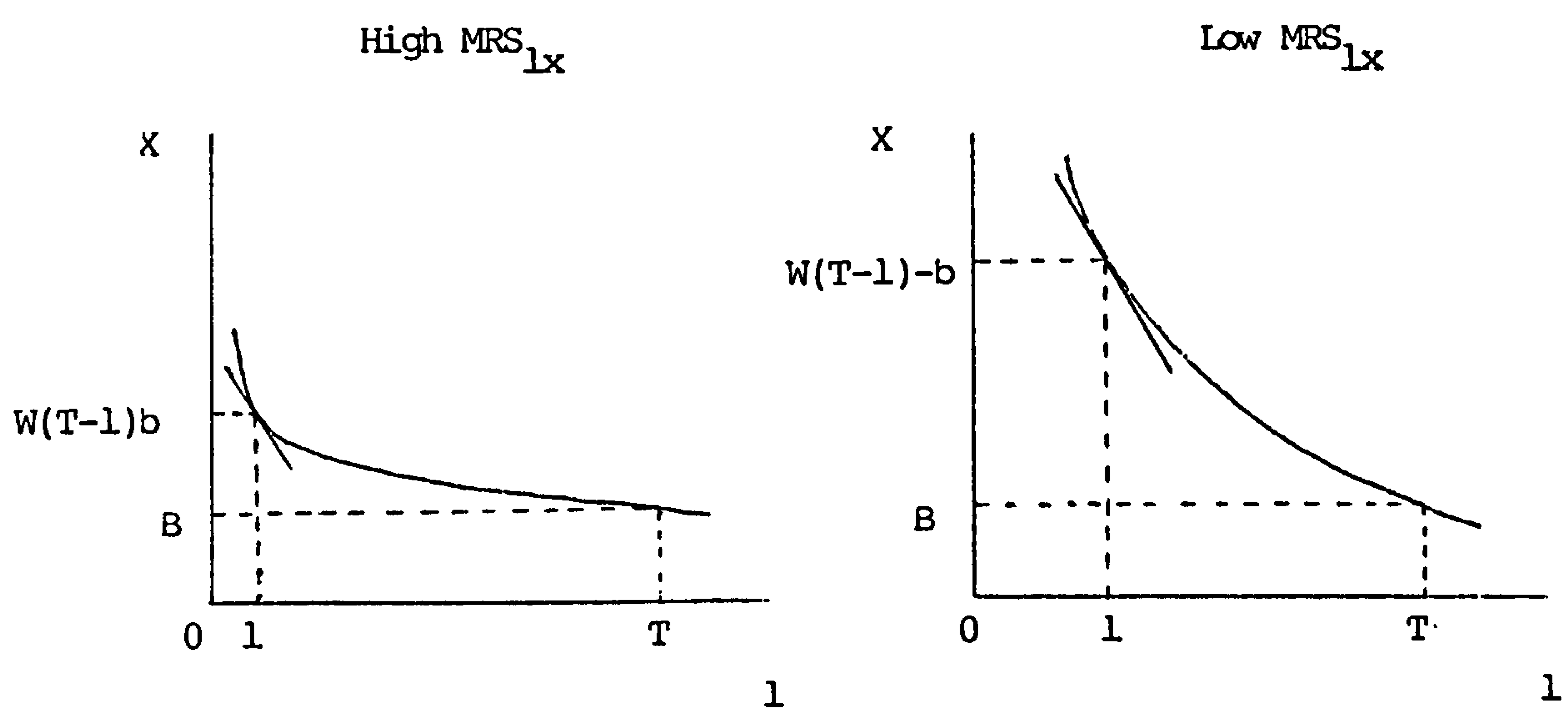
where x is consumption and l is leisure. Since leisure is a normal good, the optimum must always occur at the corner solution equating U and $\tilde{U}$ (as shown above).

Consider now the impact on the optimal replacement ratio of changing the 'weighting' parameter, δ. The demand for leisure by the employed can be written as

$$l = \frac{((1-\delta)p)^{\frac{1}{1+\theta}}(WT-b)}{p(\delta W)^{\frac{1}{1+\theta}} + W((1-\delta)p)^{\frac{1}{1+\theta}}}$$

which falls unambiguously with δ, ceteris paribus. As δ increases, the constraint on the unemployed becomes tighter, and their surplus of leisure over that of the employed increases. For the $U = \tilde{U}$ condition to hold, there must be a compensating increase in the consumption advantage of the employed over the unemployed, implying a fall in the replacement ratio.

Generally, the higher is the MRS of leisure for consumption, the lower is the replacement ratio at a corner solution optimum for a given wage rate. This can be illustrated diagrammatically as below, setting $p = 1$ so that consumption $\equiv$ income:

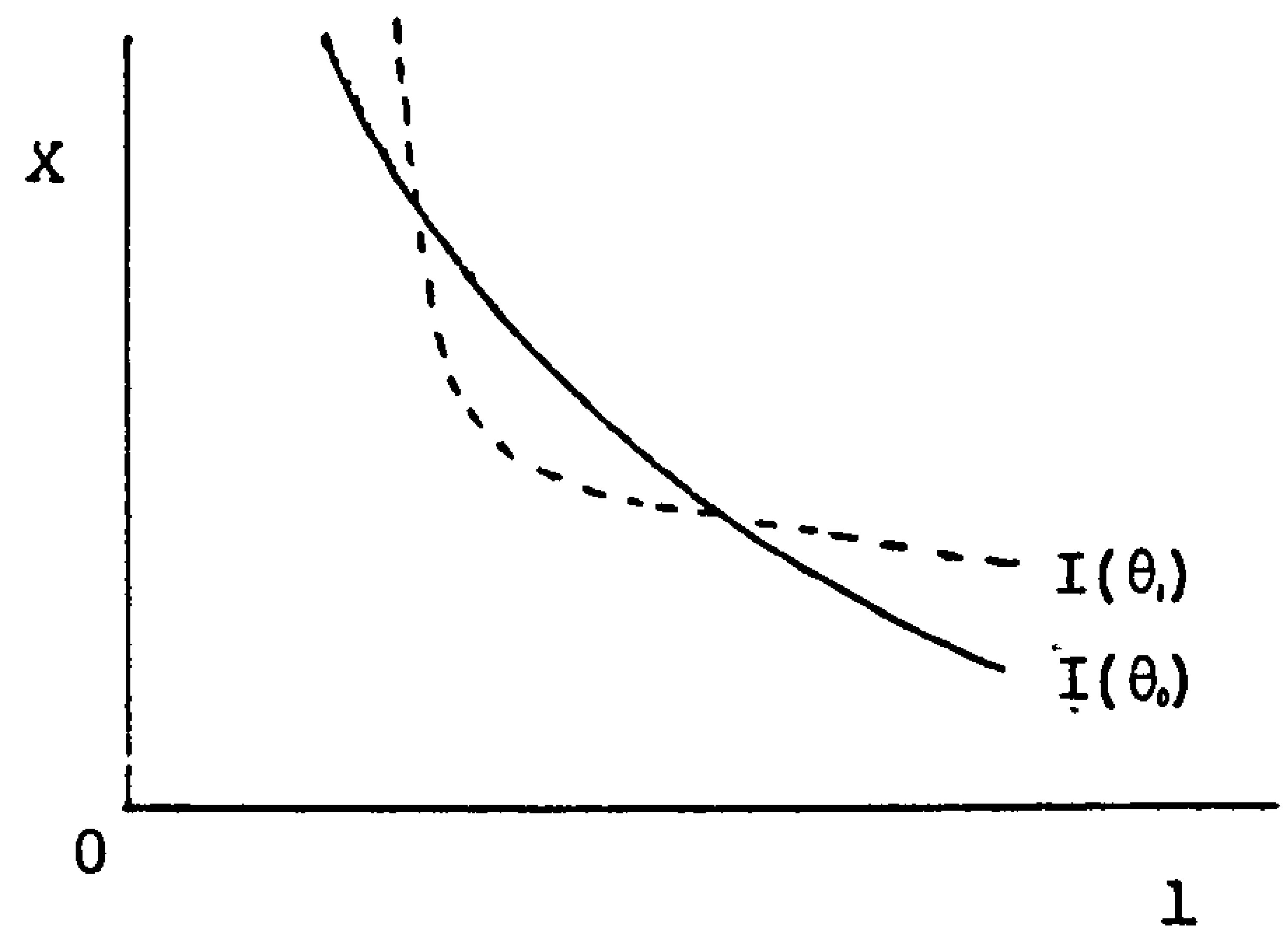


The employed and unemployed are on the same indifference curve, at the tangency with the budget line and the intersection with l=T respectively. For a given W, the income of the employed converges to that of the unemployed as the slope of the indifference curve decreases. A higher marginal valuation of leisure therefore leads to a higher optimal replacement ratio, and less divergence between the leisure consumption of the two groups.

Returning to the CES case, suppose that the substitution parameter, θ, is allowed to vary. The impact on the leisure demand function above is such that

$$\frac{\partial l}{\partial \theta} \gtreqless 0 \quad \text{as} \quad \delta W \gtreqless (1-\delta)p$$

Treating δ as a constant, the impact of θ varies with the size of the real wage, $^W/_p$. If $^W/_p$ is sufficiently high, a rise in θ tightens the constraint on the unemployed and reduces the optimal replacement ratio, and vice versa. This can be explained in terms of the effect of changes in θ on the indifference curve, as below



A higher θ (and lower elasticity of substitution) causes more 'pointed' indifference curves, raising the slope at one extremity and decreasing it at the other. The impact on the MRS_{lx} thus depends on an individual's location on an indifference curve, which is determined by the real wage, $^W/_p$. Where $^W/_p$ is high enough, the employed have high consumption relative to leisure, so that a rise in θ decreases the MRS_{lx} and increases the replacement ratio. Conversely, a sufficiently low $^W/_p$ implies a negative relation between θ and the replacement ratio. Hence the impact of the elasticity of substitution is ambiguous, varying with the slope of the budget constaint faced by the employed.

In general the optimum relies on attitudes to leisure, such that a stronger demand for leisure tends to increase the optimal replacement ratio. Apart from this observation, nothing can be said about the optimum without specifying preferences.

Comparative Static Properties

(a) A larger working population, H, ceteris paribus, results in a lower optimal benefit rate.

To demonstrate this, suppose that H rises from H to H', and denote the resulting changes in the optimal B and b likewise. From the revenue constraint $b = {}^{(H-E)B}/_E + {}^R/_E$, and at a corner solution

$$U(\underline{p}, W, WT - \frac{(H-E)B}{E} - \frac{R}{E}) = \tilde{U}(\underline{p}, B, T) = U(\underline{p}, \tilde{W}, B + \tilde{W}T)$$

Hence $$U(\underline{p}, W, WT - \frac{(H'-E)B}{E} - \frac{R}{E}) < U(\underline{p}, \tilde{W}, B + \tilde{W}T)$$

At the new optimum it must be true that

$$U(\underline{p}, W, WT - \frac{(H'-E)B'}{E} - \frac{R}{E}) = \tilde{U}(\underline{p}, B', T) = U(\underline{p}, \tilde{W}', B' + \tilde{W}'T)$$

and from the properties of the utility function this can apply only if B' < B.

A similar argument shows that b will rise.

For an interior solution the situation is essentially the same, with

$$\frac{-\partial U}{\partial b}(\underline{p}, W, WT - \frac{(H'-E)}{E}B - \frac{R}{E}) > \frac{\partial U}{\partial B}(\underline{p}, \tilde{W}, B + \tilde{W}T)$$

implying that B' < B must hold.

This outcome is intuitively clear, because with E constant a larger H simply means an increase in the number of unemployed who have to be supported from an unchanged employment level.

(b) A higher employment level, E, ceteris paribus, increases the optimal benefit rate.

The situation is identical to that in (a), except that a rise in E serves to lower rather than increase the expression $^{(H-E)B}/_{E} + {}^{R}/_{E}$. By analogous arguments it follows that B must rise and b fall.

This also is intuitive, representing a decrease in the number of unemployed to be supported, coupled with a rise in the number of employed to be taxed.

In reality the parameters H and E are not generally freely variable in the short run, and they are liable to be closely interrelated. This makes the comparative static observations above a somewhat artificial exercise, unlikely to be relevant to real policy decisions.

If variation in wage rates is permitted, then the model is liable to have the property observed by Mirrlees (1971) in the context of optimal taxation, where some individuals are better off not working at the optimum. The utility of the unemployed is then equated with that of a threshold group earning the reservation wage, $W^{\dagger}$; those with (potential) wages below W will choose not to work, while those with wages above $W^{\dagger}$ will work and enjoy utilities in excess of $\tilde{U}$. More complex models with many characteristics would produce an equivalent outcome, yielding a threshold boundary determining whether or not individuals are willing to work at the optimum.

A couple of further issues are relevant:

<u>(i) Alternative Specifications of Preferences</u>

The discussion above is based on utility function of the form $U(\underline{x},l)$, where $\underline{x}$ is consumption of commodities and l is consumption of leisure. Since working time does not appear in U, it is assumed to have no positive or negative effect on an individual's welfare, and the state of being unemployed also has no impact. This can lead to some implausible outcomes. In particular, the positive view of leisure means that, for a given commodity expenditure, unemployment raises utility. It is not obvious that this is always true in reality, and one can argue that the standard format $U=U(\underline{x}, l)$ overestimates the value of leisure to the unemployed. In practice unemployment imposes social and psychological costs beyond the presence of excessive leisure (Sinfield (1981); Seabrook (1982)). Neglect of these factors may give a misleading picture of the true social optimum.

If unemployment has a direct effect on an individual's well-being, then this may be represented by a discontinuity in preferences at the point of unemployment. In these circumstances there are two separate sets of preferences, one for the employed and one for the unemployed - the difference between them may at its simplest be a fixed utility cost attached to unemployment, or alternatively may be a complete change of preferences affecting all consumption decisions. In general one would expect unemployment to have a negative effect on utility (for given income and prices), although its effect on marginal utilities is more debatable.

A different response to the over-valuation of leisure in $\tilde{U}$ is to reformulate the role played by time in the utility function. The most basic adaptation would be to make utility a function of working time, as well as leisure, with $U=U(\underline{x}, l, T-l)$. More generally, the model could be recast in terms of 'household production' theory, as, for example, in Atkinson and Stern (1980). In this framework utility depends on a number of activities, which are in turn produced from inputs of goods. Time must be an input into all activities, and in some cases may be the only input - one possible formulation is to define 'work' as an income-bearing activity with time as its single input, and to regard the time inputs into all other activities as 'leisure'. The resulting model resembles conventional utility theory, except that there are now two constraints, in income and time. The more sophisticated treatment of time provides two possible means to downgrade the value of leisure: firstly, a positive utility valuation of work implies that any net increase in leisure imposes a utility cost from the loss of the working activity; and secondly, a rise in leisure at the expense of work reduces employment income and thus limits the opportunities to devote leisure time to those activities requiring expensive inputs of goods.

In the present context the main concern is the effect of these modifications on the optimal benefit rate and the model's comparative static properties. For the optimal benefit rate the consequences will depend heavily on the marginal utilities of income, $\partial U/\partial M$ and $\partial \tilde{U}/\partial M$. One possibility is that unemployment makes individuals relatively impervious to pleasure, so that, ceteris paribus, their marginal utility of income is less than that of the employed. This would tend to produce interior solutions with low replacement ratios - it is an example of the 'boor versus aesthete' problem, where utilitarianism favours the latter, regardless of the extent of inequality in the population. The effect could be partly or wholly offset if the social welfare function displayed a degree of inequality aversion. Should unemployment have this impact on people (as indeed may well be true), then the case for macroeconomic measures to reduce unemployment would be correspondingly strengthened. On the other hand, by invoking the general principle that low utilities accompany high marginal utilities, it can be argued that unemployment will raise the marginal utility of income. This would encourage corner solutions, and given that unemployment does lower utility, would result in high replacement ratios. Consequently, it seems that the optimal benefit rate will be highly sensitive to the particular specification of preferences, and depends on features about which it is difficult to make any definite a priori assertions. The position with the model's comparative static properties is less uncertain - they should not be significantly influenced by changes in preferences, coinciding with those outlined above.

(ii) Direct Job Creation Expenditures

Although macroeconomic measures have been ruled out by assumption, there is no particular reason why the government cannot create jobs by administrative fiat, financing them from tax revenue. It is of interest to consider the conditions under which this kind of activity is desirable.

Let the theoretical model be unchanged from above, except that employment, E, satisfies the functional relationship $E = E(Q)$, $E' > 0$, where Q is government expenditure on 'job creation'. Q may be interpreted either as the direct monetary outlay in financing new jobs (mainly the wage bill), or as an indirect means of fostering employment, for instance, by improving the institutions assisting job allocation.

When Q is present the revenue constraint becomes $E(Q)b = (H-E(Q))B + R + Q$, and the Lagrangian is

$$L = E(Q)v(U) + (H-E(Q))v(\tilde{U}) + \phi(E(Q)b-(H-E(Q))B-R-Q) + \psi(U-\tilde{U})$$

The optimality conditions for B and b are identical to those above, while the condition for Q implies that

$$\frac{\partial L}{\partial Q} = \frac{\partial E}{\partial Q} \cdot v(U) - \frac{\partial E}{\partial Q} \cdot v(\tilde{U}) + \phi\frac{\partial E}{\partial Q} \cdot b + \phi\frac{\partial E}{\partial Q} \cdot B - \phi \leqslant 0$$

$$Q\left[\frac{\partial E}{\partial Q} \cdot v(U) - \frac{\partial E}{\partial Q} \cdot v(\tilde{U}) + \phi\frac{\partial E}{\partial Q} \cdot b + \phi\frac{\partial E}{\partial Q} \cdot B - \phi\right] = 0$$

Hence either $Q > 0$ and the expression holds with equality, or $Q = 0$ and

$$\frac{\partial E}{\partial Q}\left[v(U) - v(\tilde{U}) + \phi(b+B)\right] - \phi < 0$$

Rearranging implies that optimal expenditure on job creation is zero only if

$$v(U) - v(\tilde{U}) + \phi(b+B) < \frac{\phi}{\partial E/\partial Q}$$

where ϕ is the MSU of government revenue and $^{\partial E}/_{\partial Q}$ is the marginal rate of job creation per unit of expenditure. The left and right hand sides are therefore the marginal social welfare gain and loss from job creation respectively.

When leisure is normal the optimum occurs at a corner solution satisfying $U=\tilde{U}$. In this case the expression above simplifies to

$$b + B < \frac{1}{^{\partial E}/_{\partial Q}} = \frac{\partial Q}{\partial E}$$

where $^{\partial Q}/_{\partial E}$ is the marginal cost of creating one more job. Thus, $Q=0$ only if the sum of the benefit rate and the uniform tax rate is less than the marginal monetary cost of job creation. This is intuitively appealing, since for $U=\tilde{U}$ job creation bears no utility return, and its only benefit is the revenue saving, $b + B$, to be set against the marginal monetary cost, $^{\partial Q}/_{\partial E}$. When leisure is inferior $U > \tilde{U}$ at the optimum and job creation also brings a utility return, as reflected in the full inequality condition above.

A model which ignores job creation rests on the implicit assumption that the conditions given here hold true - otherwise the government would not be acting rationally. It is by no means self-evident that the inequalities are always satisfied, and if $^{\partial Q}/_{\partial E}$ is reasonably close to the cost of net weekly wage payments, it is quite possible that $Q > 0$ will hold at a full policy optimum. Moreover, there are two additional factors which tend to strengthen the case for intervention : firstly, the government in practice is probably able to choose the wage rates associated with new jobs (instead of merely replicating existing jobs, as here), which extends the number of situations where job creation is desirable; and secondly, in a macroeconomic model job creation has expansionary demand effects, which are beneficial when unemployment is cyclical in nature.

Henceforth no further reference is made to direct job creation by the government - it should be noted, however, that the omission is justified only by the satisfaction of inequality conditions equivalent to those above or by an arbitrary restriction placed on policy tools.

In the next two sections the basic model is extended to incorporate job search, and to consider some policy issues related to search behaviour.

(3) Models Including Job Search

Section (2) assumed that the unemployment rate is exogenously fixed, which means that the unemployed cannot influence their chance of finding employment. This is a reasonably accurate description of many real situations, particularly when unemployment is cyclical in character. Nevertheless, at all times there is likely to be a frictional component to unemployment, caused by the inability to reconcile available vacancies with job-seekers. A high degree of 'search' activity by the unemployed may well be able to reduce frictional unemployment and thereby influence the total unemployment rate. In this section job search is included in the model, making the unemployment rate partly endogenous (though the basic employment probabilities are still taken to be exogenous). Generally speaking, the relative importance of both frictional unemployment and job search will tend to be greater at times of high economic activity (and low total unemployment) than in a recession - the emphasis placed on job search in this and the following section does not imply that it always has a major impact on unemployment relative to the total. Empirical work on the question is inconclusive. Examples of studies of U.K. data suggesting large search disincentives are Gujarati (1972), Maki and Spindler (1975) (criticised by Cubbin and Foley (1977) and Sawyer (1979)), Benjamin and Kochin (1979) (see also symposium in 1982 JPE), and Batchelor and Sheriff (1980). Studies of cross-section data have found either less significant disincentives (Nickell (1979a and b)), or no definite relationship (Atkinson et al. (1984)).

Once job search is included, a second type of disincentive arises, in addition to moral hazard. Specifically, a high benefit level raises the value of leisure, imposing a high cost of devoting leisure time to search activities, as well as narrowing the gap between the utility levels of the employed and unemployed. The result is a fall in the net returns to job search, leading to a reduction in the amount of job search and an increase in the rate of unemployment. The optimal benefit rate must balance the desirability of a high benefit rate on social welfare grounds against its adverse effect on (frictional) unemployment. A similar job search framework has been used previously in the optimal unemployment insurance models of Baily (1978) and Flemming (1978), although search is there defined in terms of monetary costs rather than work or leisure time.

Modelling movements between employment and unemployment requires an intertemporal framework, with a sequence of different dates. Suppose therefore that time is discrete and that the population is constant, comprising H identical individuals with an infinite work horizon (that is, no future retirement date). Employment conditions are represented by an exogeneous set of transition probabilities, resembling those in Flemming (1978). At each time period there is a probability μ of an employed individual becoming unemployed in the next period; conversely, an unemployed individual has a chance $\nu(s)$ of re-entering employment in the next period, where s is the number of hours devoted to job search. It follows that if δ_{t-1} is the probability of being unemployed at time t-1, then the probability of being unemployed at time t is given by

$$\delta_t = (1-\nu(s))\,\delta_{t-1} + \mu(1-\delta_{t-1}) = (1-\mu-\nu(s))\,\delta_{t-1} + \mu$$

Owing to the infinite horizon, search times are independent of the time period, and the expected rate of unemployment is defined by the steady-state equation $\delta = (1-\mu-\nu(s))\,\delta + \mu$, so that

$$\delta = \frac{\mu}{\mu+\nu(s)} \quad \text{and} \quad 1-\delta = \frac{\nu(s)}{\mu+\nu(s)}$$

Government policy is the same as in Section (1), with a constant benefit, B, paid to the unemployed, financed by a uniform lump-sum payment, b. The utility of the employed remains unchanged at $U=U(\underline{p}, W, WT-b)$, while that of the unemployed now becomes

$$\tilde{U} = \tilde{U}(\underline{p}, B, T-s)$$

under the assumption that job search conveys neither utility nor disutility, and influences preferences only through the loss of leisure time (this assumption could be dropped at the expense of a more sophisticated representation of time allocation, on the lines of the 'household production' approach mentioned in Section (2)). Individuals discount future utility at a rate ρ per period.

Search behaviour involves choosing s to maximise the discounted sum of expected future utility over the infinite horizon. Let U_t^* and $\tilde{U}_t^*$ denote this quantity for the employed and unemployed respectively at date t. Then U_t^* must satisfy

$$U_t^* = U_t + \frac{\mu}{1+\rho}\tilde{U}_{t+1}^* + \frac{(1-\mu)}{1+\rho}U_{t+1}^*$$

In a steady state the values are invariant over time, so that

$$U^* = \frac{(1+\rho)U + \mu\tilde{U}^*}{\mu+\rho}$$

Analogously, $\tilde{U}_t^*$ satisfies

$$\tilde{U}_t^* = \tilde{U}_t + \frac{\nu}{1+\rho}U_{t+1}^* + \frac{(1-\nu)}{1+\rho}\tilde{U}_{t+1}^*$$

so that, in a steady state

$$\tilde{U}^* = \frac{(1+\rho)\tilde{U} + \nu U^*}{\nu+\rho}$$

Substituting for U^* in the steady-state equation for $\tilde{U}^*$ yields

$$\tilde{U}^* = \frac{(1+\rho)}{\rho(\mu+\nu+\rho)} \; [(\mu+\rho)\; \tilde{U} + \nu U]$$

The optimal search time is determined so as to maximise $\tilde{U}^*$. Setting $\partial\tilde{U}^*/\partial s = 0$ implies an optimal search condition

$$\frac{-\partial\tilde{U}}{\partial s} = \frac{1}{(\mu+\nu+\rho)} \cdot \frac{\partial\nu}{\partial s} \cdot (U-\tilde{U})$$

This is the standard format for optimal search, where the left-hand side is the current cost of undertaking job search, and the right-hand side is the expected future returns from job search.

The optimal policy problem resembles that in Section (2), except for the presence of the search condition as an additional constraint. Social welfare is expressible as some function of the employment probabilities μ, ν, and the utilities U, $\tilde{U}$; for example, utilitarian social preferences lead to the social welfare function

$$V = \frac{\mu}{\mu+\nu}\, \tilde{U} + \frac{\nu}{\mu+\nu}\, U$$

The revenue constraint is defined in terms of expectations, and cannot be guaranteed to balance at any particular date; thus

$$R + H \frac{\mu}{\mu+\nu} B = H \frac{\nu}{\mu+\nu} b$$

where R is the revenue requirement per period. The government sets B and b so as to maximise the social welfare function subject to the search condition and the revenue constraint above. A solution is obtainable by the Lagrangian method, using B, b and s as instruments. Unlike Section (2) it is not necessary to include the moral hazard constraint separately, since it is implicit in the search condition (given $\partial\tilde{U}/\partial s < 0$) wherever this is binding. In cases with high search costs, $-\partial\tilde{U}/\partial s$, and/or low returns to search, $\partial\nu/\partial s$, it may be optimal to have no job search - the model then reduces to the form of Section (2), and can be dealt with in the same way.

The following comments can be made about the optimum:

Optimal Benefit Level

As before, little can be said in general about the optimal benefit level, which depends chiefly on the functional form of preferences. In fact, it is not even possible to retain Section (2)'s observation about the replacement ratio being less than unity. For example, suppose that the optimal value of s is sufficiently large that the leisure consumption, T-s, of the unemployed is less than that of the employed: in that case the virtual wage of the unemployed exceeds the wage rate of the employed ($\tilde{w} > w$). It is known that commodity expenditure, $\sum_{i=1}^{N} p_i x_i$, increases with w, ceteris paribus, as shown in Section (2). One can therefore have

$$B = \sum_{i=1}^{N} p_i x_i(\underline{p}, \tilde{w}, \tilde{U}) > \sum_{i=1}^{N} p_i x_i(\underline{p}, w, U) = w(T-1)-b$$

at the optimum, despite the fact that $U > \tilde{U}$. This means that a replacement ratio exceeding unity cannot be ruled out in the present case, even if it is not particularly likely.

Comparative Statics

Consideration of movements in μ and ν in isolation is not especially interesting, but, as in Section (2), a worsening of employment conditions (a rise in μ or fall in ν) should be associated with a lower benefit level, and vice versa.

It would be straightforward to extend the model above to include more elaborate forms of job search behaviour (as summarised in Lippman and McCall (1976)). For example, allowing for a finite work horizon or variation in wages would not significantly alter the nature of the model; the latter case would give a job search framework similar to that in Pissarides (1983), and derivation of the policy optimum would proceed as in the model above, with the addition of the reservation wage as an extra policy instrument. In reality jobs also vary in their associated employment probabilities, μ and ν, and the view of individual behaviour could be extended to include search over these parameters (if individuals are aware of them; search over job security is included on this assumption in the model of Hey and Mavromaras (1981)). Another extension would be to permit job searchers to have a more specific attitude to job search, beyond the loss of leisure incurred. For instance, a negative view of job search in preferences presumably influences search efforts, although its net effect is ambiguous, since it increases both the current costs of search and the expected returns from securing employment. Adding these various elaborations to the model of individual behaviour might be felt to increase its realism, but they would not yield any greater insights into the nature of the policy optimum, and they are not pursued any further here.

A more central issue is the question of interaction between job searchers, as occurs when search externalities are present.

Search Externalities

Many discussions of search behaviour are based on models of a single individual, and therefore do not consider the interaction between job searchers. A model of optimal unemployment benefits has to be based on a population of many individuals, so the possibility of search externalities inevitably arises. In situations with a given number of job vacancies, it is likely that greater seach efforts by one person reduce the employment probabilities of others, creating a negative externality. Such effects have an impact on the policy optimum, and ought to be included in the theoretical model.

Consider the basic homogeneous population search model described at the beginning of this section. The introduction of search externalities needs only a minor adjustment to the probability ν, to allow for the effects of search by other individuals. One possible situation is where the average probability in the population of finding a job is a fixed constant, $\bar{\nu}$; this would, for example, arise when the total number of vacancies being filled in any period is fixed independently of search behaviour. Individual search efforts can then only succeed at the expense of others, leading to a case of pure 'congestion' in job allocation. Let $\varepsilon(S)$ denote the marginal increase in an individual's employment chances resulting from job search, given that other individuals do not alter their search times (with $\varepsilon'(S) > 0$ and $\varepsilon(0) = 0$). If only the first individual searches, then the resulting employment probabilities are

$$\nu_1 = \bar{\nu} + \varepsilon(S_1) \quad \text{and} \quad \nu_i = \bar{\nu} - \left({}^{\varepsilon(S_1)}/_{m-1}\right) \qquad i \neq 1$$

where m is the number of unemployed. If the first two individuals search, then the probabilities are

$$\nu_1 = \bar{\nu} + \varepsilon(S_1) - \left({}^{\varepsilon(S_2)}/_{m-1}\right)$$

$$\nu_2 = \bar{\nu} + \varepsilon(S_2) - \left({}^{\varepsilon(S_1)}/_{m-1}\right)$$

$$\nu_i = \bar{\nu} - \left(\frac{\varepsilon(S_1) + \varepsilon(S_2)}{m-1}\right) \qquad i \neq 1, 2$$

Hence, in the general case

$$\nu_i = \bar{\nu} + \varepsilon(S_i) - \bar{\varepsilon}_{-i} \qquad \forall_i$$

where $\bar{\varepsilon}_{-i}$ is the mean value of ε for all individuals except i. Individuals therefore benefit from search to the extent that their search time exceeds the average for the rest of the population.

For a homogeneous population it is clear the $\varepsilon(S_i) = \bar{\varepsilon}_{-i}$ will always hold, and that $\nu_i = \bar{\nu}$, $\forall i$. Job search is consequently a self-defeating activity, serving only to diminish the utility level. The unemployed face a game theoretic situation, in which the best possible outcome is a co-operative agreement for nobody to undertake job search. Under Nash assumptions, however, such an agreement is unstable, as all individuals have an incentive to break it; the game is therefore of the 'prisoners' dilemma' variety. At the Nash equilibrium individuals choose s so as to maximise their sum of expected future utilities

$$\tilde{U}^* = \frac{(1+\rho)}{\rho\,(\mu+\bar{\nu}+\varepsilon(S_i)-\bar{\varepsilon}_{-i}+\rho)}\,[(\mu+\rho)\tilde{U} + (\nu+\varepsilon(S_i)-\bar{\varepsilon}_{-i})U]$$

on condition that $\bar{\varepsilon}_{-i}$ is fixed. Setting $\partial\tilde{U}^*/\partial S = 0$ and rearranging yields

$$-\frac{\partial\tilde{U}}{\partial S} = \frac{1}{\mu+\bar{\nu}+\rho}\cdot\frac{\partial\varepsilon}{\partial S}\cdot(U-\tilde{U}) \qquad (\text{given that } \varepsilon(S_i) = \bar{\varepsilon}_{-i}, \quad \forall i)$$

which is equivalent in form to the zero externality case, but with ε substituted for a genuine return to search. The unemployed therefore persist in carrying out job search, despite its complete futility. Behaviour of this nature may seem highly irrational, and yet it is unlikely that a group as large and disparate as the unemployed could ever achieve a co-operative outcome. The Nash equilibrium is liable to be an accurate picture of what would happen in reality.

From the government's point of view, job search in this model is merely a nuisance, imposing a utility cost without having any effect on unemployment. As a result the policy attitude of conventional models is exactly reversed, with the government now wishing to discourage job search. Since search times fall with higher benefits, there is no longer a policy conflict between the need to raise the incomes of the unemployed and to avoid the search disincentive. The model is constrained only by moral hazard issues, so that the optimum for a homogeneous population is the same as it would have been in a model without job search, with $U=\tilde{U}$ and $S=0$ (given normality of leisure; otherwise it may not be optimal to eliminate all search).

Constancy of $\bar{\nu}$, as assumed above, is an extreme case, implying that the general level of search times in the population has no impact on unemployment. In practice search is likely to comprise both a direct effect, which reduces frictional unemployment, and a competitive effect, which merely diminishes the employment chances of other individuals. The general case may be represented by writing the employment probabilities as

$$\nu_i = \nu_i(S_i, \underline{S}_{-i}) \qquad \forall i$$

where $\underline{S}_{-i}$ is the vector of search times of the rest of the unemployed, and the functions ν_i, $\forall i$, satisfy $\partial\nu_i/\partial S_i > 0$ and $\partial\nu_i/\partial S_j < 0$, $j \neq i$. In contrast with the model above, it is now possible that uniform non-zero search by identical individuals raises the general level of employment probabilities and reduces unemployment. Search consequently has a positive direct effect to offset the adverse effects working through search externalities and the impact on utility levels. Because externalities depend on the extent of an individual's search relative to that of other individuals, the function ν could well take a particular form

$$\nu_i = \nu(S_i, S_i/\bar{S}_{-i}) \text{ or } \nu_i = \nu(S_i, S_i-\bar{S}_i) \qquad \forall i$$

where $\bar{S}_{-i}$ is the mean level of search among the other unemployed. As before, job search under Nash assumptions involves setting S_i to maximise U_i^* subject to $\underline{S}_{-i}$ being fixed; the first-order conditions yield a similar search equation

$$-\frac{\partial\tilde{U}}{\partial S_i} = \frac{1}{\mu+\nu(S_i, \underline{S}_{-i})+\rho} \cdot \frac{\partial\nu_i}{\partial S_i}(S_i, \underline{S}_{-i}) \cdot (U-\tilde{U}) \qquad \forall i$$

With identical individuals it is known that $S_i=S_j$, $\forall$ i, j, and so the conditions above reduce to a single condition in S (otherwise it would be a set of simultaneous equations). If a direct employment effect is present, then it is no longer true that undertaking search is necessarily futile, and its net value depends on the size of the direct effects vis-à-vis utility and externality costs. Nevertheless, it remains valid that individuals will generally search longer at the Nash equilibrium than they would have chosen to do had they taken externalities into account. The government's problem is similar to the ones described above, and becomes identical in form to that in the basic model of this section when search is uniform in the population. It may or may not be desirable for the government to discourage job search, depending on the relative importance of its positive and negative aspects.

It is worth mentioning that other types of externality effect are also possible. For example, there may be efficiency returns arising from accurate 'job matching', where productivity is higher if jobs are filled by the most suitable individuals from a heterogeneous population. Diamond (1981) examines this possibility, and argues that high unemployment benefits could be justifiable as a subsidy to job search in order to improve job matching. Such an outcome does not immediately translate to the current framework, because Diamond's model does not include an individual choice of search times (so that the number of employer-unemployed contacts depends only on the unemployment rate). When choice of search times is included, a rise in benefits increases unemployment but simultaneously reduces search times, giving a more ambiguous effect on the number of contacts made. Nevertheless any productivity effects of job matching may potentially create a positive search externality, whereby increased search by one individual is of indirect benefit to others.

A further issue related to job search is the optimal time pattern of benefits. This is considered in the next section.

(4) Optimal Benefits Varying with the Duration of Unemployment

The models considered so far have arbitrarily assumed that benefits should be paid at a constant rate, independently of the length of time that an individual has spent unemployed. This neglects the possibility that benefits varying over time may provide a policy optimum superior to that with a constant benefit level. There are two main situations in which time-varying benefits can improve on constant benefits. The first is in models with job search, where a time-varying rate of payments can be used to influence the amount of search by the unemployed. The second is in models with a heterogeneous population, where it is desirable to reflect changes in the cross-sectional composition of the unemployed over time by appropriate movements in the benefit rate. Each of these two cases is discussed in more detail below.

(i) Optimal Time-varying Benefits in Homogeneous Population Models

The population is here taken to comprise identical individuals, all of whom undertake job search when unemployed. Such a model has been considered previously by Shavell and Weiss (1979), and the initial case below derives their main result that the optimal benefit schedule is uniformly declining (a finding which also emerges indirectly from Sampson (1978)).

Following Shavell and Weiss, it is analytically convenient to restrict attention to the case of a single unemployed person returning to employment; this modification does not significantly affect the qualitative nature of the outcome, and will not be retained in (ii) below. The individual in question is assumed to be unemployed at the starting date, searches for as many periods as necessary to find a job, and then keeps the job for all future time. Benefits are paid out at a varying rate, B_t, whre t denotes the number of previous discrete time periods for which an individual has been unemployed. Notation is in all respects the same as in Section (3), with a subscript t denoting the relevant unemployment duration.

A model in this form can be treated by dynamic programming arguments, with either a finite or infinite work horizon. At each duration

$$\tilde{U}_t^* = \tilde{U}_t(B_t, S_t) + \frac{\nu(S_t)}{1+\rho} U_{t+1}^* + \frac{(1-\nu(S_t))}{1+\rho} \tilde{U}_{t+1}^*$$

where $\tilde{U}_{t+1}^*$ depends on future search decisions. After rearranging the first-order condition $\partial\tilde{U}_t^* / \partial S_t = 0$ implies that

$$-\frac{\partial\tilde{U}_t}{\partial S_t} = \frac{1}{(1+\rho)} \cdot \frac{\partial\nu}{\partial S_t} (U_{t+1}^* - \tilde{U}_{t+1}^*)$$

at the individual's search optimum. The government pays out benefits, B_t, in such a way that the expected sum of benefit payments is equal to a predetermined positive constant, $-R_o$, where $R_o < 0$. Hence

$$-R_o = \sum_{i=o}^{\infty} \left(\frac{\prod_{j=o}^{i} (1-\nu_{j-1})}{(1+\rho)^j} \right) B_i \qquad \text{(where } \nu_j = \nu(S_j) \text{ and } \nu_{-1} \equiv 0)$$

for an infinite horizon. Because only present and future payments are relevant to decisions at any particular time, the revenue constraint experienced at duration t is of the form

$$-R_t = B_t + \sum_{i=t+1}^{\infty} \left(\frac{\prod_{j=t+1}^{i} (1-\nu_{j-1})}{(1+\rho)^{j-t}} \right) B_i$$

where $-R_t$ takes some constant value. The optimal benefit scheme is derived from a sequence of constrained optimisation problems in which the government can choose B_t, B_{t+1} and S_t to maximise U_t^* subject to the individual optimal search condition and the expected revenue constraint. In order to obtain definite conclusions about the optimal benefit schedule, it is necessary to impose separability of $\tilde{U}$ in B and S, so that $\partial\tilde{U}/\partial B$ is independent of S. The Lagrangian and first-order conditions are:

$$L_t = \tilde{U}_t + \frac{\nu(S_t)}{1+\rho} U^*_{t+1} + \frac{(1-\nu(S_t))}{1+\rho} \tilde{U}^*_{t+1}$$
$$+ \phi_t \left(-R_t - B_t - \sum_{i=t+1}^{\infty} \left(\frac{\prod_{j=t+1}^{i}(1-\nu_{j-1})}{(1+\rho)^{j-t}} \right) B_i \right)$$
$$+ \psi_t \left(-\frac{\partial \tilde{U}_t}{\partial S_t} - \frac{1}{(1+\rho)} \cdot \frac{\partial \nu}{\partial S_t} (U^*_{t+1} - \tilde{U}^*_{t+1}) \right)$$

$$\frac{\partial L_t}{\partial B_t} = \frac{\partial \tilde{U}_t}{\partial B_t} - \phi_t = 0$$

$$\frac{\partial L_t}{\partial B_{t+1}} = \frac{(1-\nu_t)}{(1+\rho)} \cdot \frac{\partial \tilde{U}_{t+1}}{\partial B_{t+1}} - \phi_t \frac{(1-\nu_t)}{(1+\rho)} + \frac{\psi_t}{(1+\rho)} \cdot \frac{\partial \nu}{\partial S_t} \cdot \frac{\partial \tilde{U}_{t+1}}{\partial B_{t+1}} = 0$$

$$\frac{\partial L_t}{\partial S_t} = \phi_t \frac{\partial \nu}{\partial S_t} \left(\frac{B_{t+1}}{1+\rho} + \sum_{i=t+2}^{\infty} \left(\frac{\prod_{j=t+2}^{i}(1-\nu_{j-1})}{(1+\rho)^{j-t}} \right) B_i \right) + \psi_t \left(\frac{\partial^2 \tilde{U}_t}{\partial S_t^2} - \frac{\partial^2 \nu}{\partial S_t^2} \frac{(U^*_{t+1} - \tilde{U}^*_{t+1})}{(1+\rho)} \right) = ($$

(using the individual optimal search condition)

From $\partial L_t / \partial B_t = 0$ it must be true that $\phi_t > 0$. Hence, as

$$-\frac{\partial^2 U_t}{\partial S_t^2} - \frac{\partial^2 \nu}{\partial S_t^2} \frac{(\tilde{U}^*_{t+1} - \tilde{U}^*_{t+1})}{(1+\rho)} > 0$$

(from the second-order condition for the individual search problem) it follows from $\partial L_t / \partial S_t = 0$ that $\psi_t < 0$. Rearranging the first two equations yields the expression

$$\frac{\partial \tilde{U}_{t+1} / \partial B_{t+1}}{\partial \tilde{U}_t / \partial B_t} = \frac{1}{1 + \frac{\psi_t}{1-\nu_t} \cdot \frac{\partial \nu}{\partial S_t}} \qquad (*)$$

Since $\psi_t < 0$, $(1-\nu_t) > 0$ and $\partial\nu/_{\partial S_t} > 0$, it is known that

$$\frac{\partial\tilde{U}_{t+1}/\partial B_{t+1}}{\partial\tilde{U}_t/\partial B_t} > 1 \Rightarrow \frac{\partial\tilde{U}_{t+1}}{\partial B_{t+1}} > \frac{\partial\tilde{U}_t}{\partial B_t}$$

Given that $\partial\tilde{U}/_{\partial B}$ is independent of S and shows diminishing returns, it holds true that

$$\frac{\partial\tilde{U}_{t+1}}{\partial B_{t+1}} > \frac{\partial\tilde{U}_t}{\partial B_t} \Rightarrow B_t > B_{t+1}$$

for all values of t, so that the optimal benefit scheme is uniformly declining over time. This is the basic optimality result, arising from the fact that, for a given expected total of benefit payments, a declining benefit scheme elicits a higher level of job search than a constant one.

That benefits asymptotically approach zero can be shown as follows. Suppose that as t approaches infinity, B_t approaches a finite positive limit, l, that is, $\lim_{t\to\infty} B_t = l$, $0 < l < \infty$. Then, because $\nu(s_t) > 0$ for any S_t, the expression

$$\frac{B_{t+1}}{1+\rho} + \sum_{i=t+2}^{\infty}\left(\frac{\prod_{j=t}^{i}(1-\nu_{j-1})}{(1+\rho)^{j-t}}\right) B_i$$

must always exceed zero. Consequently, from $\partial L/_{\partial B_t} = 0$ and $\partial L/_{\partial S_t} = 0$ it is known that $\phi_t > 0$ for any t and that $\lim_{t\to\infty} \psi_t < 0$. Using the expression (*) above, this means that

$$\lim_{t\to\infty} \frac{\partial\tilde{U}_{t+1}/\partial B_{t+1}}{\partial\tilde{U}_t/\partial B_t} > 1 \Rightarrow \lim_{t\to\infty} \frac{B_{t+1}}{B_t} < 1$$

This contradicts the initial assumption of convergence, however, and it cannot be true that benefits approach a positive limit. Where $t\to\infty$

benefits will decline asymptotically towards zero; where the horizon is finite, and $t \to T$ for $T < \infty$, then the terminal benefit at T will be positive, and benefits will be declining at all durations up to T.

The declining benefits result is not robust, and there are plausible circumstances in which it does not perforce hold true. Within the present homogeneous population framework, the following three influences could potentially overturn the result.

(a) Non-separability of Indirect Utility Function

As is noted by Shavell and Weiss (1979), the declining benefits result is crucially dependent on the utility of the unemployed being separable in B and S. If non-separability is present, then the optimality conditions above are slightly modified, yielding an expression

$$\frac{\frac{\partial \tilde{U}_{t+1}}{\partial B_{t+1}}(B_{t+1}, S_{t+1})}{\frac{\partial \tilde{U}_t}{\partial B_t}(B_t, S_t) - \psi_t \frac{\partial^2 \tilde{U}_t}{\partial B_t \partial S_t}} = \frac{1}{1 + \frac{\psi_t}{(1-\nu_t)} \cdot \frac{\partial \nu}{\partial S_t}}$$

Since $\psi_t < 0$ and $\partial^2 \tilde{U}_t / \partial B_t \partial S_t > 0$ (in general), the additional cross-derivative term counteracts the argument used above. Moreover, even if $\partial^2 \tilde{U}_t / \partial B_t \, S_t$ is negligible and hence $\partial \tilde{U}_{t+1} / \partial B_{t+1} > \partial \tilde{U}_t / \partial B_t$ still holds true, it is quite conceivable for this to be satisfied by a rising benefit rate and falling search times. Consequently it is not possible to reach any definite conclusions under non-separable preferences, and the outcome depends on the particular utility function in question.

(b) Employment Probabilities Changing with the Duration of Unemployment

Empirical work suggests that the long-term unemployed have less chance of finding a job than those who have only been unemployed for a short time (McGregor (1978), Nickell (1979b)). To some extent this reflects the non-homogeneity of the population with the more disadvantaged groups constituting a larger proportion of the long-term unemployed (a situation discussed below). Nonetheless, even when the population is homogeneous it remains possible that individuals who are jobless for a long period experience a diminishing probability of securing re-employment. An example is where employers view prolonged unemployment as indicative of some undesirable characteristic possessed by the individual, regardless of whether or not this is actually the case.

Diminishing probabilities can be represented by making ν a function of t, so that $\nu_t = \nu(S_t, t)$, where $\partial\nu/\partial t < 0$. The key feature of the relationship is the effect of t on the returns to search, $\partial\nu/\partial S_t$, and, since ν declines with t, it is likely that $\partial\nu/\partial S_t$ also declines with t. Accordingly it is assumed that $\partial^2\nu/\partial S_t\partial t < 0$, and that at some point, $t=\bar{t}$, $\partial\nu/\partial S_t$ reaches a lower limit of zero ($\partial\nu/\partial S_t = \partial^2\nu/\partial S_t\partial t = 0$ for $t \geq \bar{t}$). In other words, once an individual has been unemployed for a sufficiently long time ($t>\bar{t}$), ν becomes fixed at some low constant value and job search ceases to have any effect.

These assumptions do not significantly alter the analysis, and the first-order conditions are the same as those for the standard model. Hence

$$\frac{\partial\tilde{U}_{t+1}/B_{t+1}}{\partial\tilde{U}_t/\partial B_t} = \frac{1}{1 + \frac{\psi_t}{1-\nu(S_t, t)} \cdot \frac{\partial\nu}{\partial S}(S_t, t)}$$

The argument given previously implies a declining optimal benefit rate up to time $\bar{t}$; thereafter $\partial\nu/\partial S_t$ becomes zero, so that $\partial\tilde{U}_{t+1}/\partial B_{t+1} = \partial\tilde{U}_t/\partial B_t$ and benefits are constant. The optimal benefit scheme differs from the standard case in that benefits can attain a limiting value within a finite time and need not decline to zero. On the other hand, the present case can never on its own produce a rising optimal benefit scheme.

(c) Preferences Changing with the Duration of Unemployment

Unemployment is known to have strong psychological effects on those concerned, which may vary systematically with duration (Hayes and Nutman (1981), Jahoda (1982)). These effects are difficult to identify precisely, but it may nonetheless be desirable to allow for them by permitting variation in individual preferences. Changing tastes are sometimes represented theoretically by an adaptive process, in which current utility is a function of consumption levels in past periods (Hammond (1976)). In the context of unemployment, however, it is more appropriate to introduce the duration of unemployment directly as a parameter in the utility function. Suppose therefore that preferences take the form $\tilde{U}_t = \tilde{U}(B_t, S_t, t)$, where t is the duration of unemployment. The function $\tilde{U}$ is probably declining with t, although the effect may become less pronounced as t increases. For the optimal benefit scheme the most important feature is the impact of t on $\partial\tilde{U}/\partial B_t$, and there is no conclusive a priori argument as to whether this is positive or negative: a simple discounting effect of t on $\tilde{U}$ would leave it negative, but on the general principle that low utility is associated with high marginal utility it would be positive.

The optimality conditions imply that

$$\frac{\partial\tilde{U}_{t+1}/\partial B_{t+1}(B_{t+1}, t+1)}{\partial\tilde{U}_t/\partial B_t(B_t, t)} = \frac{1}{1 + \dfrac{\psi_t}{(1-\nu_t)} \cdot \dfrac{\partial\nu}{\partial S_t}} \quad \text{and} \quad \frac{\partial\tilde{U}_{t+1}}{\partial B_{t+1}} > \frac{\partial\tilde{U}_t}{\partial B_t}$$

$$\text{(with } \psi_t < 0\text{)}$$

If $\partial\tilde{U}_t/\partial B_t$ is a decreasing function of t, then the outcome is merely to strengthen the previous conclusions, and optimal benefits decline over time. If, by contrast, $\partial\tilde{U}_t/\partial B_t$ rises with t, then increasing duration is conducive to the satisfaction of the inequality above. In cases where the impact of t is large relative to that of a change in B, it is possible that the optimality conditions are satisfied by benefits increasing over time. It can thus no longer safely be concluded that optimal benefits are declining.

The effect of changing preferences is independent of job search, and is most apparent in a model where no job search takes place. In that case the optimality conditions reduce to an equating of marginal utilities over time, so that

$$\frac{\partial\tilde{U}_t}{\partial B_t}(B_t, t) = \frac{\partial\tilde{U}_{t+1}}{\partial B_{t+1}}(B_{t+1}, t+1) \qquad \forall t$$

It follows that if $\partial^2\tilde{U}/\partial B\partial t < 0$, then optimal benefits will be falling (given $\partial^2\tilde{U}/\partial B^2 < 0$), and vice versa; in other words the direction of movement of B_t is determined entirely by the effect of t on the marginal utility of income. Adding search behaviour means that a second dynamic element is superimposed on the model, and, for the reasons given above, this acts in favour of declining optimal benefits. When $\partial\tilde{U}/\partial B_t$ declines with t the dynamic influences are working in the same direction and optimal benefits are unequivocally declining; when $\partial\tilde{U}/\partial B_t$ rises with t they are working in opposite directions and optimal benefits may rise or fall, depending on which influence predominates.

The exceptions to the uniformly declining benefits rule described so far are based on a single person model, but apply equally well to more general cases. Further doubts about the desirability of declining benefits arise within heterogeneous population models, and these are discussed below.

(ii) Optimal Time-varying Benefits in Models with a Heterogeneous Population

In reality the population's characteristics are almost certain to vary in ways that are relevant to the model. A particularly important feature is the fact that individuals do not share the same employment experiences, and that some individuals are more prone to spells of unemployment than others (Disney (1979)). Such differences imply that the cross-sectional composition of the unemployed changes over time, with the proportion of unemployment-prone increasing alongside the duration of unemployment. There is much empirical evidence on the differing employment experiences of different groups in the population: expected duration can vary, for example, with age (Narendranathan, Nickell and Stern (1985)), occupation (Nickell (1980)), race (Lynch (1983)) or location (Armstrong and Taylor (1985)). Allowing for this variation introduces an additional dynamic influence into the model, which may significantly modify the optimality results obtained for a homogeneous population.

The heterogeneous case is not susceptible to a recursive treatment (as in (i)), because the population composition is always dependent on past benefit levels. The analysis is therefore in terms of the Section (3) form of model, with movements occurring between employed and unemployed states. It is analytically convenient to assume that neither individuals nor the government discount utility over the infinite future horizon. Two particular cases are distinguished, according to the presence or absence of groups who undertake no job search.

Case 1 : Disadvantaged groups do not undertake job search

Suppose that the population is divided into two groups, differentiated by their employment conditions (generalising to allow more than two groups or variation in other characteristics is straightforward). A proportion

α of the population faces the employment conditions of Section (3), with a probability $\nu(s)$ of finding a job when unemployed; the remaining $1-\alpha$ of the population are 'disadvantaged' facing the same chance μ of losing a job, but having a chance $\nu^{\dagger}$ of finding a job, where $\nu^{\dagger} < \nu(s)$ for any value of s. Hence the disadvantaged have a relatively low probability of obtaining a job and are unable to influence this probability by undertaking job search. The reasons for such a population structure are left unspecified, and are irrelevant to the model's outcome: the position of the disadvantaged could be due to exogenous institutional circumstances, or alternatively could be explained by their own personal characteristics. The disadvanted will not search for no return, and their utility when unemployed takes the $\tilde{U}^{\dagger} = \tilde{U}(B,T)$ compared with $\tilde{U} = \tilde{U}(B, T-S)$ for the non-disadvantaged. As simplifying assumptions $\tilde{U}$ is separable in B and S, and includes a utility penalty relative to U sufficient to rule out moral hazard considerations.

In order to implement policy the government must either have full knowledge of the population groups, or at least be aware of their distribution (assignment uncertainty). Under complete information it would in general be desirable to operate a separate benefit scheme for each group, but this approach is ruled out by assumption in the analysis below. In reality the very large number of groups within the population makes it impossible to segregate all the groups, leading to the inclusion of more than one type of individual within the same benefit schedule.

The duration of unemployment is denoted by t, which can range from zero to infinity. The chance of being unemployed for exactly t periods declines with t for the non-disadvantaged at a varying discount rate per period of $1-\nu(s_t)$. Thus, if $\bar{\delta}$ is the probability of being unemployed for all durations, it follows that the chance of being unemployed for t periods is given by

$$\mu(1-\bar{\delta}) \cdot \prod_{j=0}^{t} (1-\nu_{j-1}) \qquad \text{(where } \nu_j \text{ denotes } \nu(S_j) \text{ and } \nu_{-1} \equiv 0\text{)}$$

The steady-state equation defining $\bar{\delta}$ is then

$$\bar{\delta} = \mu(1-\bar{\delta}) \sum_{i=0}^{\infty} \left[\prod_{j=0}^{i} (1-\nu_{j-1})\right]$$

so that

$$\bar{\delta} = \frac{\mu \sum_{i=0}^{\infty}\left[\prod_{j=0}^{i}(1-\nu_{j-1})\right]}{1 + \mu \sum_{i=0}^{\infty}\left[\prod_{j=0}^{i}(1-\nu_{j-1})\right]} = \frac{\mu}{C} \cdot \sum_{i=0}^{\infty}\left[\prod_{j=0}^{i}(1-\nu_{j-1})\right]$$

where $C \equiv 1 + \mu \sum_{i=0}^{\infty}\left[\prod_{j=0}^{i}(1-\nu_{j-1})\right]$ (= constant)

Hence the steady-state probability of being employed is

$$1 - \bar{\delta} = \frac{1}{C}$$

and the steady-state probability of being unemployed for duration t is

$$\delta_t = \frac{\mu}{C} \prod_{j=0}^{t} (1-\nu_{j-1})$$

The situation for the disadvantaged is analogous, with the changing discount rates $1-\nu_j$, $j = 0,\dots,\infty$, replaced by the constant rate $1-\nu^{\dagger}$. If $\delta^{\dagger}$ is the probability of the disadvantaged being unemployed for any duration, it must satisfy

$$\delta^{\dagger} = \mu(1-\delta^{\dagger}) \sum_{j=0}^{\infty} (1-\nu^{\dagger})^{j} = \frac{\mu}{\nu}(1-\delta^{\dagger}) \Rightarrow \delta^{\dagger} = \frac{\mu}{\mu+\nu^{\dagger}}$$

Similarly, the probability of the disadvantaged being employed is

$$1 - \delta^{\dagger} = \frac{\nu^{\dagger}}{\mu+\nu^{\dagger}}$$

and their probability of being unemployed for exactly t periods is

$$\delta_t^{\dagger} = \frac{\mu\nu^{\dagger}(1-\nu^{\dagger})^{t}}{\mu + \nu^{\dagger}}$$

Government policy resembles that in previous models, involving a time-varying benefit B_t financed by a uniform lump-sum payment.

Individual optimisation behaviour - In the present framework job search is only undertaken by the non-disadvantaged. The infinite horizon and no discounting assumptions imply that the expected sum of future utilities is infinite, so that optimisation instead involves the steady-state utility at any given date. This quantity is independent of whether or not the individual is unemployed, and is equal to

$$U^* \equiv \frac{1}{C}\left[U + \mu \sum_{i=0}^{\infty}\left[\prod_{j=0}^{i}(1-\nu_{j-1})\right]\tilde{U}_i\right]$$

Taking B_t, $t = 0,\ldots,\infty$, to be a known benefit schedule, the search optimum can be obtained by a conventional unconstrained maximisation, using the values S_t, $t = 0,\ldots,\infty$, as instruments. At the optimum

$$\frac{\partial U^*}{\partial S_t} = \frac{-1}{C^2}\cdot\mu\,\frac{\sum_{i=t+1}^{\infty}\left[\prod_{j=0}^{i}(1-\nu_{j-1})\right]}{(1-\nu_t)}\cdot\frac{-\partial\nu}{\partial S_t}\cdot C\cdot U^*$$

$$+\frac{1}{C}\left[\mu\prod_{j=0}^{t-1}(1-\nu_j)\,\frac{\partial\tilde{U}}{\partial S_t}+\mu\frac{\sum_{i=t+1}^{\infty}\left[\prod_{j=0}^{i}(1-\nu_{j-1})\right]\tilde{U}_i}{(1-\nu_t)}\,\frac{\partial\nu}{\partial S_t}\right] = 0$$

$$t = 0,\ldots,\infty$$

After rearrangement and simplification this implies the optimal search condition

$$\frac{-\partial\tilde{U}}{\partial S_t} = \frac{1}{\prod_{j=0}^{t}(1-\nu_j)}\cdot\frac{\partial\nu}{\partial S_t}\cdot\sum_{i=t+1}^{\infty}\left[\prod_{j=0}^{i}(1-\nu_{j-1})\,(U^* - \tilde{U}_i)\right]$$

$$t = 0,\ldots,\infty$$

When benefits are constant, the condition above reduces to that of Section (3).

Government Optimisation Behaviour - Using the probabilities derived above, the expected revenue constraint takes the form

$$R = \frac{\alpha}{C}\left[b - \mu \sum_{i=0}^{\infty} \prod_{j=0}^{i} (1-\nu_{j-1})B_i\right] + \frac{(1-\alpha)\nu^{\dagger}}{\mu+\nu^{\dagger}}\left[b - \mu \sum_{i=0}^{\infty} (1-\nu^{\dagger})^{i} B_i\right]$$

where R denotes the per capita revenue requirement in each period. Social welfare is some function of the utilities and employment probabilities, and under utilitarian social preferences (assumed in the analysis below) is expressible as

$$V = \frac{\alpha}{C}\left[U + \mu \sum_{i=0}^{\infty} \prod_{j=0}^{i} (1-\nu_{j-1})\, \tilde{U}_i\right] + \frac{(1-\alpha)\nu^{\dagger}}{\mu+\nu^{\dagger}}\left[U + \mu \sum_{i=0}^{\infty} (1-\nu^{\dagger})^{i}\, \tilde{U}_i^{\dagger}\right]$$

The government wishes to maximise V subject to the revenue constraint and the individual optimal search conditions at all dates. Solution is by the Lagrangian method, with a Lagrangian of the form

$$L = V + \phi\left(\frac{\alpha}{C}\left[b-\mu \sum_{i=0}^{\infty} \prod_{j=0}^{i} (1-\nu_{j-1})B_i\right] + \frac{(1-\alpha)\nu^{\dagger}}{\mu+\nu^{\dagger}}\left[b-\mu \sum_{i=0}^{\infty} (1-\nu^{\dagger})^{i} B_i\right] - R\right)$$

$$+ \sum_{k=0}^{\infty} \psi_k \left(\frac{-\partial\tilde{U}}{\partial S_k} - \frac{1}{\prod_{j=0}^{k} (1-\nu_j)} \cdot \frac{\partial\nu}{\partial S_k} \cdot \sum_{i=k+1}^{\infty}\left[\prod_{j=0}^{i} (1-\nu_{j-1})(U^{*}-\tilde{U}_i)\right]\right)$$

using B_t, b and S_t, $t=0,\ldots,\infty$ as instruments. The first-order conditions are:

$$\frac{\partial L}{\partial B_t} = \left[\frac{\alpha}{C}\mu \prod_{j=0}^{t} (1-\nu_{j-1}) + \frac{(1-\alpha)\nu^{\dagger}\mu}{\mu+\nu^{\dagger}} (1-\nu^{\dagger})^{t}\right]\left(\frac{\partial\tilde{U}}{\partial B_t} - \phi\right)$$

$$- \left[\sum_{k=0}^{\infty} \frac{\psi_k}{\prod_{j=0}^{k} (1-\nu_j)} \cdot \frac{\partial\nu}{\partial S_k} \sum_{i=k+1}^{\infty} \prod_{j=0}^{i} (1-\nu_{j-1})\right] \frac{\mu}{C} \prod_{j=0}^{t} (1-\nu_{j-1}) \frac{\partial\tilde{U}}{\partial B_t}$$

$$+ \sum_{k=0}^{t-1} \frac{\psi_k}{\prod_{j=0}^{k} (1-\nu_j)} \cdot \frac{\partial\nu}{\partial S_k} \cdot \prod_{j=0}^{t} (1-\nu_{j-1}) \cdot \frac{\partial\tilde{U}}{\partial B_t} = 0$$

(noting that $\frac{\partial\tilde{U}_t}{\partial B_t} = \frac{\partial\tilde{U}^{\dagger}}{\partial B_t}$, from separability of $\tilde{U}$) $\quad t = 0,\ldots,\infty$

$$\frac{\partial L}{\partial b} = \left(\frac{\alpha}{C} + \frac{(1-\alpha)\nu^{\dagger}}{\mu+\nu^{\dagger}}\right)\left(\frac{\partial U}{\partial b} + \phi\right) - \sum_{k=0}^{\infty} \frac{\psi_k}{\prod_{j=0}^{k}(1-\nu_j)} \cdot \frac{\partial \nu}{\partial S_k} \cdot \sum_{i=k+1}^{\infty} \prod_{j=0}^{i} (1-\nu_{j-1}) \cdot$$

$$\frac{1}{C} \cdot \frac{\partial U}{\partial b} = 0$$

(The conditions for S_t, $t = 0,\ldots,\infty$ are omitted).

Initially it is helpful to identify the signs of the Lagrange multipliers ϕ and ψ_k, $k = 0,\ldots,\infty$. A rise in the revenue requirement, R, would unequivocally reduce social welfare at the optimum, implying that $\partial L/\partial R = -\phi < 0$ and $\phi > 0$. In a similar way, an upward shift in the utility cost of search, $-\partial U/\partial S_k$, ceteris paribus, for any duration k would certainly reduce social welfare at the optimum, so that $\psi_k < 0$, $k = 0,\ldots,\infty$. These inequalities correspond to those obtained for the model of part (i).

The expression

$$-\sum_{k=0}^{\infty}\left[\frac{\psi_k}{\prod_{j=0}^{k}(1-\nu_j)} \cdot \frac{\partial \nu}{\partial S_k} \cdot \sum_{i=k+1}^{\infty} \prod_{j=0}^{i} (1-\nu_{j-1})\right]$$

appearing in all the first-order conditions given above, is invariant with respect to t and can therefore be represented by a positive constant, D (since $\psi_k < 0$, $\forall k$)

For the purposes of comparison, suppose firstly that $\alpha = 1$ and that the population is homogeneous, as in the previous model

$\boxed{\alpha = 1}$ In this case $\partial L/\partial B_t = 0$ reduces to

$$\frac{\partial L}{\partial B_t} = \frac{\mu}{C} \prod_{j=0}^{t} (1-\nu_{j-1}) \left(\frac{\partial \tilde{U}}{\partial B_t} - \phi\right) + \frac{D}{C}\mu \prod_{j=0}^{t} (1-\nu_{j-1}) \frac{\partial \tilde{U}}{\partial B_t}$$

$$+ \sum_{k=0}^{t-1} \frac{\psi_k}{\prod_{j=0}^{k} (1-\nu_j)} \cdot \frac{\partial \nu}{\partial S_k} \cdot \prod_{j=0}^{t} (1-\nu_{j-1}) \frac{\partial \tilde{U}}{\partial B_t} = 0 \qquad t = 0,\ldots,\infty$$

Dividing by $\prod_{j=0}^{t} (1-\nu_{j-1})$ and rearranging yields

$$\underbrace{\left[\frac{\mu(1+D)}{C} + \sum_{k=0}^{t-1} \frac{\psi_k}{\prod_{j=0}^{k} (1-\nu_j)} \cdot \frac{\partial \nu}{\partial S_k}\right]}_{+ve} \underbrace{\frac{\partial \tilde{U}}{\partial B_t}}_{+ve} = \underbrace{\frac{\phi\mu}{C}}_{+ve}$$

where the individual terms can be signed as above. The only expressions varying with t are the summation and $\partial \tilde{U}/\partial B_t$ (which changes as B_t changes). Given that $\psi_k < 0$, ∀k, then the summation is decreasing with t (rising in magnitude), and the positive square-bracketed expression is also decreasing. Consequently, for the equation to hold at all durations $\partial U/\partial B_t$ must be rising with t, so that B_t is falling (if $\tilde{U}$ is separable in B and S, and satisfies $\partial^2 U/\partial B^2 < 0$). Hence the uniformly declining benefits result carries forward to this model when the population is homogeneous.

$\boxed{\alpha=0}$ A second special case is the opposite extreme to $\alpha = 1$, with a homogeneous population of non-searching individuals. The first-order conditions for this situation imply that

$$\frac{\partial \tilde{U}}{\partial B_t} = - \frac{\partial U}{\partial b} = \phi \qquad t = 0,\ldots,\infty$$

so that the optimum involves a constant benefit rate. The outcome is similar to the basic model of Section (2), with moral hazard ruled out if $\tilde{U}$ is sufficiently low relative to U.

The main focus of interest is the effect of relaxing the homogeneity assumption, by setting $0 < \alpha < 1$.

$0 < \alpha < 1$ Dividing the condition $\partial L / \partial B_t = 0$ by $\prod_{j=0}^{t} (1-\nu_{j-1})$ and rearranging yields

$$\mu \left[\frac{\alpha}{C} + \frac{(1-\alpha)\nu^{\dagger}}{(\mu+\nu^{\dagger})} \cdot \frac{(1-\nu^{\dagger})^{t}}{\prod_{j=0}^{t} (1-\nu_{j-1})} \right] \left(\frac{\partial \tilde{U}}{\partial B_t} - \phi \right) + \left[\frac{D\mu}{C} + \sum_{k=0}^{t-1} \frac{\psi_k}{\prod_{j=0}^{k} (1-\nu_j)} \cdot \frac{\partial \nu}{\partial S_k} \right] \frac{\partial \tilde{U}}{\partial B_t} = 0 \qquad (++)$$

Consider firstly the position as $t \to \infty$. From the assumption that $\nu^{\dagger} < \nu(s)$, $\forall s$, it is known that the expression $(1-\nu^{\dagger})^{t} / \prod_{j=0}^{t} (1-\nu_{j-1})$ is increasing with t. The term

$$\sum_{k=0}^{t-1} \frac{\psi_k}{\prod_{j=0}^{k} (1-\nu_{j-1})} \cdot \frac{\partial \nu}{\partial S_k}$$

also increases in magnitude with t, but must be dominated asymptotically by $(1-\nu^{\dagger})^{t} / \prod_{j=0}^{t} (1-\nu_{j-1})$ (otherwise $\lim_{t\to\infty} \partial L / \partial B_t = 0$ would imply that $\partial \tilde{U} / \partial B_{\infty} = 0$, which is a contradiction if $\partial \tilde{U} / \partial B > 0$, $\forall B$).

It then follows that $\lim_{t\to\infty} \frac{\partial \tilde{U}}{\partial B_t} = \phi$

Now consider the situation for a finite unemployment duration, $t < \infty$. In the equation (++) above, the first square-bracketed term is known to be positive. The sign of the second square-bracketed term depends on the relative size of the two elements

$$\frac{D\mu}{C} = - \frac{\sum_{k=0}^{\infty} \left[\frac{\psi_k}{\prod_{j=0}^{k} (1-\nu_{j-1})} \cdot \frac{\partial \nu}{\partial S_k} \cdot \sum_{i=k+1}^{\infty} \prod_{j=0}^{i} (1-\nu_{j-1}) \right] \mu}{1 + \mu \sum_{i=0}^{\infty} \prod_{j=0}^{i} (1-\nu_{j-1})} \quad \text{and} \quad \sum_{k=0}^{t-1} \frac{\psi_k}{\prod_{j=0}^{k} (1-\nu_{j-1})} \cdot \frac{\partial \nu}{\partial S}$$

Noting that $\dfrac{\mu \sum\limits_{i=k+1}^{\infty} \prod\limits_{j=0}^{i} (1-\nu_{j-1})}{1+\mu \sum\limits_{i=0}^{\infty} \prod\limits_{j=0}^{i} (1-\nu_{j-1})} < 1$ for $k=0,\ldots,\infty$

it follows that as $t \to \infty$ the limit of the right-hand (negative) expression must exceed the left-hand expression in magnitude. Consequently, for t sufficiently large, the second square-bracketed expression in $\partial L / \partial B_t = 0$ must become negative. If $\partial L / \partial B_t = 0$ is to hold true for all t, it is then apparent that for t sufficiently large

$$\left(\frac{\partial \tilde{U}}{\partial B_t} - \phi\right) > 0 \quad \text{and} \quad \frac{\partial \tilde{U}}{\partial B_t} > \phi$$

Thus, for some finite t, sufficiently large,

$$\frac{\partial \tilde{U}}{\partial B_t} > \lim_{t\to\infty} \frac{\partial \tilde{U}}{\partial B_t}$$

and hence $B_t < \lim\limits_{t\to\infty} B_t$

given the assumptions made about preferences. This shows that optimal benefits must eventually start to increase with unemployment duration, approaching their asymptotic limit from below. The previous uniformly declining benefits result no longer holds true, although it remains the case that optimal benefits will be declining at earlier unemployment durations.

To see the latter point, consider the equation (++) above. At t=0 this becomes

$$\mu\left[\frac{\alpha}{C} + \frac{(1-\alpha)\nu^{\dagger}}{\mu+\nu^{\dagger}}\right]\left(\frac{\partial \tilde{U}}{\partial B_o} - \phi\right) + \frac{D\mu}{C} \cdot \frac{\partial \tilde{U}}{\partial B_o} = 0$$

Since $\frac{D\mu}{C} \cdot \frac{\partial \tilde{U}}{\partial B_o} > 0$, it follows that

$$\left(\frac{\partial \tilde{U}}{\partial B_o} - \phi\right) < 0 \quad \text{and} \quad \frac{\partial \tilde{U}}{\partial B_o} < \phi \left(= \frac{\partial \tilde{U}}{\partial B_\infty}\right)$$

Consequently, the initial benefit level B_0 is higher than the terminal level B_∞ which itself is higher than benefits at finite duration B_t, for t sufficiently large. Benefits must be decreasing at earlier durations, but ultimately start to increase over the final approach to B_∞.

The outcome can be explained intuitively in terms of the expected cross-sectional composition of the unemployed population. Among the newly unemployed (duration zero), the proportions are the same as in the entire population, with 1-α disadvantaged and α non-disadvantaged. As unemployment duration increases, however, relatively more of the non-disadvantaged succeed in finding work, so that the proportions alter in favour of the disadvantaged. Formally, the expected composition of the population unemployed for duration t is given by

Expected proportion of non-disadvantaged =

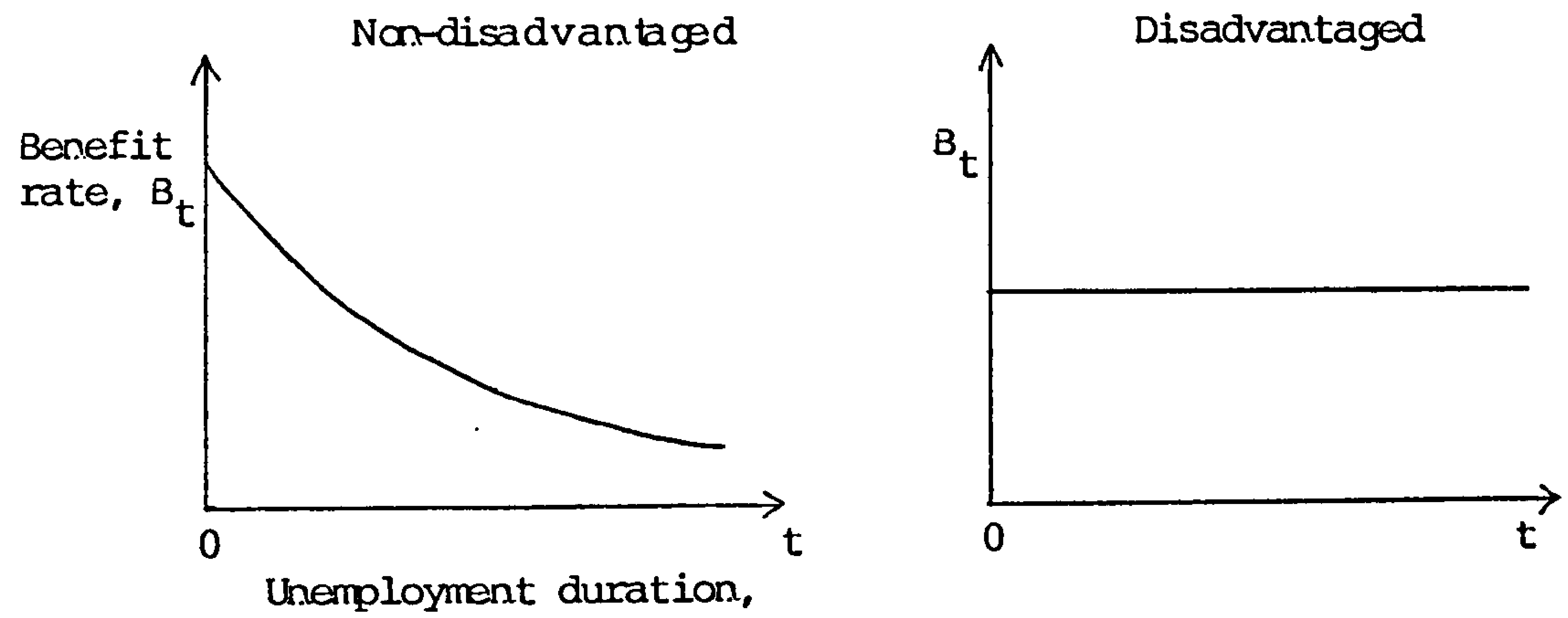
$$\frac{\alpha\delta_t}{\alpha\delta_t+(1-\alpha)\delta_t^\dagger} = \frac{\alpha}{\alpha + \dfrac{\nu^\dagger C}{(\mu+\nu^\dagger)} \cdot \dfrac{(1-\nu^\dagger)^t}{\prod_{j=0}^{t}(1-\nu_{j-1})} \cdot (1-\alpha)} \qquad (\to 0 \text{ as } t\to\infty)$$

Expected proportion of disadvantaged =

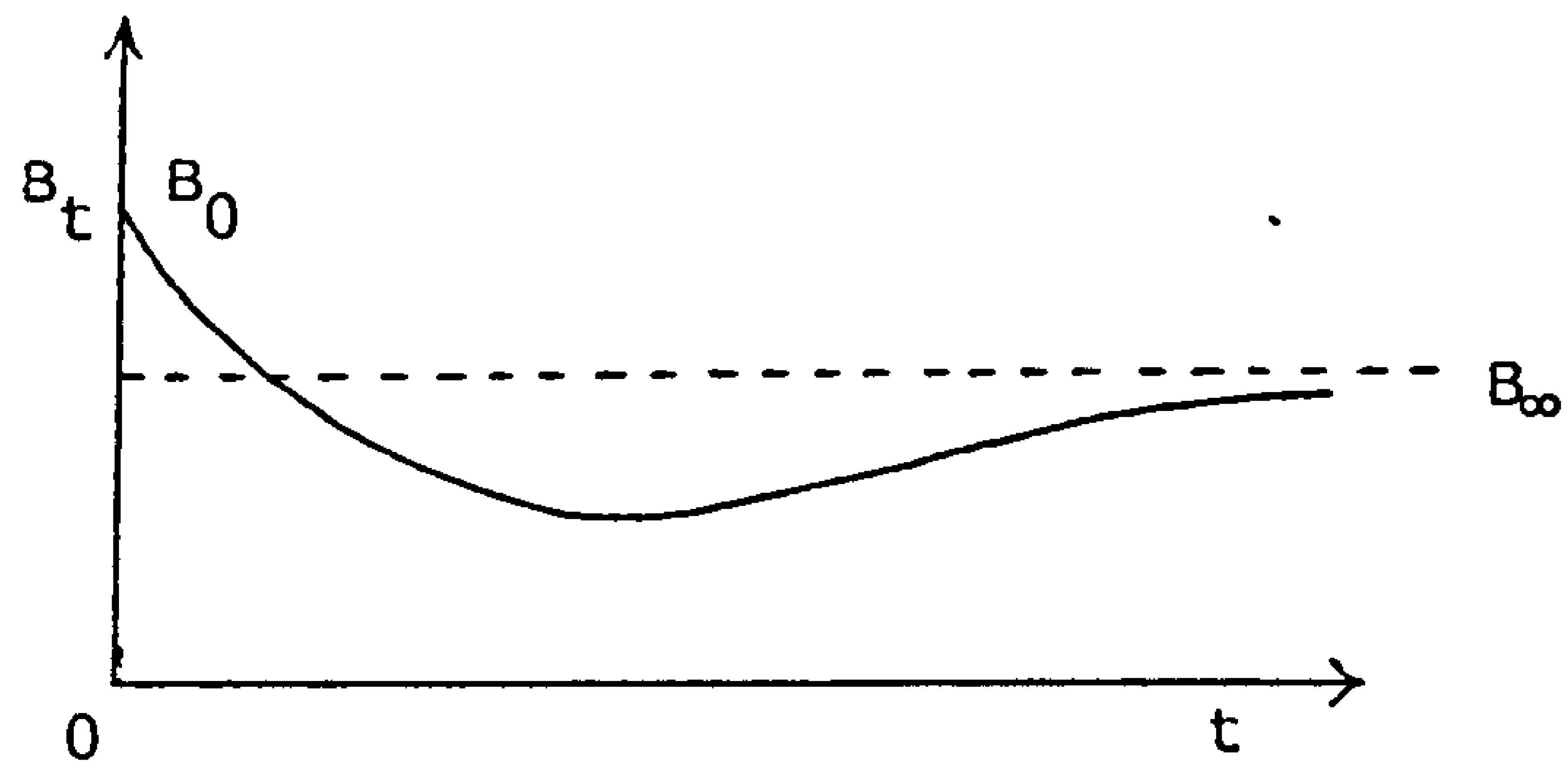
$$\frac{(1-\alpha)\delta_t^\dagger}{\alpha\delta_t+(1-\alpha)\delta_t^\dagger} = \frac{1-\alpha}{(1-\alpha) + \dfrac{(\mu+\nu^\dagger)}{\nu^\dagger C} \cdot \dfrac{\prod_{j=0}^{t}(1-\nu_{j-1})}{(1-\nu^\dagger)^t} \cdot \alpha} \qquad (\to 1 \text{ as } t\to\infty)$$

A government which could identify the two groups would ideally like to implement a separate benefit scheme for each : a constant rate for the non-searching disadvantaged, and a uniformly declining rate for the non-disadvantaged (declining because they undertake job search). In practice benefit schemes have to cater for heterogeneous populations, as was assumed in the foregoing analysis. The position may be illustrated diagrammatically as below:

Separate schemes



Unified scheme



Broadly speaking, the unified scheme is a 'weighted average' of two separate schemes, in which the weights vary with unemployment duration. As t increases, the disadvantaged form a relatively more important component of the unemployed, and the unified scheme approaches the time-invariant form associated with a homogeneous non-searching population. The unemployment duration is therefore being used as a source of information on population composition, and as an imperfect means of identifying the disadvantaged individuals.

Case 2 : All groups undertake job search

Suppose that the population is still divided into disadvantaged and non-disadvantaged groups, but that the disadvantaged are now able to influence their chance of finding work by undertaking job search. Let

$\nu^{\dagger}(S^{\dagger})$ denote the re-employment probability for the disadvantaged, where $\nu^{\dagger}(S^{\dagger}) < \nu(S)$ for all values of $S^{\dagger}$ and S. The analysis goes ahead in a similar way to case 1, except that there are now two separate optimal search conditions.

On dividing the first-order condition $\partial L/\partial B_t = 0$ by $\prod_{j=0}^{t}(1-\nu_{j-1})$ and rearranging, one obtains

$$\mu\left[\frac{\alpha}{C} + \frac{(1-\alpha)}{F} \cdot \frac{\prod_{j=0}^{t}(1-\nu^{\dagger}_{j-1})}{\prod_{j=0}^{t}(1-\nu_{j-1})}\right]\left(\frac{\partial\tilde{U}}{\partial B_t} - \phi\right) + \left[\frac{D\mu}{C} + \sum_{k=0}^{t-1}\frac{\psi_k}{\prod_{j=0}^{k}(1-\nu_j)} \cdot \frac{\partial\nu}{\partial S_k}\right]\frac{\partial\tilde{U}}{\partial B_t}$$
$$+ \frac{\prod_{j=0}^{t}(1-\nu^{\dagger}_{j-1})}{\prod_{j=0}^{t}(1-\nu_{j-1})}\left[\frac{F\mu}{G} + \sum_{k=0}^{t-1}\frac{\theta_k}{\prod_{j=0}^{k}(1-\nu^{\dagger}_j)} \cdot \frac{\partial\nu^{\dagger}}{\partial S^{\dagger}_k}\right]\frac{\partial\tilde{U}}{\partial B_t} = 0$$
$$t=0,\ldots,\infty$$

where θ_k, $k=0,\ldots,\infty$, are the Lagrange multipliers on the search constraints of the disadvantaged, and F, G are the equivalent expressions to C, D, with $\nu^{\dagger}_t$ substituted for ν_t. As before the explosive term $\prod_{j=0}^{t}(1-\nu^{\dagger}_{j-1})/\prod_{j=0}^{t}(1-\nu_{j-1})$ dominates as $t\to\infty$, but in this case it is also attached to the final expression related to the search condition for the disadvantaged. Hence asymptotically there emerges an equation

$$\left[\frac{\mu(1-\alpha+F)}{G} + \sum_{k=0}^{\infty}\frac{\theta_k}{\prod_{j=0}^{k}(1-\nu^{\dagger}_j)} \cdot \frac{\partial\nu^{\dagger}}{\partial S^{\dagger}_k}\right]\frac{\partial\tilde{U}}{\partial B_t} = \frac{\phi\mu(1-\alpha)}{G}$$

which is identical in form to that in case 1 where $\alpha=1$, and implies benefits declining to zero. Thus, unlike case 1, the eventual dominance of the search effect for the disadvantaged will cause benefits to decline asymptotically towards zero.

This outcome is intuitively appealing, since job search effects never disappear from the model, even as $t \to \infty$. The cross-section of the unemployed is approaching a homogeneous disadvantaged population, which, given that it undertakes job search, would still give rise to declining optimal benefits. Nonetheless, it remains possible that at earlier unemployment durations optimal benefits may rise over time, owing to the population composition effects which are absent from the homogeneous population case.

A model aiming to represent reality closely would have to include many different groups of unemployed people facing different search probabilities. It is also likely that one or more disadvantaged groups experience employment probabilities largely outside their control. A more extensive model would amalgamate cases one and two above, and could easily produce a very complex optimum. In these circumstances it is difficult to generalise about the optimal time pattern of benefits, a problem further compounded if (plausible) features like a non-separable indirect utility function are introduced.

(5) Other Relevant Issues

(i) Macroeconomic Conditions

The assumption of a fixed, exogenous macroeconomic situation is not necessarily appropriate, and employment conditions may be dependent on policy decisions. This issue is considered within an explicitly macroeconomic model in Chapter 6.

(ii) Insurance Premia

In a state-run system of unemployment insurance there is no obvious distinction between compulsory contributions to the insurance scheme and other forms of public finance (such as direct taxes). Hence it is not really possible to talk about insurance premia in the usual sense of the term. This notwithstanding, one can still pose the question of whether or not the state's method of financing should make use of information on an individual's employment experiences. The discussion below considers this issue, and contrasts the outcome with that under a hypothetical private insurance policy.

(a) Private Insurance

Despite the non-existence of private unemployment insurance, it is of interest briefly to consider the likely features of such schemes if they were to be implemented. The same points can also apply to state schemes, wherever they seek to imitate competitive private insurance.

For a private insurance policy to operate, it must be true that the expected returns to both insurer and insured are (or are believed to be) at least as good as the no policy situation. When individual risk patterns are observable by both parties, it is possible to tailor separate insurance schemes for each individual, and there is no theoretical reason why private insurance is not feasible. Under competitive conditions such schemes should just break even, and there would be no cross-subsidisation between different groups of the insured. In practice it is frequently the case that the insurer cannot observe the risk characteristics of the insured, leading to the problem of 'adverse selection'. As shown in Rothschild and Stiglitz (1976), the inability to distinguish different risk categories may result in the non-existence of a scheme which both breaks even and insures the whole population. The standard explanation of the absence of private unemployment insurance is in terms of such market failures.

The one obvious source of information on the risk of unemployment is an individual's previous employment record. As the current employment spell proceeds, this is continually being updated in a favourable direction, so that it becomes increasingly less likely that an individual is unemployment-prone. Thus, if private insurance was to operate (for, say, a subset of the population) it would generally be desirable for premia to be adjusted downwards with the current employment spell (from a starting point determined by past employment experience). This is what happens in other types of private insurance, where individuals with a history of making no previous claims are offered discounts, 'no claims bonuses', and so on. It is also the idea behind the 'experience rating' in the U.S. (Becker (1972)), where a firm's insurance contributions increase with the number of recent lay-offs it has imposed.

(b) State Insurance

State unemployment insurance schemes are generally compulsory, and can be financed by the imposition of taxes or other mandatory payments. Individuals with a low risk of unemployment may well make a net loss from the scheme, but are nevertheless forced to participate in it; conversely, unemployment-prone individuals are likely to make a net gain. Given that redistribution is taking place, it is rational to set the benefit rate on social welfare principles, as has been done throughout the models discussed in this chapter. On the whole the result should be superior to individualised private schemes (if these were feasible), since actuarially fair insurance preserves the initial income distribution and is desirable in social welfare terms only to the extent that the initial position is desirable.

The aim of the discussion below is to consider the optimal time pattern of compulsory state insurance contributions, in cases where they may be varied to take account of the current employment spell. It is assumed that the government cannot directly observe an individual's risk of job loss, but can use recent employment experience as an indicator of this.

Let the model take a discrete time form, with a similar preference and probability structure to those in Sections (3) and (4). Since job search is not germane to this particular issue, it is excluded for simplicity's sake. Unemployment benefits are paid out at the constant rate, B, but premium payments, b_t, can vary with the duration of the current period of employment. Utilities are therefore expressible as

$U_t = U(\underset{\sim}{p}, w, wT-b_t)$ for those employed for duration t

$\tilde{U} = \tilde{U}(\underset{\sim}{p}, B, T)$ for the unemployed

with notation as in previous models. $\tilde{U}$ includes a utility penalty sufficient to rule out moral hazard. Let U^*_t and U_t denote the expected sums of future utilities for those employed for t periods and those unemployed respectively. Then U^*_t satisfies

$$U^*_t = U_t + \frac{\mu}{1+\rho}\tilde{U}^* + \frac{(1-\mu)}{1+\rho}U^*_{t+1}$$

where μ is the chance of job loss in any period, and ρ is the per period discount rate on utility. Repeatedly substituting for U^*_{t+1}, U^*_{t+2} etc. yields

$$U^*_t = \sum_{i=0}^{\infty}\left(\frac{1-\mu}{1+\rho}\right)^i U_{t+i} + \frac{\mu}{1-\mu}\tilde{U}^* \sum_{i=1}^{\infty}\left(\frac{1-\mu}{1+\rho}\right)^i = \sum_{i=0}^{\infty}\left(\frac{1-\mu}{1+\rho}\right)^i U_{t+i} + \frac{\mu}{\mu+\rho}\tilde{U}^*$$

$\tilde{U}^*$ must satisfy $\qquad t=0..., \infty$

$$\tilde{U}^* = \tilde{U} + \frac{\nu}{1+\rho}U^*_o + \frac{(1-\nu)}{1+\rho}\tilde{U}^*$$

where ν is the fixed probability of obtaining a job in the next period. Substituting for U_o^* from the equation above and rearranging gives

$$\tilde{U}^* = \frac{\mu+\rho}{\rho(\mu+\nu+\rho)}\left[(1+\rho)\,\tilde{U} + \nu \sum_{i=o}^{\infty}\left(\frac{1-\mu}{1+\rho}\right)^i U_i\right]$$

while substituting for $\tilde{U}^*$ in the expression for U_t^* yields

$$U_t^* = \sum_{i=o}^{\infty}\left(\frac{1-\mu}{1+\rho}\right)^i U_{t+i} + \frac{\mu\nu}{\rho(\mu+\nu+\rho)} \sum_{i=o}^{\infty}\left(\frac{1-\mu}{1+\rho}\right)^i U_i + \frac{\mu(1+\rho)}{(\mu+\nu+\rho)}\,\tilde{U} \qquad t=0,\ldots,\infty$$

The last two equations are in terms of current utilities, and can therefore be used for optimisation.

The heterogeneous population is divided into two groups, differing both in their wage rates, W_1 and W_2, and their degree of job security, μ_1 and μ_2. Without loss of generality, suppose that group one has the more secure employment, so that $\mu_1 < \mu_2$. Population composition is such that a proportion α_1 are in group one and α_2 are in group two, where $\alpha_1 + \alpha_2 = 1$. These values are known to the government, but there is assignment uncertainty preventing the direct identification of the group to which an individual belongs.

Social welfare is assumed to be utilitarian, and is dependent on the steady-state employment and unemployment probabilities. Analogously to previous models, the probabilities are such that

$\frac{\mu_k}{\mu_k+\nu}$ = steady-state probability of being unemployed, k=1,2

$\frac{\mu_k}{\mu_k+\nu}\nu(1-\mu)^j$ = steady-state probability of being employed for j periods, k=1,2

The social welfare function can therefore be written as

$$V = \sum_{k=1,2} \alpha_k \left[\frac{\mu_k}{\mu_k+\nu} \tilde{U}^* + \frac{\mu_k}{\mu_k+\nu} \sum_{i=o}^{\infty} (1-\mu)^i U_i^{k*} \right]$$

$$= \sum_{k=1,2} \alpha_k \frac{\mu_k}{\mu_k+\nu} \left[\frac{(1+\rho)}{\rho} \tilde{U} + \frac{\nu}{\rho} \sum_{i=o}^{\infty} \left(\frac{1-\mu_k}{1+\rho} \right)^i U_i^k \right.$$

$$\left. + \nu \sum_{j=o}^{\infty} (1-\mu_k)^j \sum_{i=o}^{\infty} \left(\frac{1-\mu_k}{1+\rho} \right)^i U_{j+i}^k \right]$$

substituting for $\tilde{U}^*$ and U_j^*, $j=0,\ldots,\infty$, from the equations above. In a similar way the expected revenue constraint takes the form

$$\sum_{k=1,2} \alpha_k \frac{\mu_k}{\mu_k+\nu} \left(\sum_{i=o}^{\infty} \nu(1-\mu_k)^i b_i \right) = R + \sum_{k=1,2} \alpha_k \frac{\mu_k}{\mu_k+\nu} \cdot B$$

where R is the revenue requirement in each period. Solution is by the Lagrangian method with b_t, $t=0,\ldots,\infty$, and B as instruments:

$$L = V + \phi \left[\sum_{k=1,2} \alpha_k \frac{\mu_k}{\mu_k+\nu} \sum_{i=0}^{\infty} \nu (1-\mu_k)^i b_i - R - \sum_{k=1,2} \alpha_k \frac{\mu_k}{\mu_k+\nu} \cdot B \right]$$

$$\frac{\partial L}{\partial B} = 0 \Rightarrow \frac{(1+\rho)}{\rho} \frac{\partial \tilde{U}}{\partial B} = \phi$$

$$\frac{\partial L}{\partial b_t} = 0 \Rightarrow \frac{\frac{(1+\rho)}{\rho} \left(\frac{\alpha_1 \mu_1}{\mu_1+\nu} \cdot \frac{\partial U_t^1}{\partial b_t} + \left(\frac{1-\mu_2}{1-\mu_1} \right)^t \frac{\partial U_t^2}{\partial b_t} \right)}{\frac{\alpha_1 \mu_1}{\mu_1+\nu} + \left(\frac{1-\mu_2}{1-\mu_1} \right)^t \frac{\alpha_2 \mu_2}{\mu_2+\nu}} = -\phi \qquad t=0,\ldots,\infty$$

(after dividing by $(1-\mu_1)^t$ and rearranging)

Let the large expression equated with $-\phi$ be denoted by $A(t, b_t)$. If A rises with t for a given value of b, then b_t must be increasing with t in order to ensure that $A = -\phi$ holds (noting that A and $\partial U/\partial b$ are negative, and $\partial^2 U/\partial b^2 < 0$ from the assumption of diminishing marginal utility of income); conversely, if A falls with t for a given b, then b_t must be declining with t.

The direction of movement of A with t is given by the partial derivative $\partial A/_{\partial t}$, which satisfies

$$\frac{\partial A}{\partial t} = \frac{\dfrac{(1+\rho)}{\rho} \cdot \dfrac{\alpha_1\mu_1}{\mu_1+\nu} \cdot \dfrac{\alpha_2\mu_2}{\mu_2+\nu} \cdot \ln\left(\dfrac{1-\mu_2}{1-\mu_1}\right)\left(\dfrac{\partial U_t^2}{\partial b_t} - \dfrac{\partial U_t^1}{\partial b_t}\right)}{\left(\dfrac{\alpha_1\mu_1}{\mu_1+\nu} + \left(\dfrac{1-\mu_2}{1-\mu_1}\right)^t \dfrac{\alpha_2\mu_2}{\mu_2+\nu}\right)^2}$$

Hence $\frac{\partial A}{\partial t}$ has the same sign as

$$\ln\left(\frac{1-\mu_2}{1-\mu_1}\right)\left(\frac{\partial U_t^2}{\partial b_t} - \frac{\partial U_t^1}{\partial b_t}\right)$$

By assumption $\mu_2 > \mu_1$, so that $\ln\left(\frac{1-\mu_2}{1-\mu_1}\right) < 0$

and $\frac{\partial A}{\partial t} \gtreqless 0$ as $\frac{\partial U_t^2}{\partial b_t} \lesseqgtr \frac{\partial U_t^1}{\partial b_t}$

Provided that leisure is a normal good, $\partial U_t/_{\partial b_t}$ is increasing with the wage rate for all t.

Consequently $\frac{\partial A}{\partial t} \gtreqless 0$ as $W_2 \lesseqgtr W_1$

which means that at the optimum (for $\mu_2 > \mu_1$),

$$\left.\begin{matrix} W_2 > W_1 \\ W_2 = W_1 \\ W_2 < W_1 \end{matrix}\right\} \Rightarrow \text{premia} \begin{cases} \text{decreasing} \\ \text{constant} \\ \text{increasing} \end{cases}$$

In other words, if W and μ are positively correlated, premia rise with t, and vice versa.

Setting t to 0 and ∞ in the first-order condition yields

$$\frac{\partial \tilde{U}}{\partial B} = -\left[\frac{\alpha_1\mu_1}{\mu_1+\nu} \cdot \frac{\partial U_o^1}{\partial b_o} + \frac{\alpha_2\mu_2}{\mu_2+\nu} \cdot \frac{\partial U_o^2}{\partial b_o}\right] = -\frac{\partial U_\infty^1}{\partial b_\infty}$$

As $t \to \infty$, the proportion of group two individuals among the employed approaches zero (given that $\mu_2 > \mu_1$), and b approaches the level appropriate to a homogeneous group one population.

This outcome arises from straightforward redistributive considerations, and is intuitively quite obvious. On utilitarian grounds the government would like to impose higher contributions on individuals with a higher wage rate (and thus a lower marginal utility of income). But if b is restricted to be a function of t alone, then there can be redistribution among the employed only if job security is systematically related to the wage rate. In that event the expected proportion of the employed with a high wage rate rises or falls with t, allowing b_t to be adjusted accordingly. The case for moving premia is entirely motivated by the desire to make income transfers between the employed, and has little directly to do with insurance matters. It should be noted that simultaneous variation in W and μ is the only situation in this model giving rise to non-constant premia - for variations in any other parameters or in W and μ individually the optimal premia are constant (as there is no sense in which employment duration conveys useful information).

The model has close parallels both with the characteristics models of Chapter 2 and the varying unemployment benefits model of Section (4) above. In the latter case it is almost the reverse of the earlier model, with benefits now held constant and contributions allowed to vary in line with changing population composition. The following further comments can also be made.

Firstly, the model is highly simplified in its assumption of two groups and perfect correlation between μ and W. In reality there would be many different values of μ and W, and less than perfect correlation between them. Generally speaking, the lower is the correlation involved, the less will be the case for premia moving with t.

Secondly, if μ and W are directly observable, then the policy optimum implies a separate constant payment charged to each $\{\mu, W\}$ group (essentially, a first-best income transfer). Equivalently, if μ alone is directly observable, the model reduces to a special case of the optimal taxation of a fixed characteristic. For non-constant premia to be justified it must be true that direct contributions cannot be levied - the most likely reason for this is assignment uncertainty, but other forms of exogenous constraint are also possible.

Thirdly, casual empiricism suggests that the best paid employment also has greater job security, so that μ and W are negatively correlated. This implies that optimal premia rise over time, in contrast with the outcome of a hypothetical private scheme.

(iii) Saving Behaviour

In the models above it has been assumed that consumption is always equal to current income, so that there is no saving or borrowing taking place. The possibility that individuals anticipate unemployment and try to make their own provisions for it is consequently ruled out, and the unemployed are totally reliant on their benefit receipts. This view is not especially implausible, as individuals in lowly paid, insecure jobs do not usually accumulate large savings, and have no easy access to borrowing. Nevertheless, in situations with zero or negligible unemployment benefits there would be a strong incentive for individuals to save part of their income, and it is therefore of some interest to consider this issue.

The role played by savings is a central concern of the papers by Baily (1978) and Flemming (1978). As Flemming points out, in an infinite horizon model with identical individuals, a perfect capital market and a zero interest rate, there is no need for unemployment benefits. This follows simply from the fact that under these conditions individuals can freely and costlessly transfer income between employed and unemployed states, and do so in a way which maximises their expected utility, giving the same result as obtained through optimal unemployment benefits (with no moral hazard). The observation does not extend to populations varying in the wage rate and job security, as unemployment benefits can then be justified by redistributive considerations. Where there are perfect capital markets with positive interest rates, individual saving/borrowing decisions do not necessarily coincide with the government's, and therefore impose an additional constraint on the government's optimisation problem. This case is awkward to analyse in a general theoretical manner, but Flemming's numerical examples suggest that the optimal benefit level is lower when saving and borrowing is present. Such an outcome is intuitively appealing, since individual saving at the optimum is liable to reduce the need for benefits, even if it is not a particularly desirable feature from the government's viewpoint.

In relating models with rational savings to actual behaviour, there are two main points to be made:

(a) It is certainly not true that individuals face a perfect capital market, and the ability of unemployment-prone individuals to borrow is severely restricted. This implies that individuals cannot fully plan their future consumption, even if they wish to do so.

(b) The assumption of rational saving behaviour seems to overstate both the willingness and capacity of individuals to make such long-term decisions. People are likely to differ in this respect, but even the most 'responsible' and well-informed would not find it easy to make accurate planning decisions related to job security. Individuals may well be content to transfer their provision for unemployment collectively to the government.

In reality there does exist a certain amount of workers' saving and borrowing, and a truly comprehensive theoretical discussion would have to include them. These aspects are not necessarily well depicted by a perfect capital market and rational saving behaviour, however, and to extend the model in that way is not obviously to improve it. The problem of adequately representing workers' savings is difficult, and no attempt is made here to solve it, in the presumption that it is not of critical importance in the present context. Nonetheless, one should acknowledge that the presence of savings can potentially influence the welfare optimum.

(iv) Variable Size of Labour Force

It is sometimes argued that unemployment benefits have an impact on the size of the labour force, with high benefits encouraging some individuals to register as unemployed who would not otherwise have done so. The previous models in this chapter have taken the labour force to be independent of the benefit level, and have therefore excluded this effect; in practice its importance is not clear, although it probably exists to some degree. A simple variable labour force model is considered below.

For labour force variability to occur at all, the model must have two features:

(a) A significant group of individuals must have sufficient non-employment income to be able to subsist without employment or receipt of benefits. Otherwise people simply have no choice but to register as part of the working population, and thereby guarantee at least a minimum income. In reality most households have relatively low levels of unearned income, leaving little question that at least one member must seek employment.

The participation issue is therefore largely influenced by the behaviour of secondary workers in households.

(b) The population must have varying degrees of reluctance to join the labour force. If there is no cost attached to registering as unemployed, then all individuals have an incentive to claim unemployment benefits, even if they have no intention of finding employment. For this not to happen there must be some disutility attached to signing on, for example, the stigma of being unemployed or the physical inconvenience of claiming. Given that individuals differ in these attitudes, there may then be a gradual inflow into the labour force, as benefits increase sufficiently to overcome the reluctance of various groups to claim them.

By including these two properties, the following model allows for a labour force of varying size. Suppose that the basic structure is that of Section (2), with a single date and no job search. The total population is H, and employment is fixed at E, so that the remaining H-E individuals have to decide whether or not to register as unemployed. All individuals are assumed to have the same unearned income, M(>0), but they differ in their attitude to being unemployed. This is represented as a fixed monetary cost, k, to be set against the benefit receipt, B, implying that

$$\tilde{U} = \tilde{U}(\underline{p},\ M+B-k,\ T)$$

Preferences are therefore discontinuous at the point of unemployment, and k is the compensating variation measure of the disutility of being unemployed. Values of k varying between 0 and $\bar{k}$, and are distributed according to the density function f(k), satisfying

$$\int_0^{\bar{k}} f(k)\, dk = H-E$$

(there are no negative k's, so nobody actually enjoys being unemployed). The E employed individuals are not included in f(k), and are assumed to have a value k=0 when considering whether or not to declare themselves unemployed - this maximisess the incidence of moral hazard, and allows it to be imposed as a prior constraint.

Within the framework above, a person not employed will register as officially unemployed if $B > k$, that is, if the benefit payments exceed the monetary disutility of being on the dole. Given the distribution of k it follows that for any particular value of B, a total of F(B) individuals register as unemployed, where F is the cumulative distribution function associated with the density function f. By definition $F'(B) = f(B) > 0$, so that a rise in the benefit rate increases the size of the labour force, equal to E+F(B). Benefits are financed by a uniform lump sum, b, charged only to those in employment (the model could alternatively be defined with b paid by all individuals not registered as unemployed, thus increasing the incentive to enter the labour force). Population structure and preferences are such that

E = number of employed, with a single wage rate and utilities $U(\underset{\sim}{p}, w, wT+M-b)$

F(B) = total number of registered unemployed, comprising f(k) individuals in each group $0 \leqslant k \leqslant B$, with utilities $\tilde{U}(\underset{\sim}{p}, M+B-k, T)$

H-E-F(B) = number of non-participants in labour force, with utilities $\tilde{U}(\underset{\sim}{p}, M, T)$

where $\tilde{U}$ is the rationed version of U. The utilitarian social welfare function is given by

$$V = EU(\underline{p}, w, wT+M-b) + \int_0^B \tilde{U}(\underline{p}, M+B-k, T) f(k) dk +$$

$$+ (H-E-F(B)) \tilde{U}(\underline{p}, M, T)$$

and the government's revenue constraint by

$$R + BF(B) = Eb$$

To avoid the presence of moral hazard, one also has to impose $U \geqslant \tilde{U}(\underline{p}, M+B, T)$. The Lagrangian and first-order conditions are

$$L = V + \phi (Eb-BF(B)-R) + \psi(U-\tilde{U}(\underline{p}, M+B, T))$$

$$\frac{\partial L}{\partial B} = \tilde{U}(\underline{p}, M, T) f(B) + \int_0^B \frac{\partial \tilde{U}}{\partial B} (\underline{p}, M+B-k, T) f(k) dk$$

$$- \tilde{U}(\underline{p}, M, T) f(B) - \phi F(B) - \phi Bf(B) - \psi \frac{\partial \tilde{U}}{\partial B} (\underline{p}, M+B, T)$$

$$= \int_0^B \frac{\partial \tilde{U}}{\partial B} (\underline{p}, M+B-k, T) f(k) dk - \phi(F(B)+Bf(B))$$

$$- \psi \frac{\partial \tilde{U}}{\partial B} (\underline{p}, M+B, T) = 0$$

$$\frac{\partial L}{\partial b} = E \frac{\partial U}{\partial b} + \phi E + \psi \frac{\partial U}{\partial b} = 0$$

$$\frac{\partial L}{\partial \psi} = U-\tilde{U} \geqslant 0 \qquad \psi(U-\tilde{U}) = 0$$

These differ from the standard case only in the condition $\partial L/_{\partial B} = 0$, where B appears as the upper limit of the integral and there is the additional term $-\phi Bf(B)$. Thus, a rise in benefits increases the size of the labour force, extending the marginal utility effect via the $\partial \tilde{U}/_{\partial B}$ terms but also incurring an extra revenue cost $\phi Bf(B)$. As previously, the outcome is likely to be at the corner solution $U=\tilde{U}(p, M+B, T)$ if leisure is a normal good, although the presence of the term $\phi Bf(B)$ can potentially give an interior solution.

There is no particular reason to suppose that a variable labour force model produces a dramatically different and less favourable outcome than the static case. In social welfare terms it can even have a beneficial effect, as it widens the range of individuals over which redistribution takes place - from this perspective the fact that it may lower the optimal benefit rate and increase the rate of registered unemployment is irrelevant. On the other hand, the outcome might be less favourable when there is wide variation in unearned incomes (unlike the constant M above) and society is inequality averse. An inflow of more prosperous non-working individuals into registered unemployment as benefits increased might then constitute a genuine burden on the 'bona fide' unemployed and on social welfare.

The decision on whether or not to include labour force variation as a major influence in the models depends on its prominence in reality. Actual unearned incomes are not very high in most cases, and this will tend to reduce the relative costs involved in being unemployed. For example, in the model above the worst thing that can happen to an individual is to be on the unemployment register with a benefit rate of zero, thus facing the cost k but receiving no benefit to compensate - if net income equivalents never become negative (as seems reasonable), this means that $\bar{k} < M$ and that the general level of k values tends to decrease with M. Where M is insignificant, there is a probability that k is also small, and that the extent of labour force variation is limited. On this basis one can say that models with a varying labour force are more general than the fixed case (since the latter is included as a special case of the former), but that their additional complexity is not necessarily capturing an important part of the problem. There are some situations in which labour force movements can play a much more prominent role. For example, the Harris-Todaro model of developing economies (Todaro (1969)) is formally similar to the case considered above, and represents a situation where rises in urban incomes provoke a large scale inflow into the urban labour force from rural areas. The effects may be large enough to nullify any attempts to improve urban conditions, and are thus of an entirely different magnitude from those pertaining to a developed economy.

(v) Worksharing

One final issue is noteworthy. Given an exogenous macroeconomic situation with a fixed total number of working hours in each occupation, the optimal policy response is to share out the work available over the entire group eligible to do it. At the optimum the marginal social utility of extra working time is equated for all individuals within a particular occupation, an example of the quantity allocation of Section (3). Worksharing of this sort does not take place in reality, and unemployment is concentrated among a small group of individuals who are completely idle. Such an arrangement is the least efficient possible in social welfare terms, whence the need for unemployment benefits.

Worksharing would be practicable only in situations of long-term unemployment, caused by chronic demand deficiency or structural problems (and the former case ought to be avoidable by rational intervention). Its absence from the foregoing optimality models can therefore be justified if they are supposed to be representing short-term frictional unemployment. For other situations worksharing is in theory preferable to redistribution by unemployment benefits, so the optimisation framework does not itself explain the use of a benefit system: benefits only arise as the result of an exogenous fixity of existing work patterns. As with the question of macroeconomic conditions, this demonstrates the partial nature of the models being considered.

(6) Conclusion

There is no single conclusion, but the following points merit re-emphasis:

(i) Even in the very simplest models, little can be said about optimal benefit levels without specifying the functional forms of individual preferences. Such an outcome is a common feature of optimal policy models, encountered, for example, in optimal taxation theory.

(ii) The discussion has sought to indicate the wide range of features which may influence optimal unemployment benefits (and the coverage is probably not comprehensive). To include them all simultaneously in a unified theoretical model would be well nigh impossible, so one is left with several alternative simplified cases highlighting different aspects of the problem. In trying to represent an actual economy, it would be necessary to make an empirically based judgement on which are the most important influences.

(iii) The format of optimisation used is not necessarily appropriate to all unemployment situations. In particular, positions with chronic demand deficient or structural unemployment generally need additional policy tools, whose inclusion may significantly alter the nature of the policy optimum. Models of optimal unemployment benefits are most easily compatible with the presence of only frictional unemployment, and to depict positions which do not fall into this category a different theoretical approach may be preferable.

CHAPTER 6 : MACROECONOMICS, THE UNEMPLOYED AND UNEMPLOYMENT BENEFITS

(1) Introduction

Macroeconomics had its origins in a single policy issue, the chronic unemployment of the nineteen-twenties and thirties. Keynes's 'General Theory' was written as the formalisation of certain (already existing) policy beliefs in a way that would be acceptable to the economics profession (Keynes (1936), Chapter 1); it did not arise from a spontaneous desire to reformulate economics, but from the wish to further a particular view of policy. The structure of macroeconomic theory reflects these beginnings, in that it depends on a number of casual empirical observations and is not constructed on axiomatic lines. To the extent that it preserves its original identity, macroeconomics is still chiefly geared to the problems of chronic unemployment and excess capacity.

In view of this background it is a little surprising that unemployment does not feature more strongly in Keynesian models. The original neoclassical approach to macroeconomics believed in an equilibrium tendency towards full employment, and understandably had little to say about the unemployed. But Keynesian theory rests on the idea that there is no such tendency, and that unemployment is a common occurrence in unregulated capitalist economies. Consequently it might well be expected that Keynesian modelling has unemployment as a major element, and incorporates the unemployed as a significant group within the population. This is not the case, however, and one looks in vain for any mention of the unemployed in the conventional formulation of macroeconomics.

The discussion below has a simple and limited objective : to introduce the unemployed into a Keynesian income-expenditure model, thereby highlighting their central role in national income adjustments. Section (2) describes the reasoning behind this, and sets out the basic model. Section (3) considers the impact of unemployment benefits, both directly and through job search effects. Section (5) discusses the long run position, and Section (6) concludes.

(2) A Simple Keynesian Model with Unemployment

Keynesian economics is largely a model of national income, explaining how the level of national income (and thus economic activity) is determined in the short run. At any given time total income must be equated with total expenditure, and any injections to or withdrawals from the circular flow of income must balance. In the simplest closed economy with no government, investment expenditure, I, is the only injection and savings, S, the only withdrawal, yielding an ex post requirement that I=S. Expenditure is usually divided between consumption and investment, with the latter assumed to be determined exogenously by businessmen and inherently volatile in nature. Consumption, C, is systematically related to national income, Y, in such a way that $^{\partial C}/_{\partial Y} < 1$; their relationship can be summarised by the linear consumption function, C = a+bY, where b(<1) is the marginal propensity to consume. Because I is fixed exogenously the equality of income and expenditure is brought about by adjustment of the other two linked variables, Y and C. The system is as follows:

$$Y = C + S \equiv C + I$$

$$C = a + bY$$

$$\Rightarrow Y = \frac{a+I}{1-b}$$

National income settles at the level above, implying that any rise in I has an effect on Y that is magnified by the 'multiplier' $^{1}/_{1-b}$ (>1). There is no reason to believe that Y represents full employment, which is assumed to occur at some predetermined income level.

The derivation of national income outlined in preceding paragraph rests on two main foundations; the notion of the consumption function and the idea that national income adjusts to equate savings and investment. Consider firstly the consumption function. As utilised in macroeconomic analysis this is defined in aggregate terms, relating national income to total consumption expenditure in the economy. The same is not true of the reasoning behind the function, which is usually expressed in terms of individual behaviour. Keynes's preferred functional form (derived in Chapter 8 of the 'General Theory') depends on a combination of introspection and casual observation, and is not formally related to any theory of choice behaviour; the resulting 'psychological law' states that consumption rises with income, but in a lesser proportion. This seems to be a reasonably accurate view of reality based on plausible premises. To move to an aggregate formulation it is necessary to take account of the distribution of income; Keynes was aware of this, but assumed that the aggregate curve would display characteristics similar to the individual curves. For a given income distribution this should generally be true. When employment is varying, however, the income distribution does not remain static.

The other main feature of the Keynesian model is variations in national income. Unlike expenditure, income is not disaggregated in Keynes's formulation of macroeconomics and is depicted as the single variable, Y. When Y changes,the incomes of different members of the population must also be changing, mostly in the same direction. One is prompted to ask how this happens. Whose income changes when the economy faces cyclical booms and recessions? It is easier to start by identifying whose income is not changing. A wage earner who remains employed through a recession does not face a major fluctuation in income; life just continues as normal, doing the same job for the same wage. Simple Keynesian models usually assume constant prices and money wages, but even if these are moving they are likely to do so together, in such a way that real wages do not show much short run variation. This fact is a familiar property of recessions. Those who kept their jobs in the nineteen-thirties did not face falling real incomes or living standards; average real wages rose slightly, and there was little sign of hardship among the employed (Mowat (1955), Chapter 9). Casual observation suggests that the same is true in the current recession, and that the majority of the employed have not faced significant changes in living standards. The experience of the unemployed is strikingly different. A person laid off during a recession loses all wage income and becomes dependent on social security receipts. The fall in income sustained depends on the generosity of unemployment benefits, as summarised in the average replacement ratio; at current and historical U.K. benefit levels the income loss is substantial. Statistics and empirical studies have generally confirmed that unemployment significantly increases income inequality: see, for example, Lydall (1959), Blinder and Esaki (1978), and Gramlich (1974). The fortunes of the working population are therefore sharply divided according to employment status, and income adjustments are concentrated among the jobless. Since a large majority of the population is dependent on wage income, this division defines the position of most people during a recession; it also accounts for the largest part (albeit a smaller proportion) of national income. The other main income source is profits, which can also be expected to vary over the business cycle. Keynes said little about profits, and they do not appear in conventional macroeconomic models. More attention is given to them in Kalecki's version of macroeconomics (Kalecki

(1971)), where they account for a roughly fixed proportion of national income, determined by the 'degree of monopoly'. Other things equal profit income can be expected to move procyclically, increasing in a boom and falling in a recession. One is left with a well defined picture of how cyclical variations in national income affect individuals. Those who remain employed carry on more or less as normal, and adjustments in income are shared between the unemployed and recipients of profit.

The divergence in experience is concealed by conventional Keynesian theory, which is defined in terms of the aggregated variable, Y. By using Y in a consumption function based on notions of individual behaviour,the implication is that it accurately reflects the income changes faced by individuals. In other words, the distribution of income is assumed to remain unchanged, so that all individuals face equivalent income adjustments and the aggregate consumption function is defined by similar behavioural parameters to individual consumption functions. This is a long way from reality. Changes in national income are closely linked to predictable variations in the distribution of income and to movements in the employment level. There is no such thing as a general fall in incomes; national income falls through the income losses of those becoming unemployed and through reductions in profit. It follows that fluctuations in national income are inextricably bound up with changes in the employment level, and that the two should be considered together. Employment is, in fact, the agency by which movements in national income are accomplished.

In order to allow for the points made above, employment changes must be integrated with the usual Keynesian analysis of national income determination. This can be done quite straightforwardly, as below.

(i) Basic Model

In the following discussion national income varies through changes in employment income, which in turn adjusts by changes in the number of people employed. Profit incomes are ignored, although introducing them would alter little if they are a constant proportion of income, on Kaleckian lines. The fixed working population of L individuals is divided between the employed receiving wage income and the unemployed receiving social security payments. For simplicity's sake these are taken to be homogeneous groups, represented by an 'average' wage and a uniform unemployment benefit. The wage rate does not have to be identical for all the employed but unemployment has to affect all individuals with the same probability, regardless of their wage. In that case the wage distribution is independent of the employment rate, and employment changes can be discussed in terms of the mean wage in the distribution. The implications of dropping this assumption on wages are considered below. Two steps are required to set up the model, looking firstly at the composition of aggregate income and then at the composition of aggregate expenditure.

The initial step is to set out the relationship between employment and the level of national income. Under the assumption in the preceding paragraph this takes its simplest possible form. Let E denote the employment level (so that L-E individuals are unemployed) and M the average gross income in the invariant wage distribution. National income is then the sum of employment incomes in the population. In other words

$$Y = EM$$

where Y is national income. Income receipts by the unemployed are counted as transfer payments and do not appear in the national income assessment. For a given M, aggregate income can thus be treated as directly proportional to employment.

It remains to specify the nature of aggregate expenditure. Following Keynes, individual consumption is based on the 'psychological law' that expenditure rises with income but at a lesser rate. Consumption adheres to the pattern

$$C = a + bD$$

for all individuals, where $b(<1)$ is the marginal propensity to consume, a is a positive constant, and D is the individual's disposable income. If the employed are taxed at the rate t, then their mean disposable income is given by $(1-t)M$. The $L-E$ unemployed individuals receive a uniform social security payment of $\tilde{M}$, expressible as ϕM, where $\phi(<1)$ is the average (gross income) replacement ratio. To assume a broadly uniform benefit is a reasonable approximation to reality, although $\tilde{M}$ can alternatively be interpreted as the mean of a range of benefit payments. This makes no difference to the model's outcome, and the fact that ϕ is less than unity is beyond any doubt. Total consumption expenditure is the sum of the expenditure of the employed and unemployed, so that

$$\begin{aligned} C &= E(a+b(1-t)M) + (L-E)(a+b\tilde{M}) \\ &= E(a+b(1-t)M) + (L-E)(a+b\phi M) \\ &= L(a+b\phi M) + (1-t-\phi)bEM \end{aligned}$$

This relation is equivalent to the aggregate consumption function in the usual formulation, writing EM as Y. It differs in one important respect, namely that aggregate consumption depends explicitly on employment as well as on the level of disposable incomes. It is interesting in this regard to note Keynes's comment on p.90 of the 'General Theory', to the effect that the consumption function should really be defined in terms of employment.

This was not pursued further because Keynes felt that the distribution of employment between industries could be assumed approximately constant, which implies a unique relation between Y and E. The same assumption is made here, since M does not alter with E, but there is also the presence of the unemployed, who enter as another 'industry' into aggregate expenditure. Unlike the situation within the employed population, the distribution of expenditure between the employed and unemployed cannot reasonably be assumed invariant, hence the need for the revised model suggested here. Other expenditures are divided between investment, I, and government spending, G; the economy is assumed to be closed to foreign trade to simplify the model. In the usual manner these expenditures are autonomous, with investment as the most volatile component of aggregate demand.

The national income level is derived as in the standard Keynesian model. At any time aggregate income must be equated with aggregate expenditure, and their identity is achieved through variation in the income level. In this case, however, income variations are specifically associated with changes in the employment rate. The full model and solution are as below:

Income $\qquad Y = EM$

Expenditure $\qquad X = C + I + G$

$$= L(a+b\phi M) + (1-t-\phi)bEM + I + G$$

L, M, ϕ and t are assumed constant in the period considered.

Income = Expenditure

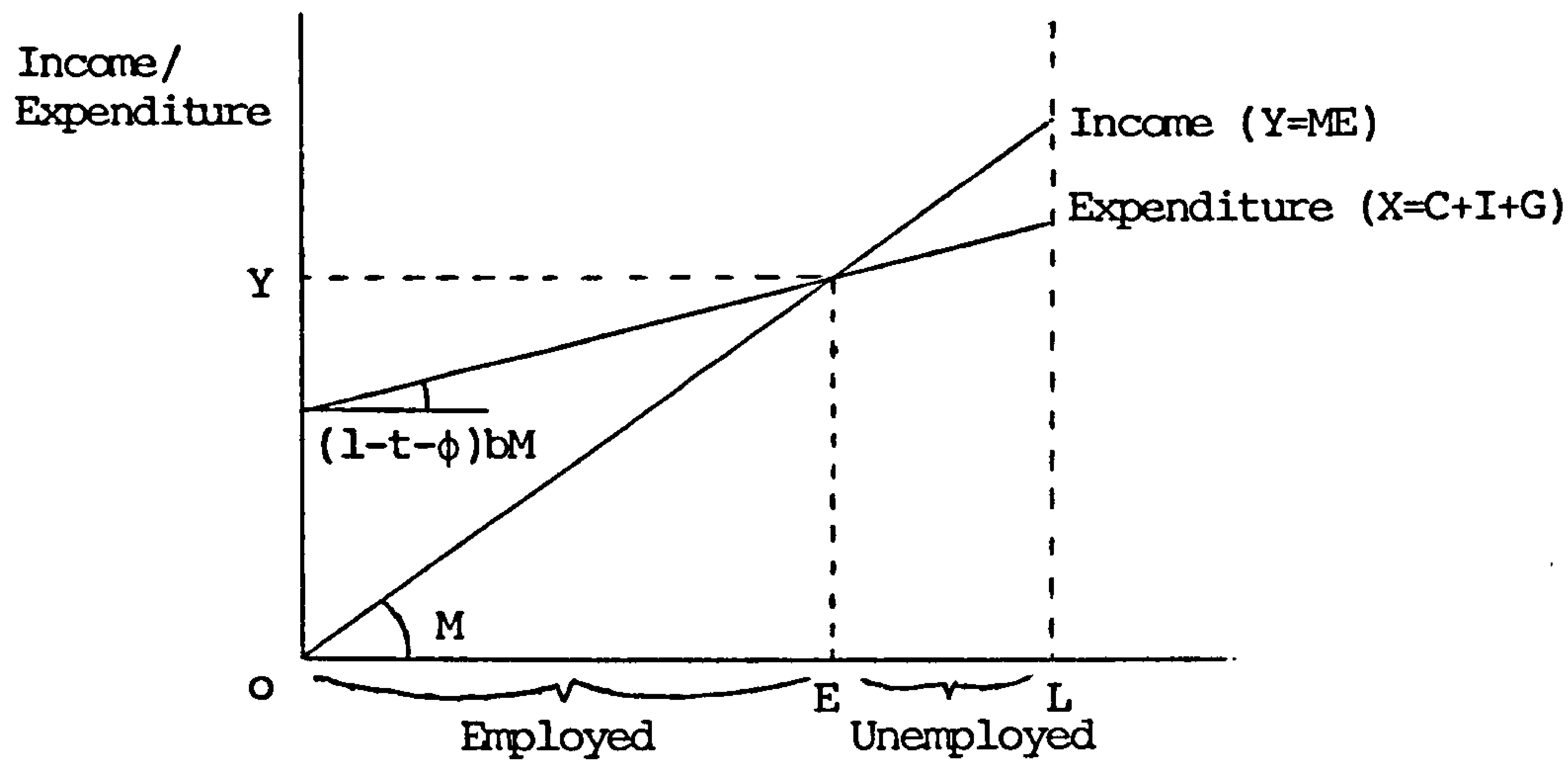
$$\Rightarrow \quad Y = X = C + I + G$$

$$\Rightarrow \quad EM = L(a+b\phi M) + (1-t-\phi)bEM + I + G$$

Solving for Y and E yields

$$Y = \frac{L(a+b\phi M)+I+G}{1-(1-t-\phi)b} \quad ; \quad E = \frac{L(a+b\phi M)+I+G}{M(1-(1-t-\phi)b)}$$

Y satisfies an equation similar to that in conventional models; in this case E is also defined by the model, such that $E = {}^{Y}/_{M}$. The outcome can be illustrated diagrammatically as below:



Instead of plotting income against itself and using the 45^{o} line as an income curve, it is now possible to plot income directly against employment. The income curve shows the positive relation between income and employment, and has a slope equal to the mean employment income. The expenditure function is akin to the standard consumption function, with a slope dependent on the marginal propensity to consume. Y and E are determined at the intersection of the two curves; in the usual Keynesian fashion there is no guarantee that E is at the full employment level, giving rise to unemployment L-E on the bottom axis.

An increase in either I or G shifts the expenditure curve upwards by the amount involved. Y then rises in greater proportion, subject to the multiplier relation

$$\frac{\partial Y}{\partial I} = \frac{1}{1-(1-t-\phi)b}$$

The employment multiplier can be defined as M $\frac{\partial E}{\partial I}$, relating the number of jobs implicit in the initial investment expenditure to the total number of jobs created. This takes the same value as the income multiplier, owing to the linear relationship between Y and E.

In other respects the model reproduces the conclusions of the simple Keynesian analysis. A rise in G has the same impact as a rise in I, implying that increased government expenditure is a possible way of preventing chronic cyclical unemployment. Reductions in t also have an expansionary effect, and can be seen as a potential policy tool. The impact of ϕ is discussed in Section (3).

The diagram above resembles the aggregate supply/demand analysis of Weintraub (1958), Davidson (1962) and Davidson and Smolensky (1964), although its interpretation is rather different. In particular, the income curve is not tied to any behavioural view of the labour market, and merely represents the variation of income through employment changes. One would not expect this relationship always to be linear in reality, but the exact nature of its curvature is not immediately apparent. Two main alternative views can be put forward:

(a) Income and Expenditure Curves Concave

In practice job losses are unlikely to be representative of average wages in the economy. Casual empiricism indicates that most redundancies occur among low paid, unskilled workers, whereas senior positions are relatively more secure. This view is formalised in hierarchical or organisational theories of the firm (Simon (1957), Williamson (1975)), which see both remuneration and job security increasing as the summit of the hierarchy is approached. Similar conclusions are implicit in dual labour market theories (Gordon (1972) , Doeringer and Piore (1971)), which contrast the secure jobs in the 'internal' labour market with unskilled ones in the secondary labour market. Alternatively, one can appeal to purely neoclassical theory, based on diminishing marginal productivity in technical production. Either way round the prediction is that average wages and incomes increase as national income falls, because of the loss of below average incomes. The result is a 'batting average' effect: removing the tail-enders increases the team's batting average but decreases its aggregate score. Such arguments were used by Keynes to suggest that real wages move counter-cyclically (Keynes (1939)). Keynes's analysis is based on marginal productivity considerations, although similar conclusions can also follow from alternative approaches to the labour market.

(b) Income and Expenditure Curves Convex

Empirical work does not always confirm that real wages move counter-cyclically, as was pointed out soon after the 'General Theory' was published (Dunlop (1938), Tarshis (1939)). The effects in (a), while still present, may be outweighed by other factors working in the opposite direction. The chief possibility here is the impact of the business cycle on wage bargaining. If a recession impairs the bargaining power of employees, it is possible that real wages and earnings tend to fall in a recession and rise in a boom.

Cases (a) and (b) merely alter the curvature of the income and expenditure curves, without changing the basic format of the model. In (a) the employment multiplier exceeds the income multiplier, and vice versa in (b). The central point remains the interdependence of income and employment adjustments : this can be generalised to allow for alternative arrangements, as in (ii) below.

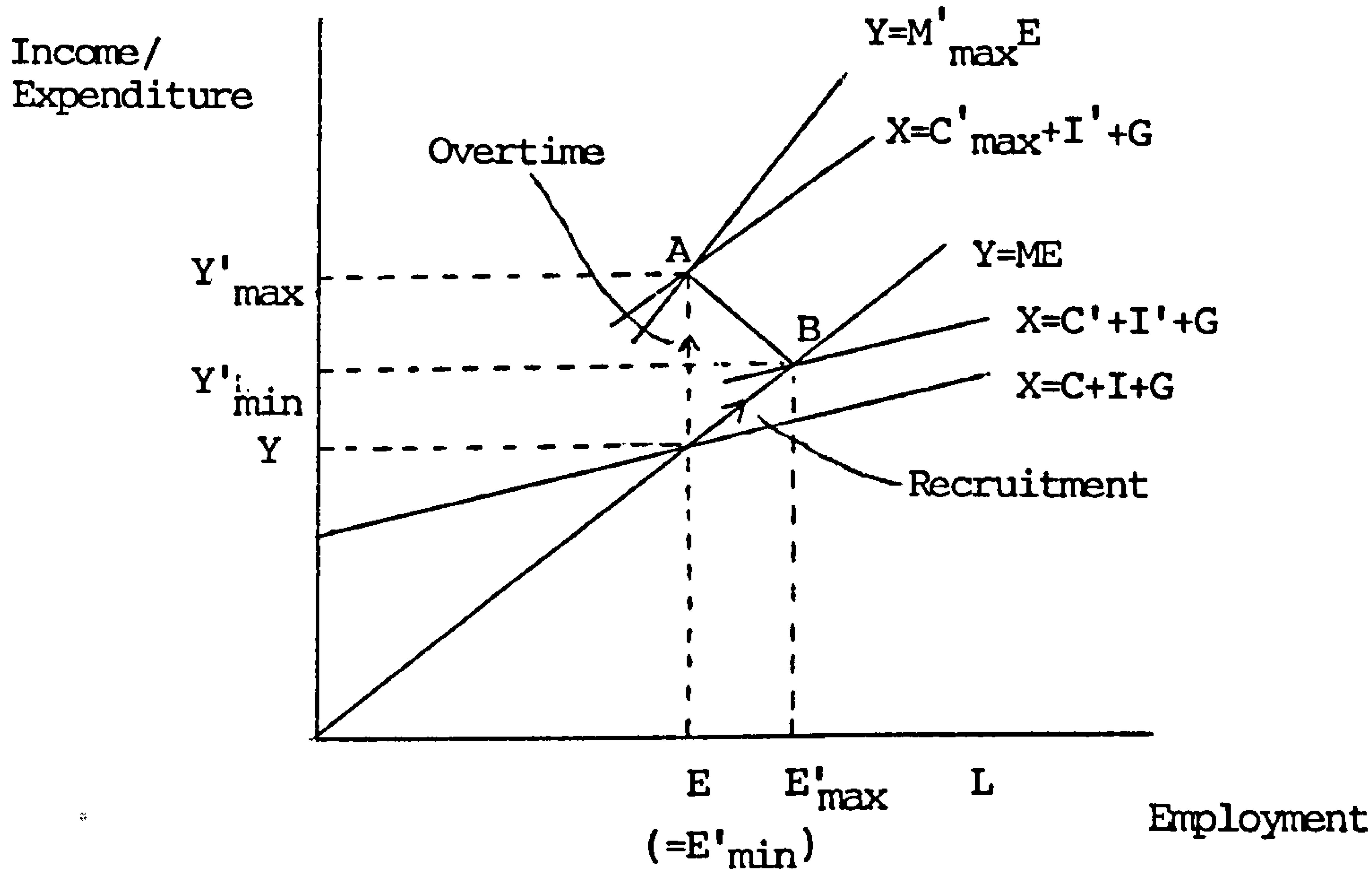
(ii) Alternative Forms of Employment Adjustment

Employment changes are not necessarily accomplished by recruitment and redundancies. For a small change in output firms may prefer to alter the working hours of their existing labour force, by overtime or short-time working. This often true, for example, when labour has firm-specific skills involving prior training. Newly recruited labour then has to acquire these skills, implying the presence of a fixed cost discouraging variation in the level of employment. In the terminology of labour economics, labour becomes a 'quasi-fixed factor' (Oi (1962)), with some of the characteristics of investment rather than consumption. Whatever the reason behind it, a reluctance of firms to vary employment must influence the relation between employment and national income. The model below makes no assumptions as to why some firms do not vary their labour force, and is compatible with a range of interpretations.

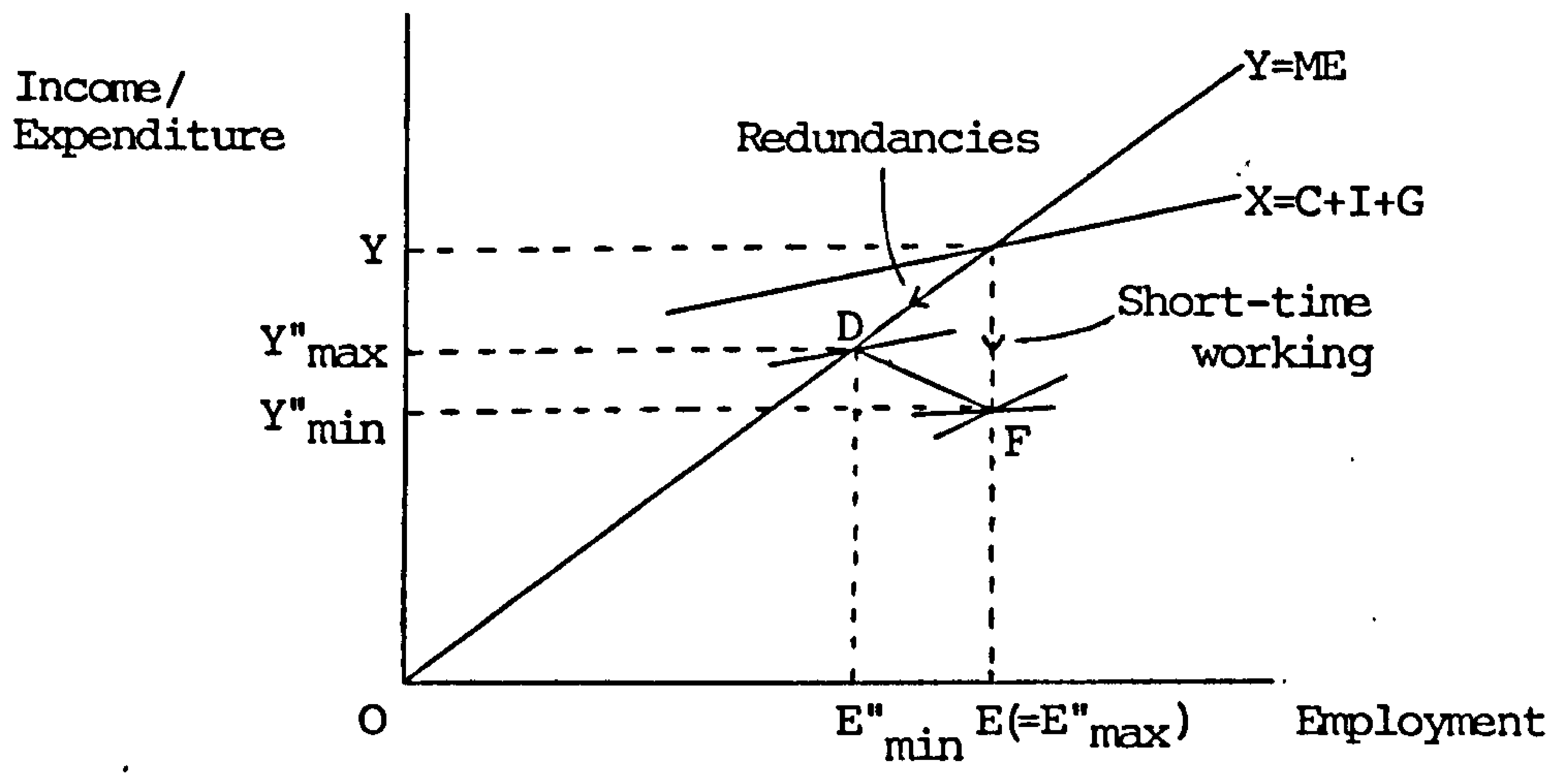
Any adjustment of working hours means that M is no longer constant, and that Y is not uniquely related to E. There are several possible combinations of M and E which satisfy the income-expenditure identity. To show this diagrammatically, suppose that unemployment benefits remain fixed at $\tilde{M}$. If investment rises from I to I', the set of potential end points is given by

$$Y' = \frac{L(a+b\tilde{M})+I'+G}{1-(1-t)b} - \frac{b\tilde{M}}{1-(1-t)b}E' \qquad \begin{array}{l} M' \geq M \\ E' \geq E \end{array}$$

The relation between Y' and E' is obtained by holding E' constant, and allowing movements in M to equate income and expenditure; the inequalities impose the assumption that a rise in expenditure reduces neither M nor E. Hence there is a negative linear relationship between Y' and E', bounded by the extremes of constant E or M. Diagrammatically one obtains



where the line segment AB is the set of possible outcomes. Adjustment by overtime alone causes a vertical movement to point A, with no change in E and the maximum possible increase in Y. Conversely, adjustment by recruitment alone produces a shift along the income curve to B, giving a lesser increase in Y and the largest possible rise in E. Between these cases any point on the line AB can be the outcome of the rise in I, depending on how employment changes occur. The more adjustment that takes place through recruitment, the closer the end result will be to point B, and vice versa. The final values Y' and E' can fall anywhere in the ranges $\{Y'_{max}, Y'_{min}\}$ and $\{E'max, E'min\}$ above, and there is no unique outcome until the nature of employment changes is specified. Much the same applies for a fall in investment, which is illustrated below:



The set of possible outcomes is bounded by points D and F, corresponding to adjustment by redundancies and short-time working respectively. National income and employment can settle anywhere in the Y" and E" ranges given in the diagram. The result again depends on how the expenditure shift is accommodated in terms of employment changes. Adjustment is not necessarily symmetrical, so a rise in income may be accomplished differently from a reduction.

The model above is a slightly rearranged version of the 'Keynesian cross'. Traditional Keynesian models depend on a partial disaggregation of expenditure (into consumption, investment, etc.) but leave income as a single parameter, Y. Such an approach implies that the income distribution is stable and independent of the general level of economic activity. Post-Keynesian critics of conventional theory stress the importance of the income distribution through arguments about differential propensities to consume (Kregel (1973), Eichner (1979)). Their case is strengthened when unemployment is taken into account. It is possible to imagine an income distribution between wages and profit that is stable over the business cycle (an assumption often made). It is not possible for the distribution of income between the employed and unemployed to remain constant as unemployment varies. In this case the distribution of income and employment is built into the workings of the system and cannot reasonably be neglected. Conventional formulations of the consumption function have to allow for this, even if they do not mention it explicitly. It seems preferable to have a genuine income curve to set alongside the expenditure relation, instead of the device of a 45^{o} line : with the luxury of a spare axis it is sensible to use it. The outcome helps to emphasise the roots of macroeconomic analysis in the national income accounting identities.

(3) The Role of Unemployment Benefits

The model of Section (2) has the virtue of including the replacement ratio, ϕ, explicitly in expressions for the multiplier. This rectifies the imbalance with the taxation parameter, t, which is seen much more frequently in multiplier expressions; it also emphasises the central position of unemployment benefits, instead of lumping them together with other (less variable) transfer payments. The discussion below considers the macroeconomic impact of unemployment benefits, both directly and through possible effects on job search.

(i) Direct Impact of Varying the Replacement Ratio

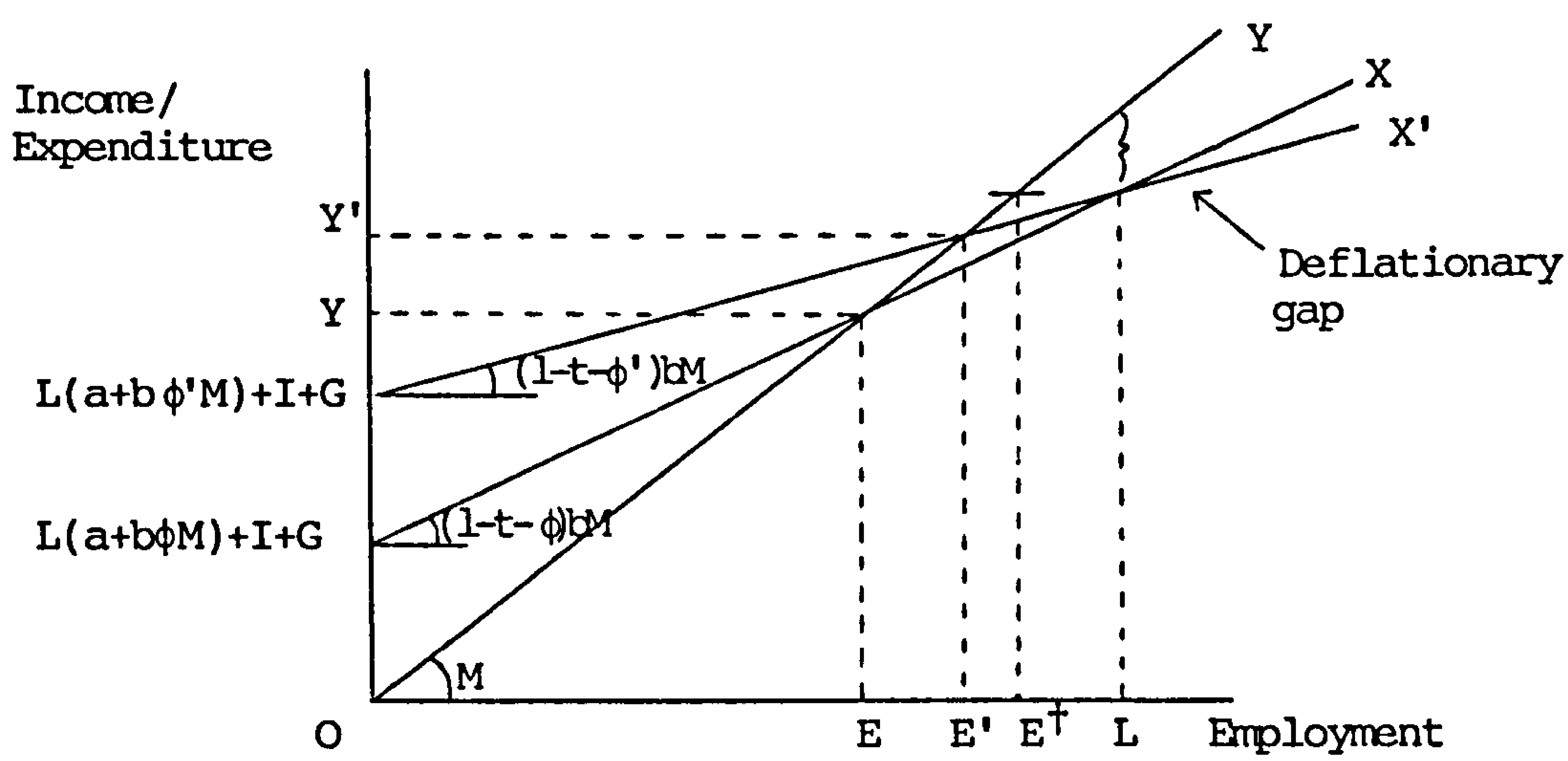
Suppose that income variations occur mainly through changes in employment, as in the basic model of Section (2). A change in unemployment benefits is described by a variation in the replacement ratio ϕ, ceteris paribus, so that the effect on Y and E is given by the derivatives of their steady state values with respect to ϕ, that is

$$\frac{\partial Y}{\partial \phi} = M\frac{\partial E}{\partial \phi} = \frac{b(ML(1-(1-t)b)-aL-I-G)}{(1-(1-t-\phi)b)^2}$$

The sign of the derivatives depends on the bracketed expression in the numerator, which can be rearranged as

$$ML - L(a+(1-t)bM) - I - G = \text{Deflationary Gap} > 0$$

Since ML is total income at the full employment position (E=L) and L(a+(1-t)bM) + I + G is total expenditure, the expression is equal to the deflationary gap at employment level E. This must always be positive, since unemployment benefits are only received when E < L : hence raising benefits is always expansionary. The size of the impact depends on the size of the deflationary gap and the number of unemployed, diminishing as full employment is approached. On the income-employment diagram the position is as below:



The replacement ratio is raised from ϕ to ϕ', producing a new expenditure curve X' with a higher intercept and lower slope than X. Raising benefits therefore produces a higher 'base' expenditure of the unemployed, but reduces the responsiveness of total expenditure to changes in employment. The expenditure curves for different ϕ values always intersect where E=L, so varying ϕ has the effect of rotating the expenditure curve about this point. It can be seen that a lower slope for the X curve always increases Y and E, with the rate of increase diminishing as the curve approaches the horizontal. The upper limit to the replacement ratio can be viewed as $\phi^\dagger = 1-t$, at which point the net average incomes of the employed and unemployed are equated and there is no monetary incentive to work. When $\phi^\dagger$ prevails the corresponding $X^\dagger$ curve is horizontal and employment attains a level of $E^\dagger$, the maximum obtainable from raising the expenditures of the unemployed. This shows that although raising ϕ can increase E, it cannot on its own restore full employment. The reason is that shifts in unemployment benefits can never remove the deflationary gap present at E=L. $E^\dagger$ depends directly on the size of the gap, so the expansionary impact of unemployment benefits is circumscribed by the general demand deficiency in the economy.

As well as directly influencing employment, the replacement ratio also affects the economy's stability in the face of shifts in exogenous expenditures. Unemployment benefits are often seen as an 'automatic stabiliser', given that a higher ϕ reduces the multipler. Such a role encourages relatively high benefits, and gains in importance as autonomous expenditures become more volatile.

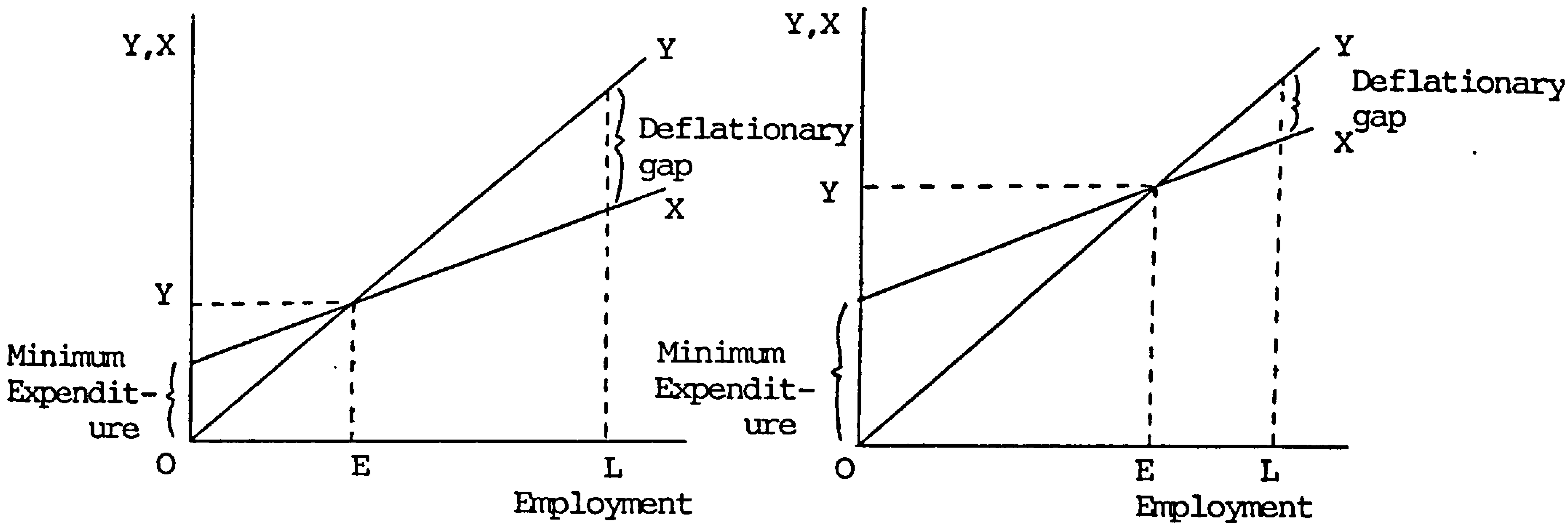
The interaction between unemployment benefits and taxes is less straightforward. Variation in t influences Y in the manner below

$$\frac{\partial Y}{\partial t} = M\frac{\partial E}{\partial t} = \frac{-b[L(a+b\phi M)+I+G]}{(1-(1-t-\phi)b)^2} < 0$$

As ϕ changes, the effect is such that

$$\frac{\partial^2 Y}{\partial t\partial\phi} = M\frac{\partial^2 E}{\partial t\,\partial\phi} = \frac{-b^2[(ML-L(a+(1-t)bM)-I-G) - (L(a+b\phi M)+I+G)]}{(1-(1-t-\phi)b)^3}$$

The sign depends on the square-bracketed expression whose first term is the deflationary gap at E=L and second term is expenditure at E=O. Raising ϕ therefore enhances the impact of tax changes when the deflationary gap exceeds minimum expenditure, and vice versa. The position is as below



ϕ increases effect of t

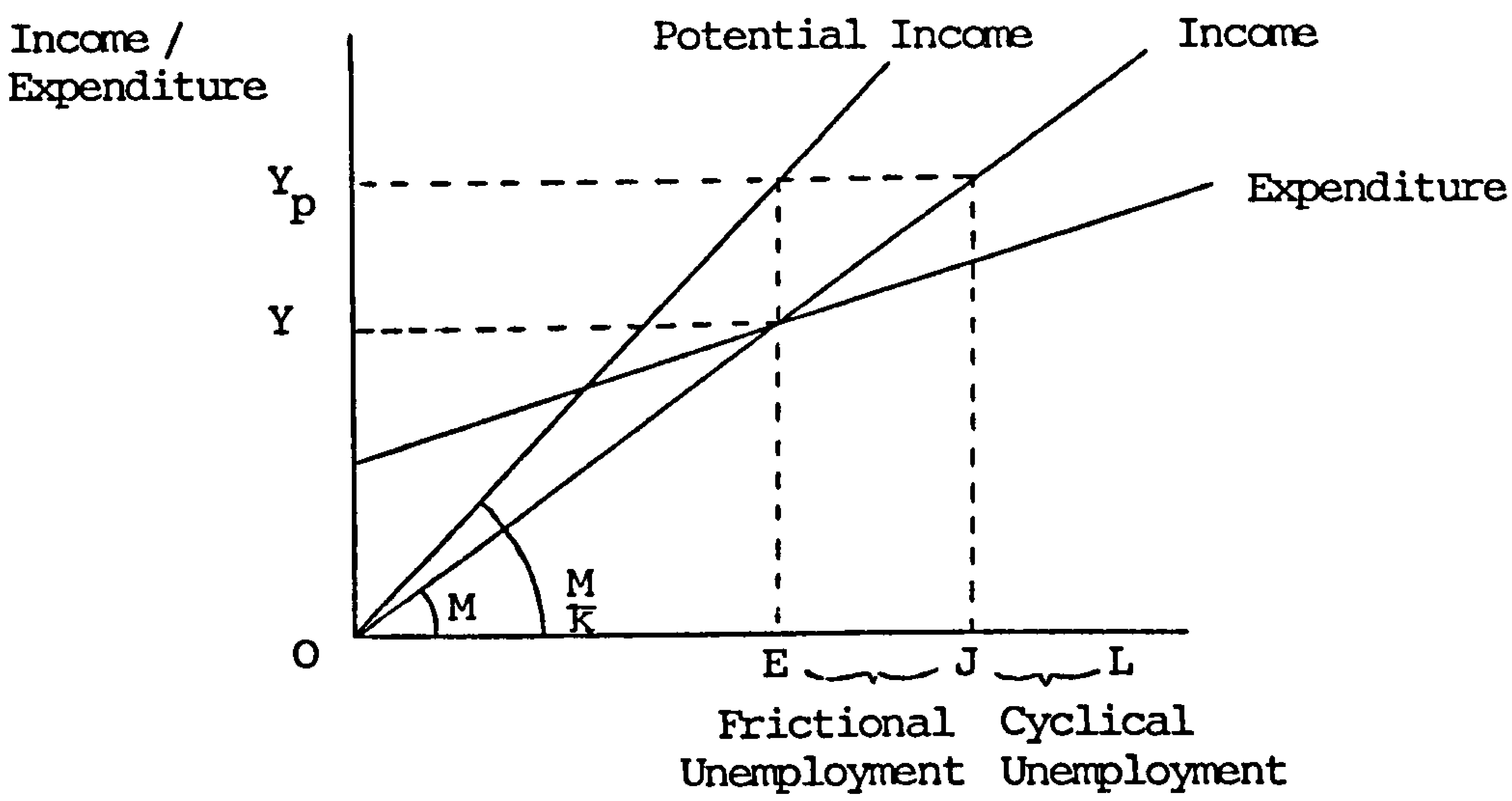
ϕ decreases effect of t

In general the left-hand case can be taken to represent lower income and expenditure levels, so that ϕ starts by strengthening the impact of taxation but eventually comes to weaken it as Y and E increase. If the economy is fairly close to full employment, ϕ has a stabilising effect on taxation changes, as well as on autonomous expenditure movements. This is a good thing where taxes are aimed at raising revenue with a minimum possible disturbance to the national income level. It does reduce the impact of taxes as a means of altering national income, although the effects of autonomous expenditure variations are scaled down in a similar way.

<u>(ii) The Impact of Job Search</u>

Much discussion of unemployment benefits concerns their possible adverse effect on job allocation and frictional unemployment. The issue is considered below within an explicitly macroeconomic framework.

To introduce frictional unemployment into the model, a distinction between employment and job availability is required. If job vacancies are high relative to the number of unemployed, then unemployment can be classified as mainly frictional : the number of jobs in existence is apparently sufficient to accommodate most of those seeking employment. In such circumstances a superior process of job allocation proffers hope of alleviating unemployment. These considerations can be allowed for by defining the total number of jobs in existence, J, as distinct from the level of employment, E. When E<J there exist J-E unfilled vacancies, which can be interpreted as the extent of frictional unemployment. Let the parameter k denote the ratio of employment to jobs, so that E=kJ and k indicates the importance of frictional unemployment. For each level of E and J it is possible to define 'potential' national income, Y_p, as $Y_p = MJ = (^M/_k)E$; this is the income level that would ensue if all vacancies were filled and frictional unemployment were reduced to zero. The situation can be depicted as below:



As usual, Y and E are determined by the intersection of the income and expenditure curves. The extent of frictional unemployment is summarised by k, producing a potential income curve with slope $^M/_k$ (steeper than the income curve, since k<1). At employment level E, Y_p can be read off the potential income curve, and the number of jobs is given by the intersection of $Y=Y_p$ with the actual income curve. On the lines stated above, the extent of frictional unemployment is then equal to J-E, while cyclical unemployment accounts for the remaining L-J. A fall in k leads to a higher slope for the Y_p curve, raising the values of Y_p and J and increasing frictional unemployment relative to cyclical.

The disincentive effects of unemployment benefits can be represented by variations in the value of k. If the replacement ratio is high, the unemployed have a reduced incentive to accept employment, implying that some vacancies remain unfilled and k takes a relatively low value. Conversely, if the replacement ratio is low, most vacancies will be filled and k will be close to unity. These effects can be represented by making k a decreasing function of ϕ, that is, $k=k(\phi)$, $k'(\phi)<0$. From observing the $E=k(\phi)J$ relation, this would seem to suggest that a direct disincentive effect of ϕ tends to raise the unemployment rate. But before reaching any conclusions one has to consider the macroeconomic position.

Two cases can be distinguished, depending on what is happening to aggregate expenditure:

(a) Direct Filling of Vacancies

It is here assumed that everything else is held constant as changes in ϕ and k occur. To allow for macroeconomic effects the level of potential expenditure as well as potential income must be included; as vacancies are filled by reducing frictional unemployment, the economy finishes at a point where potential income and expenditure are equated. The situation is such that

Potential Income $$Y_p = MJ = \frac{M}{k(\phi)} \cdot E$$

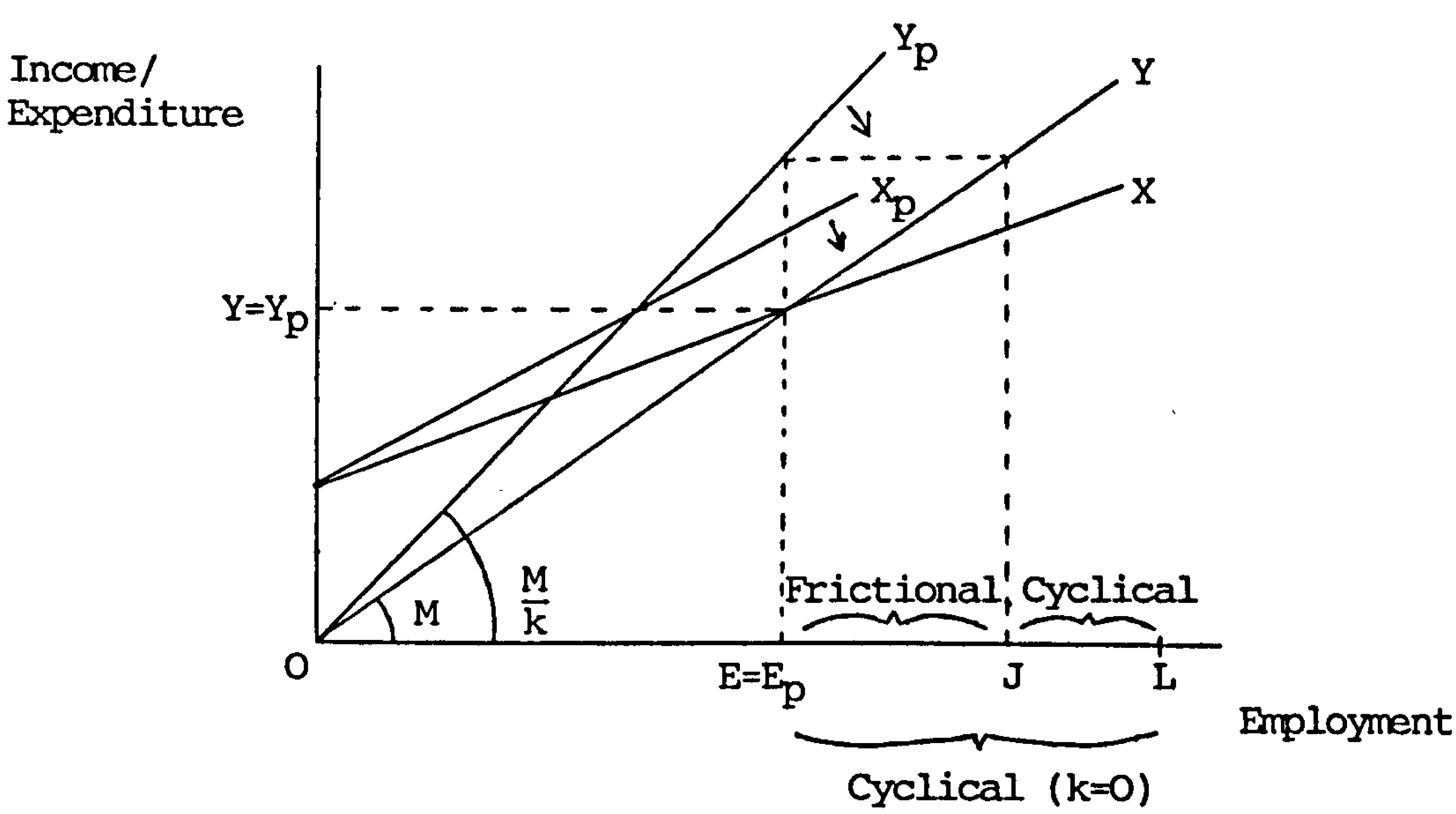
Potential Expenditure $$X_p = L(a+b\phi M) + (1-t-\phi)bJM + I + G$$

$$= L(a+b\phi M) + (1-t-\phi)b\frac{M}{k(\phi)} \cdot E + I + G$$

$$Y_p = X_p \quad \Rightarrow \quad Y_p = \frac{L(a+b\phi M)+I+G}{1-(1-t-\phi)b} = Y$$

$$E_p = \frac{Y_p}{M} = \frac{Y}{M} = E$$

Hence, when aggregate demand effects are allowed for, the potential income obtainable from filling vacancies is identical to the current income level, and the same goes for employment. The outcome is entirely independent of k, so there is nothing to be gained by improving job allocation by reducing k. A lower ϕ would have a direct deflationary effect on Y, but would produce no net gain from any increase in k. The only effect of raising k is to influence the classification of the unemployed towards cyclical rather than frictional, while leaving the total number unchanged. Diagrammatically the position is as below:



It is convenient here to consider changes in k on their own, without the associated change in ϕ. At employment level E the potential income from filling vacancies exceeds the potential expenditure. The 'potential' of $Y_p = {}^{Y}/_{k}$, E=J cannot be fulfilled, and raising k has a deflationary impact on income and employment. As k increases the Y_p and X_p curves rotate downwards, always crossing on the $Y=Y_p$ line so that total employment never varies from E. In the limit, when k=1, the curves coincide with the Y and X curves, and the removal of all frictional unemployment has left total unemployment unchanged. Thus, the positive impact of reducing k for given demand conditions is exactly offset by the negative macroeconomic effects. All that is achieved is that the value of J converges to E, reducing the number of unemployed classified as frictional. Although the J initially vacant jobs appear to be a potential net gain, the process of filling them, ceteris paribus, requires an equivalent number of jobs to be lost.

This demonstrates a point often overlooked in discussions of unemployment: filling job vacancies is deflationary. Economic models based on search disincentives tend to assume that an improvement in job allocation must reduce total unemployment. As far as frictional unemployment goes that should always be true, but an adverse impact on cyclical unemployment can also be anticipated. The grounds for this are intuitively clear. When a person takes a vacant job, recorded national income rises by appropriate discrete amount; a parallel shift occurs in national expenditure, but to a lesser degree because of the incidence of savings and taxes and because the person was earlier receiving unemployment benefits. If repeated on a large scale, this will lead to national income exceeding expenditure and consequently to deflationary pressure. In other words new employment incomes have to be received as some type of expenditure, and the spending of particular workers is not enough to justify their own jobs. The implication is not that filling jobs directly causes unemployment - causality need not be interpreted that way. But one can say that when jobs are being filled there must be some accommodating changes if income and expenditure are to balance. For example, some other people may be losing their jobs at the same time to counteract the effect, or there may be increases in autonomous expenditures that serve to maintain the expenditure level. The latter is the best way out of the impasse, and is considered as the second main possibility.

(b) Vacancies 'Justified' by Autonomous Expenditure Changes

In view of the outcome of (a), it is relevant to ask how vacancies arise in the first place. The key question is whether or not they are linked with changes in autonomous expenditure that permit net increases in employment. It is quite easy to imagine such changes happening. Employers may be wishing to undertake investment projects, but are unable to find suitable labour. This would lead to vacancies which remain unfilled and appear as frictional unemployment. An improvement in job allocation might then enable the investment to go ahead, so that filling vacancies automatically releases autonomous expenditures which were waiting to be made. In this situation a filled vacancy is 'justified' by investment, and reflects planned future spending at least equal to the income attached to the job (and probably exceeding it). Not all vacancies can be placed in this category (as, for example, when people are simply changing jobs), but a certain proportion can probably be viewed as 'new' jobs backed by planned new expenditures.

Such effects can be included in the model of (a) by allowing investment to increase as vacancies are filled. Let investment now be written as $I(k(\phi))$, where $I'(k) < 0$. A reduction in k produces a rise in I, resulting from expenditures which are able to proceed when vacancies are filled. The strength of the relation is left indeterminate, and could be anything from virtually zero to quite a strong stimulus to demand. If most filled vacancies are already existing jobs, then the extent of new investment will be small, and k will have little impact on I. On the other hand, if most vacancies are tied to planned new investment, the associated rise in expenditure could easily exceed the incomes directly attached to the jobs; together with subsequent multiplier effects, this could produce a substantial expansion in demand. The nature of the model consequently depends on the particular circumstances in question, and the relation between I and k is in no way a general one. Bearing this in mind, analysis can proceed as in (a). Equating potential income and expenditure gives an outcome as below:

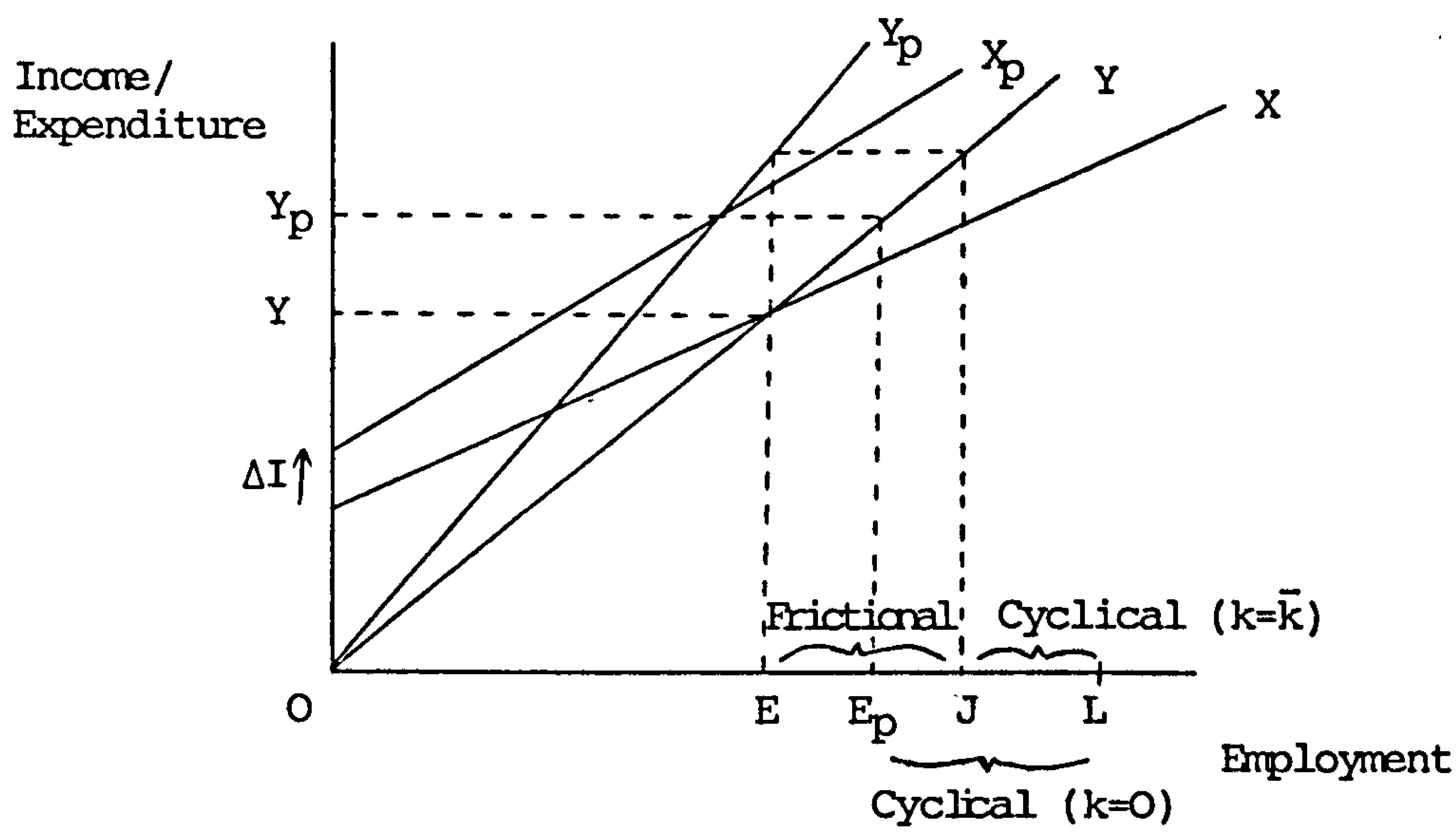
Potential Income $\quad Y_p = MJ = \frac{M}{k(\phi)} E$

Potential Expenditure $\quad X_p = L(a+b\phi M) + (1-t-\phi)bJM + I(k(\phi)) + G$

$$= L(a+b\phi M) + (1-t-\phi)b\frac{M}{k(\phi)}E + I(k(\phi)) + G$$

$$Y_p = X_p \quad \Rightarrow \quad Y_p = \frac{L(a+b\phi M) + I(k(\phi)) + G}{1-(1-t-\phi)b}, \quad E_p = \frac{Y_p}{M}$$

The important difference from (a) is that Y_p and E_p now depend on k, so that a rise in k can produce a genuine increase in employment. Diagrammatically the situation is as below:



The Y,X and Y_p curves have similar interpretations to the previous case; the only difference is that a reduction in k now entails a vertical shift in the X_p curve arising from changes in I. This means that the potential expenditure from filling current vacancies is increased by the amount of investment brought forth by a shift in k from its initial value, $k=\bar{k}$, to k=0. The X_p curve plotted therefore has the same slope as in (a), but is translated upwards by the additional investment resulting from removing all vacancies. Intersection with the Y_p curve occurs above the previous level, so that $Y_p > Y$ and $E_p > E$ now hold true. There is consequently a net income and employment gain to be made from improving job allocation, with Y_p and E_p

representing the maximum levels attainable. It is possible for E_p to be less than, equal to, or greater than J. In the case illustrated E_p is less than J, implying that the rise in expenditure is insufficient to offset fully the deflationary impact of filling job vacancies. A larger stimulus to investment could potentially raise E_p above J, so that the total employment return from increasing k is greater than the initial number of jobs observed to be available. E_p and J will be equated only by accident, and there is no particular reason to view them as being approximately the same thing. This perhaps runs counter to one's immediate impression, which sees declared vacancies as the employment 'lost' by failures of job allocation. In practice the elimination of vacancies is certain to be linked with changing demand conditions (one way or the other), producing a more fluid situation than is suggested by static job and vacancy figures.

An increase in k is in this model unambiguously desirable, raising both national income and employment. Measures that seek to influence k directly (improving information, etc.) are therefore potentially useful, provided they have no adverse effects, derived, say, from the way they are financed. The situation with the replacement ratio is less clear. If search disincentives are important a reduction in ϕ could serve to raise k and thereby produce an expansionary effect on national income and employment; at the same time, however, it also has a deflationary impact on the lines of (i) above when M is held constant (as well as reducing the living standards of the unemployed). A government wishing to decide on the best level of ϕ faces a trade-off between the positive distributional and demand effects of a high replacement ratio and its possible adverse impact on job allocation. This can be used to obtain a policy optimisation in the manner of Chapter 5 , albeit with a much different set of constraints and variables. For example, Y and E are maximised where

$$\frac{\partial Y}{\partial \phi} = \left.\frac{\partial Y}{\partial \phi}\right|_{k \text{ constant}} + \frac{\partial Y}{\partial k} \cdot \frac{\partial k}{\partial \phi} = 0$$

The direct effect of ϕ is positive and declines with ϕ, while the impact through k is negative and (possibly) increasing with k; it is consequently quite feasible for an interior solution to exist. Where a value is placed on equality between the employed and unemployed, the objective would differ from a simple maximisation of income and would encourage a higher level of ϕ. It should be said here that policy optimisation is not easily compatible with the macro model being used, since autonomous expenditures are subject to continuous change. There is no single 'optimal' value of ϕ which is the best possible choice in all situations. It remains true, however, that a decision has to be made on benefits, and that the type of considerations outlined here are relevant. The government has to decide on what seems the best 'average' value, bearing in mind the different effects that can be associated with benefits.

Under the conditions assumed here the arguments about an excessive replacement ratio increasing unemployment can potentially have some substance. To get such conclusions, however, it is necessary to link an improvement in job allocation with changes in autonomous expenditure. If unemployment is reduced by lowering ϕ, it is not the superior job allocation that is responsible for the change, but the associated rise in expenditures. There are consequently no grounds for the idea that simply filling jobs is a way to increasing employment; some other expansionary changes are required to guarantee that the level of national income is maintained. This emphasises why autonomous expenditures play such a central role in determining economic activity. Even views based on 'microeconomic' analysis of job allocation depend on increased autonomous expenditures if they are to hold water.

It should also be noted that a reduction in ϕ need not be identified with a reduction in benefits. A rise in real wages with constant benefits can have the same effect without reducing the living standards of the working population. This may even happen spontaneously in the conditions depicted here. Disincentives only bite when they are preventing autonomous expenditures from taking place, so that planned expansion is being obstructed. If firms are unable to recruit labour for investment projects, there may be upward pressure on real wages, while benefits remain fixed. The assumption made above of fixed employment incomes is not then appropriate and the need to reduce unemployment benefits is correspondingly reduced. Should a rise in real employment incomes fail to occur, it is still possible for the government to try to impose it, as is straightforward if disincentives apply to net rather than gross wages. Hence, even when disincentive arguments have some plausibility, they do not necessarily lead to a case for cutting unemployment benefits. An increase in real wages is an alternative response.

This section has considered the position of unemployment benefits in a simple macroeconomic model. Raising the replacement ratio is expansionary, providing a possible rationale for higher benefits when cyclical unemployment is present (although other means are available for raising demand). The importance of disincentives is less obvious in a macroeconomic than microeconomic approach. To be of any significance they have to be linked with expenditure changes; this can happen, but it puts a different complexion on the discussion. A valid disincentive has to be 'blocking' expenditure increases, so that investment or other injections are unable to take place. In this sense all unemployment is due to demand deficiency; there may be 'frictional' reasons for the deficiency, but a deficiency is necessary all the same.

(5) The Long Run Position

Discussion has so far concerned short run income fluctuations determined by shifts in aggregate expenditures. Over a longer period one would expect other forces to operate. In particular technical progress should lead to a rising long-term trend in real incomes, on which the short run income movements are centred. It is straightforward to depict this in terms of the models of Sections (2) and (3).

Technology has not played a central part of the consideration of short term income and employment determination above. This is in the nature of the model, rooted in the circular flow of national income defined in aggregated monetary terms. Income and expenditure are partially disaggregated in the course of macro analysis, but the model is never reduced to a discussion of physical quantities. Technical production, although always occurring, is not a major influence on cyclical fluctuations. The same cannot be said of the long run, where time-series statistics usually show a chronic rise in real national income and the average wages of the employed. These trends are normally explained by technical progress, which produces a parallel increase in all forms of income. This contrasts markedly with the selective impact of short run income variations. One cannot expect Bowley's Law to hold exactly, but long-run income movements should be more neutral distributively than short-run ones. The discussion below is based on two postulates:

(i) In the long run the effects of shifts in autonomous expenditure are dominated by the impact of technical change. It is therefore preferable to base the model on an exogenous real income trend rather than demand variations.

(ii) Long-run income changes are distributively relatively even. They are not closely associated with employment changes, and therefore do not incur the distributional consequences of unemployment.

Imposing these properties does not in any way change the model's basic nature. The intention is merely to show what will be observed in the long run under the conditions just described.

Let the model take the form of Section (2), with the same notation. Consider firstly the form of the traditional aggregate consumption function (relating aggregate consumption expenditure to aggregate income) in the short run. Consumption satisfies the equations

$$C = L(a+b\phi M) + (1-t-\phi)\, bEM$$

$$= L(a+b\phi M) + (1-t-\phi)\, bY$$

where the lower relation corresponds to the usual notion of the consumption function. This has the conventional form of a positive intercept and positive slope less than unity. It is not quite an inflated version of the individual function $C = a+bM$ because of the presence of taxes (lowering the slope) and unemployment benefits (lowering the slope and raising the intercept). Implicit in the form of the function is the fact that M remains approximately constant and that variations in Y are linked with variations in E. In the long run Y is on an upward trend arising mainly from an exogenous rise in real incomes, M, rather than shifts in E. It can be assumed that employment fluctuates about some non-zero long-run average, denoted by $\bar{E}$; this value will be treated as a constant, so that employment variation is ironed out over the period being considered. L, ϕ and t are also held constant. Under these conditions the long-run consumption function can be written as

$$C = L(a+b\phi M) + (1-t-\phi)\, bM\bar{E}$$

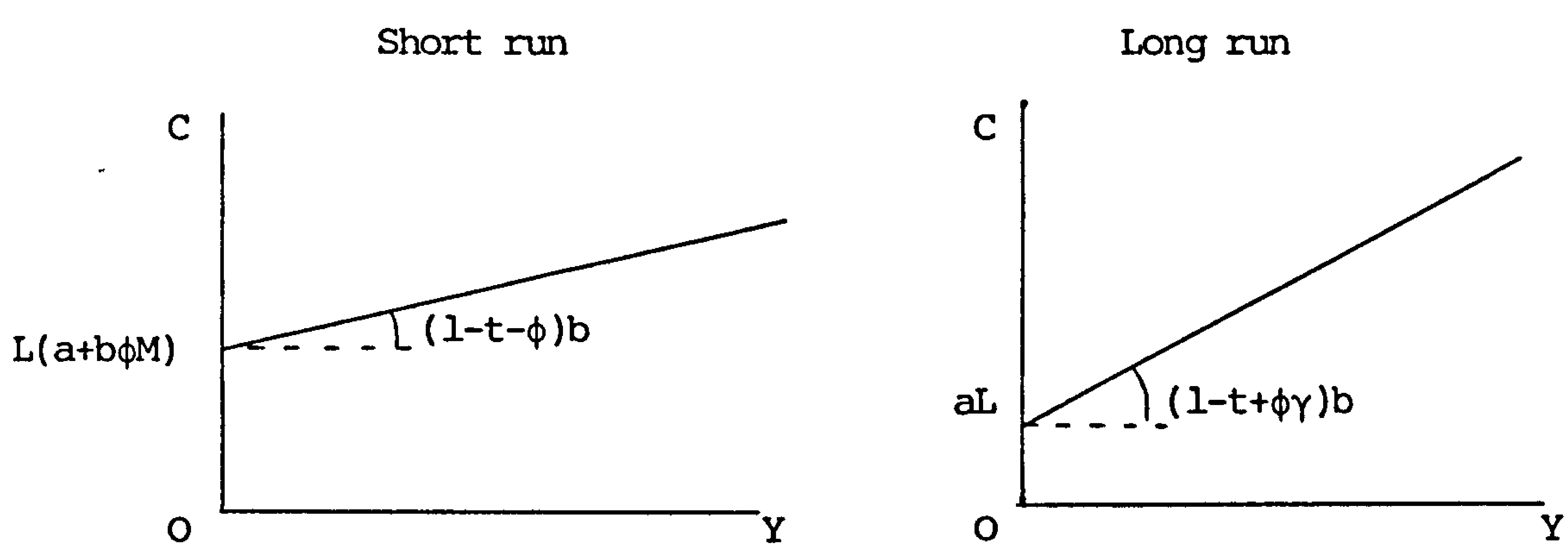
$$= L\left(a+b\phi\frac{Y}{\bar{E}}\right) + (1-t-\phi)\, bY$$

$$= aL + \left(1-t+\phi\left(\frac{L}{\bar{E}}-1\right)\right) bY$$

$$= aL + (1-t+\phi\gamma)\, bY$$

where $\gamma \equiv \frac{L-\bar{E}}{\bar{E}}$ = long-run ratio of unemployed to employed.

The function is derived from the fixity of $\bar{E}$, which allows M to be written as $Y/\bar{E}$. It follows that $\bar{E}$ enters into the final expression for the consumption function, although it is convenient to define the parameter γ representing the ratio of unemployed to employed in the long run. Depicting the short and long run functions diagrammatically, one obtains



For the same set of behavioural parameters, the long-run consumption function has a lower intercept and greater slope than the short-run version. The difference arises because of the role of unemployment in income changes. In the short run income movements are linked with employment variation, so that their impact on expenditure is lessened by the presence of unemployment benefits; those moving in and out of employment have a net change in income receipts which is less than the associated variation in national income, the difference depending on the extent of unemployment benefits.

In the long run a national income rise goes straight into disposable incomes with no significant changes in employment patterns. This contrast is illustrated by the effect of the replacement ratio, ϕ, on the two curves. Raising ϕ causes a flattening of the short-run consumption function, increasing the intercept and decreasing the slope. The opposite is true in the long run, where the slope depends positively on ϕ (since higher unemployment benefits do not affect Y, but still raise expenditure). The greater the replacement ratio, the greater the divergence between the two curves; when ϕ=0 they coincide. The slope of the long-run function increases with the level of unemployment, as represented by γ. Setting γ to zero makes the long-run curve identical in form to individual behaviour, such that

$$C = aL + (1-t)\,bY = aL + b(1-t)\,ML = L(a+bD)$$

The long-run curve is consequently a closer representation of individual behaviour than the short-run version : the difference when $\gamma \neq 0$ occurs only because the expenditures of the unemployed are from receipts which are not included in national income assessments.

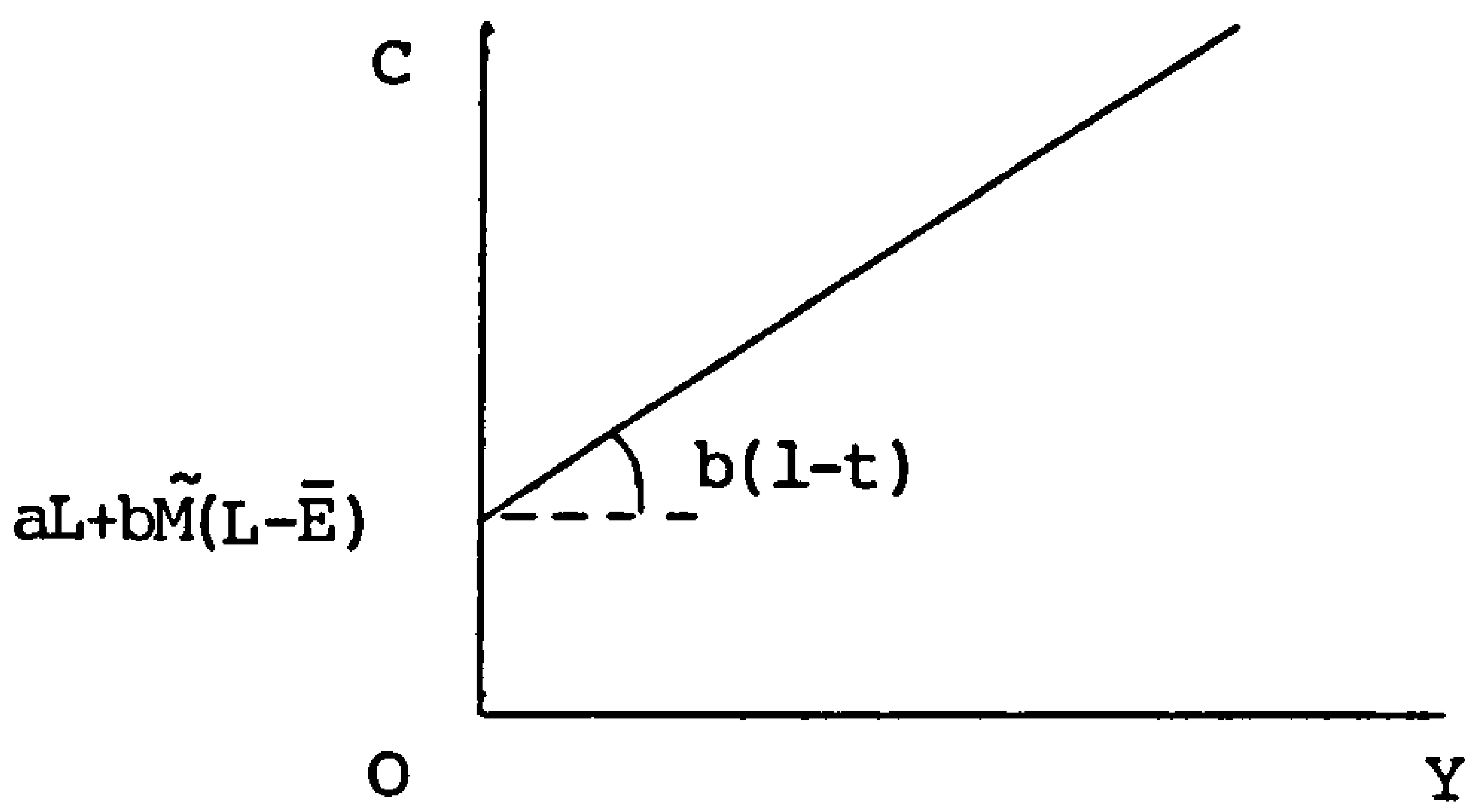
Suppose now that the replacement ratio does not remain constant in the long run, and that the unemployed receive a standard 'subsistence' payment irrespective of the national income level. In this case a steady upward trend in M leads to a rising Y, while unemployed incomes remain fixed at $\tilde{M}$. The long-run consumption function can then be derived from the short-run curve as follows

$$C = L(a+b\tilde{M}) + b((1-t)M-\tilde{M})\bar{E}$$

$$= L(a+b\tilde{M}) - b\tilde{M}\bar{E} + b(1-t)M\bar{E}$$

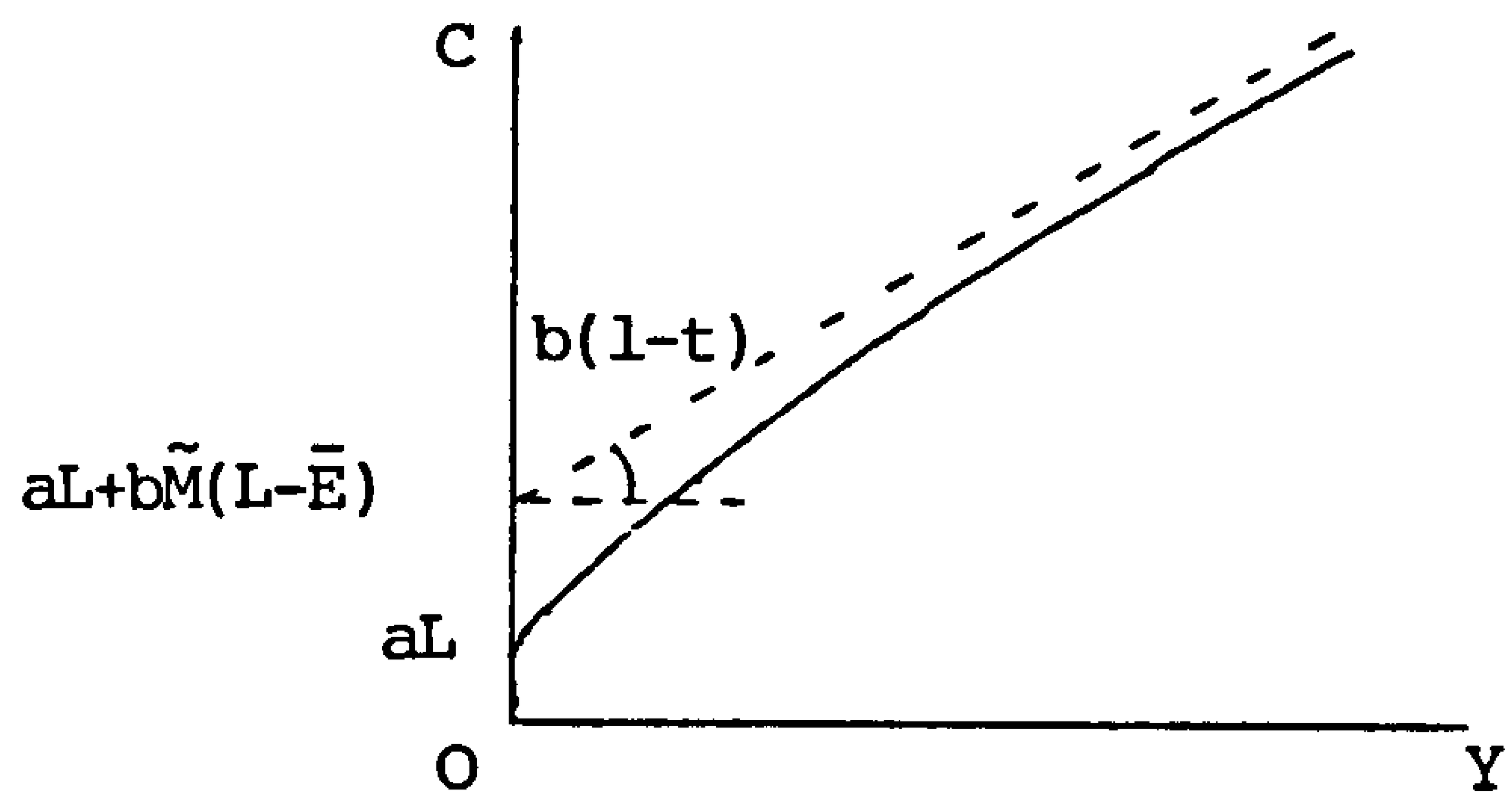
$$= [aL + b\tilde{M}(L-E)] + b(1-t)Y$$

Instead of rising parallel to M, the $\tilde{M}$ components remain static and are included in the intercept. Graphically the resulting function is such that



Long-run consumption function ($\tilde{M}$ invariant)

Comparison with the two curves above shows that this case falls between them. The function is steeper than the short-run curve, with a lower intercept; on the other hand, it is flatter than the previous long-run curve, with a higher intercept. Unlike the previous curves, the slope is independent of unemployment benefits, representing individual behaviour directly. The situation is therefore 'neutral', in the sense that unemployment benefits are constant ($\bar{E}$, $\tilde{M}$ fixed) and have neither a positive nor negative influence on the slope of the consumption function. In practice it does not seem likely that unemployment benefits will remain absolutely fixed in a period of economic growth. A position between the limiting cases is possible, where benefits increase at a rate lower than the growth rate of employment incomes. This can be represented by assuming that the replacement ratio decreases as Y and M rise over time, that is, $\phi = \phi(Y)$, $\phi'(Y) < 0$. In that case the long-run consumption function will be non-linear, with a diminishing slope as Y rises. If unemployment benefits approach an upper limit of $\tilde{M}$, then the situation is as below



Long-run consumption function (ϕ variable)

The resulting curve is a hybrid of the ϕ constant and $\tilde{M}$ constant cases; it starts with the lower intercept of the ϕ constant curve, and its slope gradually decreases (asymptotically) to that of the M constant curve. An outcome of this kind, somewhere between the two extremes, appears to be the most probable situation.

A further possibility is for the work force, L, to show long-run variation, probably as the result of demographic factors. The situation most frequently encountered is a chronic increase in population, which is reproduced in the size of the labour force. Where population is rising significantly, an increase in national income does not necessarily mean rising average real incomes; growth in national income is needed merely to maintain the same average living standards. If real incomes are to increase over time, national income growth must exceed that of population. In order to relate these factors to the aggregate consumption function, consider the extreme case where all national income growth is matched by changes in the labour force. Average real incomes therefore remain constant, and rising national income is absorbed by maintaining the new population at the customary income levels. It is assumed that the employment rate is roughly constant, equal to a given proportion δ of the labour force (so $E = \delta L$). The replacement ratio is also taken to remain approximately fixed in the long run. National income and the labour force are linked in such a way that

$$Y = ME = M\delta L \quad \Rightarrow \quad L = \frac{Y}{\delta M}$$

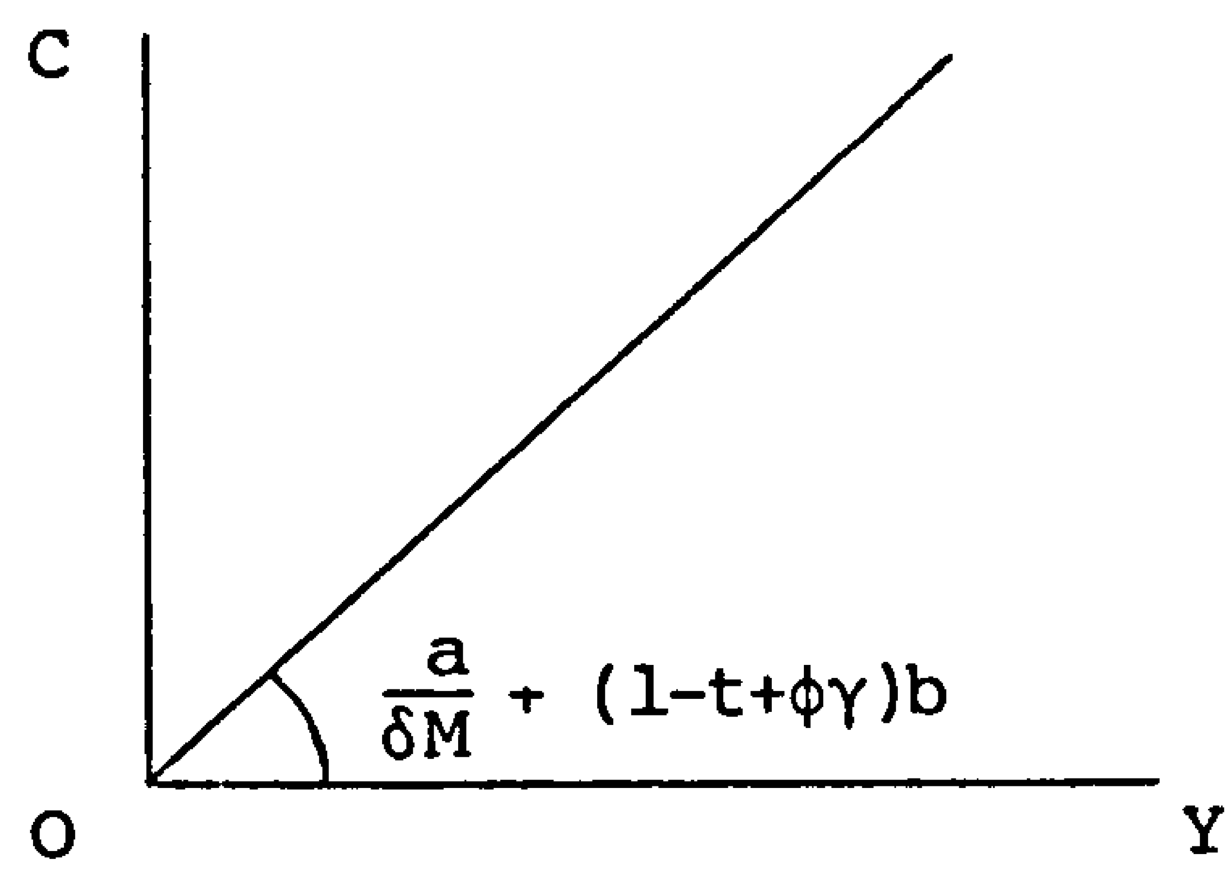
Hence the long-run consumption function can be derived as below

$$C = L(a+b\phi M) + (1-t-\phi)bY$$

$$= \frac{Y}{\delta M}(a+b\phi M) + (1-t-\phi)\,bY$$

$$= \left[\frac{a}{\delta M} + (1-t+\phi\gamma)\,b\right] Y$$

where $\gamma = {}^{1-\delta}/\delta$. Graphically this implies



Long-run consumption function (M constant, L varying)

The curve no longer has a positive intercept, and passes through the origin. It also has a steeper slope than any of the previous versions, differing from the L constant, M variable case by the term ${}^{a}/_{\delta M}$. Consequently there is a greater divergence, the higher is the long-term rate of unemployment. In reality it can be expected that simultaneous movements (generally increases) in M and L will occur. The outcome is then a combination of the two limiting cases considered here, with a slope steeper than when L is constant but flatter than when M is constant. As more income variation is accounted for by population change, the long-run consumption function becomes increasingly steep.

Several long-run consumption functions are possible, depending on the precise form of income changes. Their common feature is a lower intercept and greater slope than the short-run function defined in terms of changes in employment. This accords with expectations based on empirical findings. The accepted pattern of aggregate consumption function (as originally found in Kuznets (1946)) is that long-run time-series data give curves through the origin, while short-run (or cross-section) data lead to curves with a positive intercept and lower slope. Such findings are potentially explicable on the lines set out above, although the analyses usually offered are somewhat different. It is worth briefly comparing the present model with the conventional explanations.

Three competing explanations are commonly put forward : the 'Relative Income Hypothesis' (Duesenberry (1949)), the 'Permanent Income Hypothesis' (Friedman (1957)) and the 'Life Cycle Hypothesis' (Modigliani and Brumberg (1954)). These share the characteristic that they are based on individual behaviour. Consumption depends on some reference point or long-run income concept, which remains invariant in the short run. Consequently the MPC of individuals is higher in the long run than in a cross-section or in a short run time-series. This may well be true, and the reasoning on which the theories are based seems plausible enough. It is not strictly necessary, however, to appeal to individual behaviour to explain aggregate consumption functions. In the model used here differences in aggregate consumption functions arise directly from the way that the economy operates, even when individual behaviour is identical in the short and long run. Differential responses to short and long-run income changes would increase the difference between the consumption functions, but are not essential. If a theory based solely on individual behaviour is adopted, it remains to explain why individual incomes are varying in the short run. The example par excellence of a short-run income fluctuation is unemployment. In a sense the individual based theories rely on varying macroeconomic conditions to generate the short-term income movements on which they depend. It seems preferable to have an explanation centred directly on macroeconomic principles, rather than on a magnification of individual behavioural responses; variations in individual behaviour can then be appended to the model if they are felt to be appropriate. After all, the aggregate consumption function is supposed to be a macroeconomic relationship, and as such represents more than just individual behaviour patterns.

The discussion of this section has not attempted to set out a full long run analysis that can explain economic trends. All underlying changes have been assumed exogenous and the causes behind varying technology or population have not been considered. The objective has been more limited, to compare the different patterns of income variation that are likely to be found in the long run and the short run.

(5) Conclusion

One suspects that in practice short-term national income variations are accomplished mainly by the income changes of the unemployed. If that is the case, then it may be desirable to construct macroeconomic models on corresponding lines, as above. The adjustment is easy, and leaves the basic Keynesian conclusions unchanged. It does have a couple of implications, however:

(i) The simultaneous determination of employment and income downgrades productivity considerations. No technical production function is required to obtain the employment associated with observed national income.

(ii) If workers save at all, then they always create an aggregate

> demand deficiency for their own services. This can only be rectified by the various forms of non-employment demand, especially investment. Increased job search is futile unless it can call forth new autonomous expenditures.

To repeat the point made in the introduction : if we are trying to explain unemployment perhaps we ought to mention the unemployed in our model.

CHAPTER 7 : OPTIMAL PENSION AND RETIREMENT PRACTICES

(1) Introduction

Pension and retirement policies are complicated in nature, broaching redistributive and other issues. A formal pension or retirement policy is a constraint on individuals, so there is an immediate question of whether the benefits justify the costs. As a redistributive tool pensions are not the ideal choice, and other methods might well be preferred. This leaves no immediate theoretical case for formal state pensions or retirement, and a consequent scope for further analysis. If pensions are justified, there arises the additional problem of their optimal design. The discussion below addresses these matters.

The method resembles that in previous chapters. A government has the power to introduce a state pension scheme and formal retirement practices, and wishes to implement the best possible scheme. The objective is to choose a pension policy which maximises social welfare, subject to a given revenue constraint. Several interpretations of pension policies are possible. The usual distinction is between pensions financed on insurance principles and those permitting intergenerational income redistribution; it is convenient here to subdivide the first category, giving three cases:

(i) Insurance-based Pensions

The traditional view of pensions is as insurance, financed by contributions made during employment. Individuals are expected to claim in retirement a sum which, when discounted, is equal to the discounted sum of their contributions. Each person thus accumulates a separate fund which is enough to cover the requirements anticipated during retirement. Although the account for each individual does not in general balance exactly, total revenue is equated on average with total expenditure and the scheme breaks even (given accurate knowledge of the actuarial probabilities involved). Similarly, despite the fact that some individuals draw pensions exceeding their contributions and others less, depending on life span, the scheme is expected to be non-redistributive with no systematic income redistribution taking place. The result resembles private saving, where each individual makes an independent provision for future consumption during retirement. Formal insurance-based pensions (state or private) are largely paternalistic in spirit, with some outside agency acting as custodian of at least part of an individual's savings. Issues relevant to such pension schemes are considered in Section (3) below.

(ii) Redistribution within a Generation

The government may not in practice have the information required to achieve non-redistributive pensions : paying the same pension to people with different characteristics inevitably introduces redistribution. Moreover, if the initial income distribution is seen as inequitable, the government may wish to treat pensions as an instrument of redistributive policy. Viewing each generation separately, an appropriate combination of pensions and their associated contributions can bring about income redistribution. The result is then 'collective insurance' with each generation financing its own pension, but in a way felt to be socially equitable. Some individuals can expect to make a net gain from their pension revenues, subsidised by the expected losses of others. The outcome is more general than (i), sanctioning the redistribution of income within a single generation, while stopping short of redistribution between generations. Pensions schemes involving intragenerational redistribution are considered in Section (4).

(iii) Redistribution between Generations

The final possibility is that the pension receipts of one generation depend on the contributions of other generations. Redistribution can then occur between individuals of different ages, as well as those of the same age. Finance on intergenerational lines takes more than one possible form. For example, a requirement that the revenue constraint must balance at each separate date limits the range of redistribution to people with overlapping life spans. If, on the other hand, revenue restrictions are less stringent and do not require a balanced budget at all dates, it is possible to transfer incomes over a much longer time span. The latter case is equivalent to a 'social' pension fund, to which each generation contributes part of its income and from which they all draw their pension receipts. These two possibilities are compared and contrasted in Section (5).

The analysis of optimal pension practices can be based on any of the three structures described above, although the most general approach would be to allow full redistribution within and between age groups. None of these cases yields simple conclusions, and they share a number of features which inevitably complicate the discussion.

Firstly, all pension models have to be intertemporal, covering at least a single life span and possibly much longer. Decisions made at one date have an impact on future welfare, and must be based on forecasts of future events. Such predictions cannot be made in reality with complete accuracy, and both uncertainty and miscalculation play a part in actual policy decisions. Neither is easy to incorporate in models of decision making, nor readily compatible with optimality. No attempt is made to include them in the models below, and individuals are throughout assumed to have known fixed life spans and income patterns. That is not to say that the full range of necessary information is available in actual policy making, as is almost certainly not the case. But there remain substantial policy questions to be resolved even with full information, and it is as well to consider initially what the government would like to do under ideal conditions.

Secondly, the range of policy tools impinging on pensions is exceedingly wide. Underlying the whole question is individual retirement behaviour, and in particular the timing and speed of retirement (if it occurs at all). The population does not necessarily want a fixed universal retirement date, and any constraints imposed centrally should be justifiable in some way. This question is raised in Section (2), but is not carried over into the rest of the chapter. The remaining sections assume a fixed retirement date in line with current practice, although it is quite possible that a full social optimum would allow for individual choice over the timing and nature of retirement. Pension finance is another significant issue. Besides the insurance/intergenerational finance division, pension revenue can be raised by virtually any form of taxation. The models below largely assume a uniform lump-sum tax on the employed, but the results are not necessarily immune to changing the method of finance. Pensions themselves can also take a variety of forms (for instance, constant or moving over time, earnings-related or otherwise), in addition to differences in the pension level. A fully general model would have to permit simultaneous variation in all the possible policy parameters. The models below are more limited in scope, but serve to consider a number of the relevant issues in isolation.

Another feature of pensions policies is that they are confronted with individual saving behaviour, raising the problem of disincentives. In practice individual savings are a small and diminishing part of retirement incomes (Kay (1985)), but this may itself be due to the disincentive effects of widespread state pension. A number of empirical studies suggest that state provision replaces private savings to some extent (Munnell (1974), Feldstein (1974,1977)). The possible impact of disincentives is considered in the discussion below; in the extreme case it calls into question the rationale for state pensions, and otherwise may significantly influence the nature of the optimum. It is also possible that strong disincentives lead to the spreading of state provision in situations where it is not strictly necessary. Hence, while the low current level of private savings does not point to their having overriding importance, it is still worth including private saving in the various models.

Much debate on pensions has centred on the question of public as against private provision. The post-war period has seen a rapid growth of private occupational pensions in the U.K. and elsewhere, particularly among the more prosperous sections of the population. Recipients of such pensions are able to opt out of the state scheme, although their retirement practices and pension receipts are similar to state provision in many ways (at a somewhat higher level). It has been argued that private pensions are superior to state provision and that an extension of the private pension sector is desirable (Buchanan (1968)); the opposing case can also be made (Bosanquet (1983), Chapter 10). The discussion below is on the assumption that pensions are state-provided and does not directly investigate the public versus private sector debate. Many of the questions considered are common to public or private pension schemes, and the precise nature of the scheme is immaterial to the points made. The major exception is income redistribution, which is outside the scope of insurance-based private pensions.

A final aspect of pension design is the administration needed to operate the system. This also is likely to be large compared with some other forms of government policy. In common with the approach of previous chapters no administrative or institutional detail is included in the models below. It is possible that in practice administrative issues are sufficiently important to influence the nature of the policy optimum.

All these considerations make pensions a weighty policy problem. It is not even clear that formal pensions and retirement have a watertight justification, let alone a well-defined optimal structure. This chapter aims only to assess some of the salient issues.

(2) Retirement and the Individual

Most lives in developed countries are divided into three distinct stages, with an initial period of childhood and education followed by a working career and then retirement. The ages of transition between the stages are usually predictable to within a few years, and are frequently influenced by legislation on school leaving ages and pension entitlements. Individual choice can be exercised only to a limited extent at the transition points, with a number of people prolonging their education beyond the legal minimum or choosing to retire a few years before or after the statutory retirement age. Choice rarely extends to avoiding the tripartite education/work/retirement pattern, despite the freedom of any person not to enter the labour force. Participation is a real decision only for secondary workers in a household or for the minority of individuals whose unearned income is sufficient to support themselves without working. Rising female labour force participation since the Second World War (Joseph (1983)) implies that the three-stage life cycle is increasingly common and moving towards universality. Most individuals now enter the labour force at some point in their lives, and later retire from it.

The question arises of whether this is a spontaneous outcome of individual choice. On the basic issue of working against not working, the need for people to work is governed by the general level of labour productivity in a society relative to the society's total wants. Current productivity is nowhere near enough to supply total wants without major inputs of labour, and most of the population must consequently work to maintain the customary standard of living. Society's need to undertake productive activity is conveyed to individuals by their lack of non-employment earnings, leaving little choice but to participate in the labour force. Higher labour productivity should permit an increase in the general level of employment incomes (provided that the benefits are evenly distributed), allowing a decrease in total working hours; at the extreme of a fully automated society we could all afford to be rentiers and live off the returns to our collective investments. The need to work is therefore a fact of economic life, and the remaining question is how best to organise a working existence.

A major element in any career structure is the timing of departure from the labour force, and the manner in which this is effected. Workers ought presumably to have strong opinions on this issue, given that it has a major impact on the later years of their life span. The discussion below concentrates on the implications of conventional labour supply models for the retirement question and for the notion of an 'optimal' retirement policy.

Neoclassical theory treats labour supply as a decision by the worker, and analyses it in the same way as decisions on consumption. Labour supply at any time depends on individual preferences over leisure and consumption goods, and responds to changes in the wage rate and other exogenous variables. In practice individuals may not be at liberty to adjust their labour supply, facing a fixed working week and a commitment in the number of hours they work. Against this apparent obstacle, individual choice can be claimed to occur in the long run (being responsible for the secular decline in working hours) or indirectly (by the initial choice of job or by changing jobs). Much effort has been devoted to estimating the size of labour supply effects, as summarised in Killingsworth (1983). Despite some doubts, there is a belief arising from conventional theory that work patterns are influenced by individual choice.

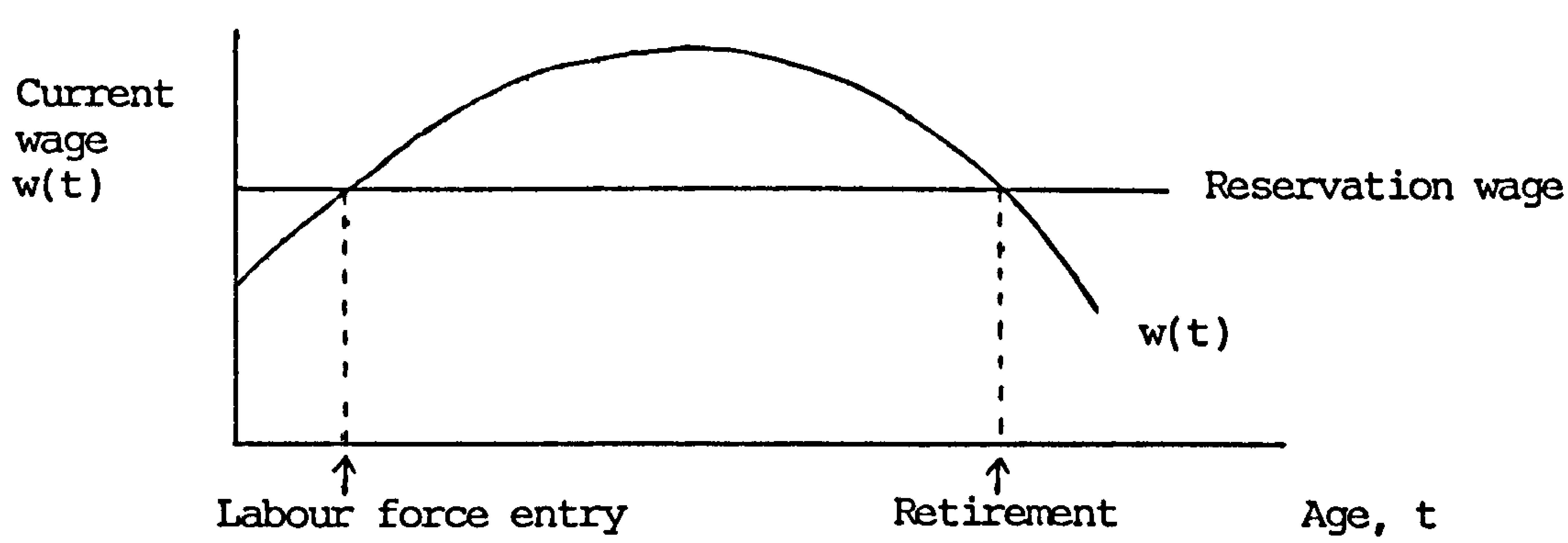
Labour supply models are most commonly addressed to the short-term problem of the working week, and are also sometimes used to account for the decision on labour force participation. They appear less often in discussions of retirement, although if the individual is influencing labour supply at all parts of the life cycle it follows that the retirement decision is also governed by the same theory. It is consequently of interest to consider what neoclassical labour supply theory implies about retirement, and the following discussion does this within a life-cycle model. Several cases can be distinguished, according to the presence or absence of saving or investment in human capital.

(i) Models with Zero Saving

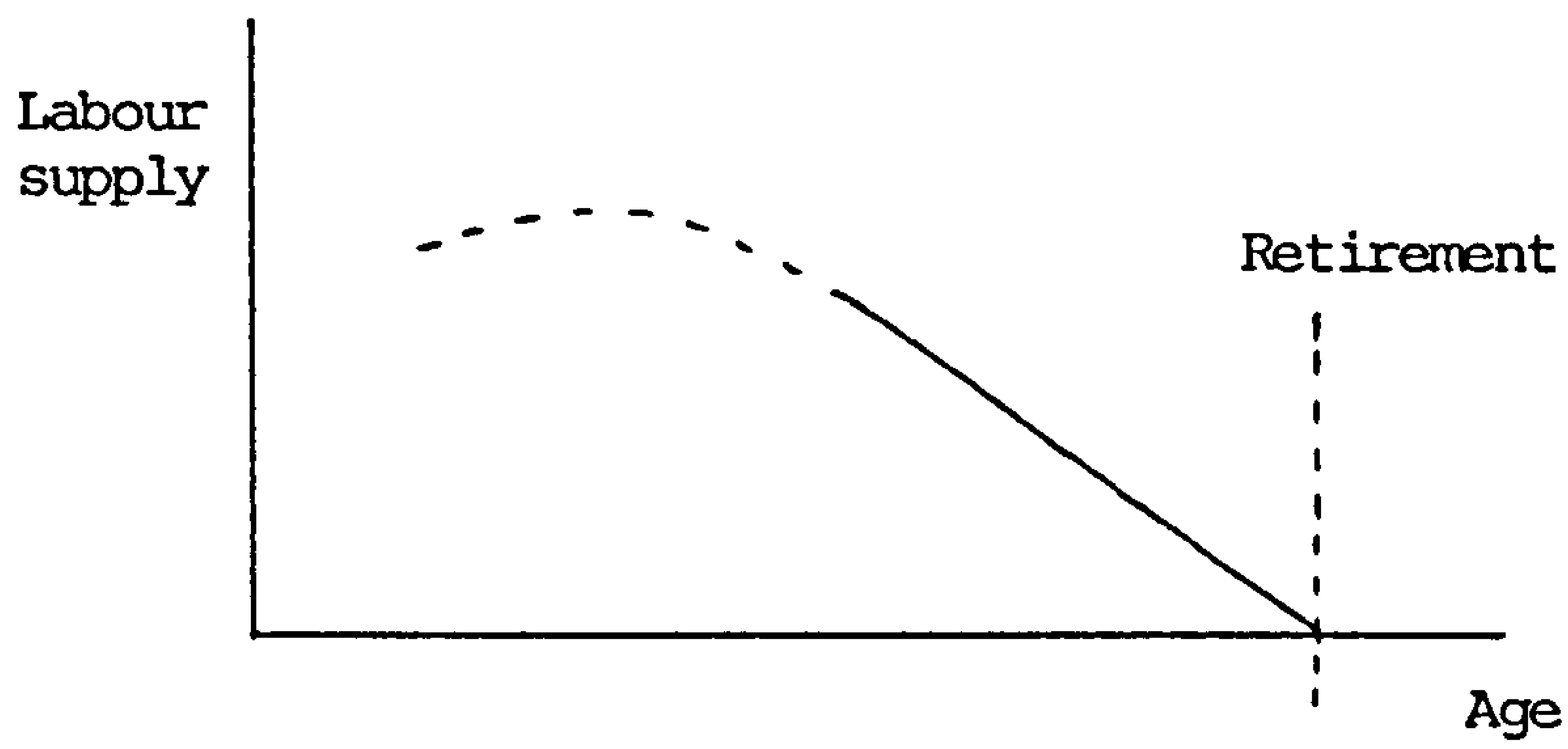
Savings are usually omitted from the simple labour supply models found in textbooks on labour economics (such as Hamermesh and Rees (1984), Chapter 2, or Addison and Siebert (1979), Chapter 3). Leisure, l, is treated as a good like any other, with a demand function dependent on other prices, $\underline{p}$, the wage rate, w, and non-employment income, M. Participation decisions are governed by the 'reservation wage' $\tilde{w}$, defined such that $l(\underline{p},\tilde{w},M) = L$, where L is the available number of hours per week. An individual takes up employment only if a job is available such that $w \geq \tilde{w}$; conversely, voluntary retirement occurs when the wage rate falls below $\tilde{w}$.

With zero saving, it can safely be assumed that non-employment income is at a low level, showing little variation over the life cycle. The sole exogenous factor determining retirement is then movements of the wage rate over time, as summarised by an individual's age-earnings profile. Manual workers are normally assumed to face a 'hump-shaped' profile, while non-manual workers receive continuously rising wages and earnings (Welford (1958), Gordon and Blinder (1980)). In both cases the standard explanation is that wages respond to the individual's 'human capital', that is, the general degree of skill and working capacity (Mincer (1958, 1974)). Manual workers gain little from work experience and face declining physical strength, whereas non-manual workers can benefit from accumulated knowledge throughout their careers. In the following discussion the time pattern of wage rates is taken to be exogenous, although relaxing these assumptions would not alter the conclusions of this section.

With zero savings and preferences separable over time there is no link between different time periods, and life-cycle labour supply can be treated as a sequence of separate single period decisions. Over a known life span $t \varepsilon [0,T]$, labour supply at time t will be determined by the prevailing wage rate w(t), where w(t), $t \varepsilon [0,T]$, describes the exogenous time path of wages. Voluntary retirement will occur only if $w(t) < \tilde{w}$; otherwise the individual still wishes to work for a certain number of hours, even though working time may be declining. The position can be depicted as below:



In the absence of changes in preferences or non-employment income, w(t) must decrease over time before individuals choose to retire voluntarily. The fact that many workers do not experience falling wages is not easily compatible with voluntary retirement on the lines of this model. A smooth function w(t) also implies that the approach to retirement is gradual, displaying a steady fall in working hours until the point of retirement is reached. The situation is as below



where the earlier path of labour supply could take any form (according to the balance of income and substitution effects), but the final approach to retirement is a continuous downward trend. Sudden movements from full-time working to retirement (frequently observed in practice) can only occur in this model if there is a discontinuous change in individual preferences or unearned income.

Once can conclude that existing retirement practices do not coincide with the retirement predicted by simple labour supply models. Non-gradual retirement is chosen only if induced by some outside agency, for example, by the provision of pension income beyond a certain age (especially when conditional on ceasing work). In these models single date retirement has to be imposed by the structure of pensions, and does not arise spontaneously from individual labour supply.

(ii) Models with Non-Zero Saving

A possible deficiency of the model in (a) is the absence of saving behaviour, which can generally be expected to have some influence on retirement choices. If rational saving is introduced, the outcome is a model of dynamic labour supply (Weiss (1972), Ghez and Becker (1975)). The effect of this on retirement is considered below.

A dynamic model is soluble in a similar way to optimal saving models. The individual maximises intertemporal utility, expressible as

$$U^* = \int_0^T e^{-\rho t} U(C_t, l_t)\, dt + \phi(K_T)$$

where C_t, l_t, K_t are consumption, leisure and wealth at date t, and $\phi(K_T)$ is the value placed on the terminal wealth holding. Leisure and consumption are chosen over the life cycle to maximise U* subject to the constraints

$$\dot{K}_t + C_t = w_t (L-l_t) + rK_t \qquad \forall_t$$

where w_t is the exogenous wage rate at time t, L is the maximum work/ leisure hours available at any time, r is the interest rate, and the dot denotes differentiation with respect to time. Leisure cannot exceed the maximum L hours available (so that $l_t \leq L, \forall_t$) and all other time is spent working, implying a labour supply of $L-l_t$. There is a fixed initial wealth of $\bar{K}_0$ units. The problem is soluble by optimal control methods, with C_t and l_t as instruments and K_t as a state variable. The Hamiltonian and optimality conditions are

$$H = e^{-\rho t} U(C_t, l_t) + e^{-\rho t} \lambda_t (w_t(L - l_t) + rK_t - C_t) + e^{-\rho t} \mu_t (L - l_t)$$

$$H_c = 0 \Rightarrow \frac{\partial U_t}{\partial C_t} - \lambda_t = 0$$

$$H_l = 0 \Rightarrow \frac{\partial U_t}{\partial l_t} - \lambda_t w_t - \mu = 0$$

$$H_k = -e^{-\rho t} \dot{\lambda}_t - \rho e^{-\rho t} \lambda_t = e^{-\rho t} r \lambda_t \Rightarrow \dot{\lambda}_t = (r - \rho) \lambda_t$$

with a transversality condition $U'(C_T) = \phi'(K_t)$. Combining the first and third equations yields a standard optimal saving condition of the form

$$\dot{C}_t = \frac{-\partial U / \partial C_t}{\partial^2 U / \partial C_t^2} (r - \rho) \qquad \forall t$$

Similarly, where the leisure constraint is not binding, the second and third equations imply the condition

$$\dot{l}_t = \frac{-\partial U / \partial l_t}{\partial^2 U / \partial l_t^2} (r - \rho - \dot{w}_t) \qquad \forall t$$

which has the same form as the condition for consumption except for the presence of $\dot{w}_t$. Leisure demand therefore resembles the demand for an ordinary commodity, subject to a variable price, w_t, and an upper bound, L.

Retirement in this model occurs when the constraint on L_t becomes binding, leaving a labour supply of zero. As in the previous case, the retirement date (if any) depends on the time path of w_t, and virtually any life time pattern of work can be obtained from the model. The main difference from the zero saving model is in the relation between the wage rate and retirement, where the previous model suggested that retirement can only arise when the wage rate is declining over time. This is also true in the present case when $r = \rho$, since $\dot{l}_t$ then has the opposite sign to

$\dot{w}_t$ and labour supply falls only when the wage rate is also falling. Making the assumption $r=\rho$ rules out intertemporal effects (and income effects are eliminated by the ability to transfer income across periods) so labour supply reacts in the expected way to changes in the wage rate, yielding the same outcome as previously. In practice, however, it may well be that $r>\rho$, producing positive net returns to saving. This means that some value is placed on working early in the life cycle to accumulate wealth, and that a rising wage rate is compatible with falling labour supply (where $r-\rho > \dot{w}_t$). The model can therefore produce voluntary retirement without decreasing wage rates, which overcomes one of the obstacles to depicting existing retirement practices as the outcome of individual choice.

The other obstacle remains : that is, a retirement decision will in general be preceded by a gradual reduction in working hours. Given smooth movements in w_t and other variables the typical pattern is for l_t to increase (and labour supply to fall) until the point $l_t=L$ is reached, beyond which the constraint on l_t is binding and retirement can be said to have occurred. A jump from positive labour supply to zero requires a jump in one of the other variables, for example, in the wage rate or non-employment income. This could well apply to some individuals, but it is not going to arise in the same way and at the same time for all individuals (unless it stems from an exogenously imposed pension scheme raising non-employment income at a given date). The same comments can be made as in (a), namely that the model does not yield a voluntary pattern of retirement akin to the retirement practices prevalent in reality.

Further elaborations of the labour supply model are also possible. One such is to introduce human capital into the analysis, making wages an endogenous return to earlier investments (Blinder and Weiss (1976)or Heckman (1976)). Another possibility is to extend the utility function from the simplest two variable cases to a more general form, on the lines mentioned in Chapter 5, Section (2). These changes might perhaps be regarded as desirable, but they do not in general alter the main point made above. Under normal assumptions models of life-cycle labour supply produce a steady reduction in hours worked towards zero, not a single date retirement.

There are two alternative responses to this divergence between retirement in theory and practice. One is to try to narrow the gap by emphasising that actual retirement does possess some of the features predicted by the theory. The other is to conclude that there is no strong link between individual choice and work patterns, and that working practices are largely determined by institutional structures outside the domain of individual preferences. Such a view is the one preferred here, and does not necessarily diminish the value of labour supply theory. In fact, the theory probably does give a reasonable account of people's preferences about their working lives; it is probably true that most people would prefer a gradual reduction in working hours to a sudden withdrawal from work. If one accepts these conclusions, then it follows that something else is determining retirement behaviour. Rather than playing its usual role as a self-contained explanation of working practices, labour supply theory can serve to emphasise the differences between what individuals would choose to do and what happens in practice.

A desire for gradual retirement exacerbates the government's policy problem, which cannot even be characterised as a straight choice between single date retirement policies and uniform graduated retirement schemes. Any uniform scheme constrains some individuals, so the initial question is whether or not individuals should be given a free hand in their retirement decisions. The case for flexible retirement has often been put (Sauvy (1969), de Beauvoir (1970), Palmore (1972), Walker (1980), Parker (1982)), and it seems convincing enough on purely individualistic grounds. The matter is not in the conventional mould, and involves a decision whether or not to constrain individuals, instead of the setting of existing policy parameters. To impose a common retirement scheme on rational individuals must always reduce social welfare below the no policy position, since it merely constrains individual behaviour. Whatever the reason for having formal retirement and pensions, it must lie outside the realms of utility-maximising individuals with perfect foresight. A full theoretical analysis of this problem is probably not feasible, but it is undeniably important as a policy issue. The following sections are based on a fixed formal retirement date, as is usually observed in practice. This assumes implicitly that any advantages of formal retirement are sufficient to justify the constraints imposed on a large number of individuals.

(3) Pensions and the Individual

Pensions can be seen variously as income maintenance, social insurance or a redistributive tool. Perhaps the commonest idea is that they are a social security measure, designed to guarantee an acceptable living standard for the aged. This accords with their central role in the welfare state, although it is sometimes argued that state pensions induce people to leave employment and accept a lower income level, thus actually increasing poverty (Townsend (1981), Walker (1980)). An alternative rationale for pensions is as social insurance, geared specifically to disability in old age rather than old age itself. Pensions would then be paid to those unable to work, and would happen to be received by the elderly only because old age is highly correlated with disability. Schemes operating on this basis are best treated as a special case of general social insurance, as in Diamond and Mirrlees (1978). A further possibility is that pensions serve a redistributive function both within and between generations. Discussion of this aspect is deferred until Sections (4) and (5).

The current section adopts the income maintenance interpretation, in that it sees pensions as part of the individual's life-cycle pattern of income receipts. Nevertheless, the yardstick employed is the maximisation of life-cycle utility, not the 'satisficing' implicit in maintaining minimum income standards. There is no real insurance in the face of uncertainty, and individuals are taken to have a fixed life span and knowledge of their future retirement date.

Two central questions need to be considered. Initially comes the decision on whether state pensions are justified at all, or whether individuals should be left alone to make their own provision for old age. If pensions are to be advocated, there occurs the further question of the form they should take and their relationship with retirement practices. The discussion below looks at these issues, using simple models of individual life-cycle saving.

It is convenient to distinguish three main cases, a single individual with optimal savings, a single individual with suboptimal saving, and the presence of many individuals.

(i) Optimal Individual Saving

In a world of perfect foresight and rational agents, individuals plan their future consumption optimally. Any government policies on pensions or retirement are incorporated in this assessment, and result in an appropriate adjustment of individual behaviour.

The most basic case, imposing the minimum number of arbitrary constraints on the individual, is the dynamic labour supply model outlined in Section (2), part (a). Allowing a free choice of retirement date means that complete retirement may or may not occur, according to individual preference, and that working hours are liable to be reduced gradually.

Fully flexible provision of pensions can be represented by allowing the government to make a net lump-sum payment b_t at any date t, where b_t corresponds to taxation when negative and to a pension when positive. One may assume that the government's payments satisfy

$$\int_0^T e^{-rt} b_t \, dt = 0$$

as there is no income redistribution taking place. A pension policy of this type has nothing to offer the rational individual, and leaves the final consumption path unchanged. In particular, because the taxes have a lump-sum form they do not interfere with the optimality conditions for consumption and labour supply, limiting their effect to any change in lifetime income. The requirement that payments and receipts must balance over the life cycle means that the net impact on lifetime income is zero, so there is nothing to be gained from introducing the pension scheme. Any taxation displaces the same amount of private savings, and results in later consumption being financed out of pension receipts instead of savings. The individual is indifferent to the form taken by the b_t payments (from the infinite number of possibilities), and there is no sense in which any particular structure of payments is superior to the others. Indeed, if government policy involves any administrative costs, or if individuals prefer to manage their own affairs, then the optimal position is to do without pensions. A similar outcome occurs when individual decisions are limited to saving, subject to a known pattern of retirement and labour supply.

The position resembles that in the permanent income or life-cycle models (Friedman (1957), Ando and Modigliani (1963)), where consumption responds to changes in lifetime income. Under optimal saving a pension scheme only influences individual utility and consumption levels to the extent that it alters lifetime income; if the individual's total contribution and receipts balance over the life cycle, then there is no net effect, and the individual's utility remains unaltered. Even when pensions do raise lifetime income, their impact will tend to be an upward shift in the whole of the consumption path, rather than an increase during the period of high pension receipts. Paying pensions to the elderly would not necessarily have much impact on either their consumption or labour supply. Under optimal saving the specific structure of any state pension becomes irrelevant.

If the government happens to disagree with individual preferences, then its ideal optimal saving path will differ from the individual's optimal saving plan. Pension policies are not much help in this case, because whatever pension structure the government chooses to impose, individuals can always adjust their saving to return to their own preferred consumption path. The government can only use pensions to influence individual behaviour by either dictating the level of all personal savings or by threatening constantly to adjust pensions as a counter-response to adjustments in individual saving. Neither of these are attractive policy options, so there is no real rationale for state pensions. A better means of influencing individual saving would be to use interest rate taxation to vary the net returns on personal wealth (see Chapter (4)).

The general conclusion is that formal pensions are not easily compatible with individual rationality.

(ii) Suboptimal Individual Saving

Suppose now that optimality is defined by the government's preferences, and that individual saving fails to reach the optimal path. The government may wish to implement a pension scheme to try to achieve its preferred saving pattern. Its success is governed by the nature of the individual reaction. There are three main possibilities:

(a) Fixed ex ante plans

In this case the individual has a fixed plan for private saving, determined before government policy is introduced and adhered to. An example would be the classical assumption of zero saving, holding regardless of the level of taxation. There are no resulting difficulties for government policy, since the passive individual behaviour creates no disincentive effects and allows the optimal consumption path to be reached. Taxes and pensions will be set equal to the difference between the individual's initial saving plans and the optimal saving plan.

(b) Plans responsive to policy

A second case is where individual saving plans are dependent in some way on the government's pension scheme. This can be represented by making individual consumption or savings a function of government policy measures, that is, $C = C(b_1, \ldots, b_T)$, $\forall_t$ where C_t is individual consumption at time t and b_t denotes the government's pension/tax policy (as in (i) above). The function C_t forms a genuine constraint on the policy optimisation problem, and will have to be taken into account if utility is to be maximised. It is possible that the unconstrained optimal saving path will still be attained (for example, if individual savings are a fixed proportion of the government's forced savings), but in general the presence of the disincentive will prevent a fully optimal savings pattern being attained. Nevertheless, it will usually be true that government intervention can improve on the original savings plan.

(c) Fixed ex post plans

Individuals here have a given saving plan which they maintain in the face of formal pension provision. If pensions are introduced, then individuals merely adjust their plans so that consumption and net saving correspond to their preferred life-cycle pattern. The situation resembles that with rational saving when individual preferences diverge from those of the government, involving a set of pension plans which are independent of the government's policy. Given that individuals can always choose to return to their own suboptimal plans, there is little that the government can accomplish by introducing pensions (short of dictating saving behaviour or using the threat of pension changes to bargain with individuals). This is the only case in which state pensions are not clearly justifiable when individual saving behaviour is suboptimal.

Hence, under the conditions assumed, there is a corrective role for state pensions as a means of bringing suboptimal individual saving back on to the optimal path. One may, of course, question the state's right to override individual choices and determine 'optimality' : paternalism is a controversial topic, and the general pros and cons carry over to this particular case.

(iii) Many Individuals

In practice, a formal pension scheme usually applies to many individuals of diverse circumstances and preferences. This may lead to redistribution (see Section (4)),and raises several other issues. Two of them, the creation of pension dependence and the impact on aggregate output, are considered below.

(a) The Creation of Pension Dependence

Given rational individual saving and a government without redistributive ambitions, there is little to be gained from formal pension schemes. One might then wonder why state pensions are so widespread, as they are rarely defended explicitly on the grounds of correcting individual irrationality or as a redistributive measure. The usual reasoning behind pensions is as income maintenance for the aged, and yet such a role would suggest a more limited pension incidence than is in fact the case. The question arises of whether any other factors have influenced the spread of pensions.

When saving is rational, pensions or other social security should generally discourage private savings. Arguments have been put forward that this has indeed occurred, resulting in a decrease in aggregate capital accumulation (Feldstein (1974)). Others have maintained that the effect is not particularly important empirically (Green (1981)). Whatever the truth in practice, in economic theory the rational response of individuals to social security schemes would be to reduce the extent of saving. The implication is that non-redistributive pensions contribute little to economic welfare and should be reduced in extent. There is something of a paradox here, however. Where many individuals are present, strong saving disincentives are liable to encourage the spread of state pension provision, despite its undesirability. A society which dabbles in social security for the aged may easily find itself with a population universally reliant on state 'pensions'.

The creation of pension dependence can be illustrated by a simple two period model of life-cycle saving. Suppose that there is a single generation of identical individuals who all work in the first period and are retired in the second. Utility is

$$V = U(C_1) + \frac{U(C_2)}{1+\rho}$$

where C_1, C_2 are consumption in the first and second periods and ρ is the discount rate on future utility. Income received from working in the first period is given by M, so individuals choose C_1 and C_2 to maximise V subject to a budget constraint

$$M = C_1 + \frac{C_2}{1+r}$$

where r is the interest rate. Rational individual saving implies the condition

$$U'(C_1) = \left(\frac{1+r}{1+\rho}\right) U'(C_2)$$

which, along with the budget constraint, defines the individual's choice of the optimal C_1^*, C_2^*. These values provide the socially optimal outcome, and cannot be improved on by government intervention in pension provision. Now assume that the government introduces a limited income maintenance scheme, guaranteeing a minimum current income of m. If $m < C_2^*$, then the initial optimal saving plan would not qualify for state assistance. Nevertheless, rational individual behaviour could still lead to saving disincentives, since individuals may deliberately run down their savings in order to claim the benefit. The government might not be aware of such responses, but even if it was, it would find it difficult to refuse the provision of a subsistence income on humanitarian grounds. Individuals therefore have a choice between maintaining their initial position C_1^*, C_2^* or reducing their savings to zero and claiming the full benefit m (there being no point in retaining a small positive saving level and claiming less than the full benefit). If each individual ignores the behaviour of others, the benefit is claimed when the condition

$$U(M - \frac{m}{H}) + \frac{U(m)}{1+\rho} > U(C_1^*) + \frac{U(C_2^*)}{1+\rho}$$

holds true, assuming that benefits are financed by lump-sum taxes, $^m/_H$, where H is the size of the population. For a large H the effect of one person claiming benefits on the level of taxation is negligible, and can be ignored by the individual in making a decision : a possible income gain of m units is then being weighed against the inconvenience of rearranging the intertemporal consumption pattern. Even an income maintenance level considerably below observed income levels can potentially lead to a substantial decrease in private savings. Any ex post benefit from this behaviour depends whether the others respond in the same way. If some have scruples about claiming benefits, then the non-savers may end up prospering at their expense. Purely economic rationality demands that all individuals behave in the same way, so private saving in the economy would collapse and the benefit m would become a universal state pension. The net outcome would be inferior to the initial position, unless it so happens that $m = C_2^*$

The difficulties experienced are similar in nature to 'free rider' problems in the financing of public goods by voluntary contributions. To illustrate this, consider a numerical example with two identical individuals. Preferences are $U = C^{\frac{1}{2}}$ in both periods, and it is assumed for simplicity's sake that $\rho=r=0$. Income arbitrarily takes the value M=200. With no policy intervention, consumption will be split equally between periods, so that $C_1^* = C_2^* = 100$ and V = 20 for both individuals. Suppose now that a minimum income guarantee of m is introduced. The threshold value of m governing saving disincentives is given by

$$\left[200 - \frac{m}{2}\right]^{\frac{1}{2}} + m^{\frac{1}{2}} = 20$$

and solving this yields $m \approx 44.44$. If $m > 44.44$, then individuals are better off not saving and claiming the benefit, given the assumption that others do not react likewise; if $m \leq 44.44$ individuals continue to save as before. Let m be sufficiently generous to fall in the range above 44.44, and assume that, say, m = 50. The possible range of outcomes can be depicted in game theoretical terms, producing a set of utility payoffs as below:

		Individual B	
		Save	Not Save
Individual A	Save	$V_A = 20$ $V_B = 20$	$V_A = 18.71$ $V_B = 20.30$
	Not save	$V_A = 20.30$ $V_B = 18.71$	$V_A = 19.32$ $V_B = 19.32$

The dominant strategy for both players is to reduce savings to zero and claim the benefit, producing a Nash equilibrium in the bottom right-hand corner. This is a standard case of the 'Prisoners' Dilemma', in which a superior result can be obtained if the players agree to maintain their original saving patterns. The co-operative solution is unstable, however, given that the individuals always have an incentive to break it and profit at the expense of the other player. Society is liable to end up with state 'pensions' which are not designed for that role and cause a reduction in social welfare. In contrast to the usual kind of policy problem, a co-operative agreement is here required to avoid the spread of public provision. Without restraint in the claiming of state benefits, the social security system will become universal, producing results inferior to an absence of pension provision. It is therefore conceivable that state pensions grow spontaneously in circumstances where they have no social value. The effect could be avoided by setting m at a sufficiently low level, reducing the generosity of the social security system.

This argument becomes rather stronger when the population is allowed to differ in its initial income levels. Let the income distribution be depicted by the continuous function f(M), such that

$$\int_{\underline{M}}^{\bar{M}} Mf(M)dM = Y$$

where $\bar{M}$, $\underline{M}$ are the maximum and minimum levels of individual income and Y is the aggregate income of the population. The model otherwise remains as before, with the same two-period structure of saving and retirement. Income maintenance again takes the form of a minimum guaranteed current income, m, raised from taxing employment earnings, M. In order to abstract from redistributive questions it is assumed that taxes are proportional to income, thus maintaining the initial pattern of income levels. If $M^{\dagger}$ denotes the threshold income below which saving falls to zero and the benefit is claimed, then the number of claimants will be

$$\int_{\underline{M}}^{M^{\dagger}} f(M)\, dM = F(M^{\dagger})$$

where F(M) is the cumulative distribution function associated with the density f(M). The proportional rate of taxation is then equal to

$$t \equiv \frac{mF(M^{\dagger})}{Y}$$

that is, to the total level of benefit payments divided by aggregate income. A range of different income levels makes the income maintenance policy rather more credible than in the identical population case above, and it can be assumed that the benefit is claimed by at least the lowest income group in the population : social security is redundant if it does not cover even the poorest members of society. Benefits are claimed if the condition below applies

$$U\left(\underline{M}\left(1-\frac{mF(M)}{Y}\right)\right)+\frac{U(m)}{1+\rho} > U(\underline{C}_1^*)+\frac{U(\underline{C}_2^*)}{1+\rho}$$

where $\underline{C}_1^*$, $\underline{C}_2^*$ are the optimal consumption levels associated with income $\underline{M}$. Consider now what happens as individuals decide whether or not to claim the benefit. If there exists a threshold income, $M^{\dagger}$, at which an individual is indifferent between saving and not saving, then it must satisfy

$$U\left(M^{\dagger}\left(1-\frac{mF(M^{\dagger})}{Y}\right)\right) + \frac{U(m)}{1+\rho} = U(C_1{}^{\dagger *}) + \frac{U(C_2{}^{\dagger *})}{1+\rho}$$

where $C_1{}^{\dagger *}$, $C_2{}^{\dagger *}$ are consumption levels chosen with the net income

$$M^{\dagger}_{net} = M^{\dagger}\left(1-\frac{mF(M^{\dagger})}{Y}\right) = M^{\dagger}(1-t)$$

This condition can never be stable, however, because one group's decision to claim the benefit always means that those with a marginally higher income level will wish to follow suit. Specifically, as individuals at an income $M^{\dagger}$ decide to claim benefits, the tax rate will rise and the effect on after-tax income is

$$dM^{\dagger}_{net} = -M^{\dagger}\frac{\partial t}{\partial M^{\dagger}}dM^{\dagger} = -\frac{M^{\dagger}m\ f(M^{\dagger})}{Y}\ dM^{\dagger} < 0$$

In terms of the threshold condition the fall in the left-hand side is $U'(M^{\dagger}_{net})\ dM^{\dagger}_{net}$, while the fall in the right-hand side is

$$\left(U'(C_1^{\dagger *})\frac{\partial C_1^{\dagger *}}{\partial M^{\dagger}_{net}} + \frac{U'(C_2{}^{\dagger *})}{1+\rho}\frac{\partial C_2{}^{\dagger *}}{\partial M^{\dagger}_{net}}\right) dM^{\dagger}_{net}$$

$$= U'(C_1{}^{\dagger *})\left(\frac{\partial C_1{}^{\dagger *}}{\partial M^{\dagger}_{net}} + \frac{1}{1+r}\frac{\partial C_2{}^{\dagger *}}{\partial M^{\dagger}_{net}}\right) dM^{\dagger}_{net} = U'(C_1{}^{\dagger *})\ dM^{\dagger}_{net}$$

using the optimal saving condition and the adding-up property. Since $C_1{}^{\dagger *} < M^{\dagger}_{net}$ and $U'' < 0$, it follows that $U'(M^{\dagger}_{net})\ dM^{\dagger}_{net} < U'(C_1^{\dagger})\ dM^{\dagger}_{net}$, and the left-hand side now exceeds the right-hand side. Individuals with a slightly higher income than $M^{\dagger}$ are induced to claim the benefit as well, so that $M^{\dagger}$ cannot be a stable threshold income level. The same argument applies right up to

$\overline{M}$, so in the end the whole population decides not to save and to claim the minimum income, m. In others words, given non-redistributive taxes and a continuous income distribution, any operational income maintenance scheme (regardless of how low m is) leads to a collapse of all private saving and a universal reliance on the minimum 'pension' level. The only real exception within the model's own terms is if there is a gap in the income distribution sufficiently large for the richer group to be better off not claiming the benefit (though in practice income distributions are virtually continuous until very high income levels). Thus optimal saving behaviour in this model is not easily compatible with a limited scale of state welfare provision. Introducing income support measures causes a spreading of benefit claims which is not liable to produce a socially optimal outcome and leads to an extensive and systematic redistribution of income between individuals.

Two possible solutions can be identified. One would be for the government (which is content with the initial income distribution) to withdraw the income maintenance schemes and let individuals do their own saving. Some people would be left at low income levels, but a wholesale dependence on social security benefits during old age would be avoided. While potentially a desirable solution within the model, it is difficult to see this response occurring in reality; even the most committed libertarians support some kind of income maintenance (for example, Friedman (1962), Chapter 12). The alternative is to adopt a more full-blooded redistributive policy, designing both taxation and pensions on the basis of a more explicit notion of social welfare. Effects as above can be prevented if revenue is raised by progressive taxes which apply only to the higher income earners in society. For once the rich might actually do better when paying higher taxes, since they are the major losers from a general spread of social security dependence throughout the population (even a first-best income redistribution might be better for them than the outcome above). There is a sense in which the government must decide whether or not to intervene in a decisive way. To implement limited income maintenance suggests some concern for income redistribution without a full commitment to it; the government is satisficing rather than optimising. On the lines of the example above the measures could spread further than intended, leading to an accidental outcome which is not going to be particularly desirable. It would be better for the government either to step back completely from social security provision, or to grasp the nettle and formulate a more ambitious redistributive taxation/pension scheme.

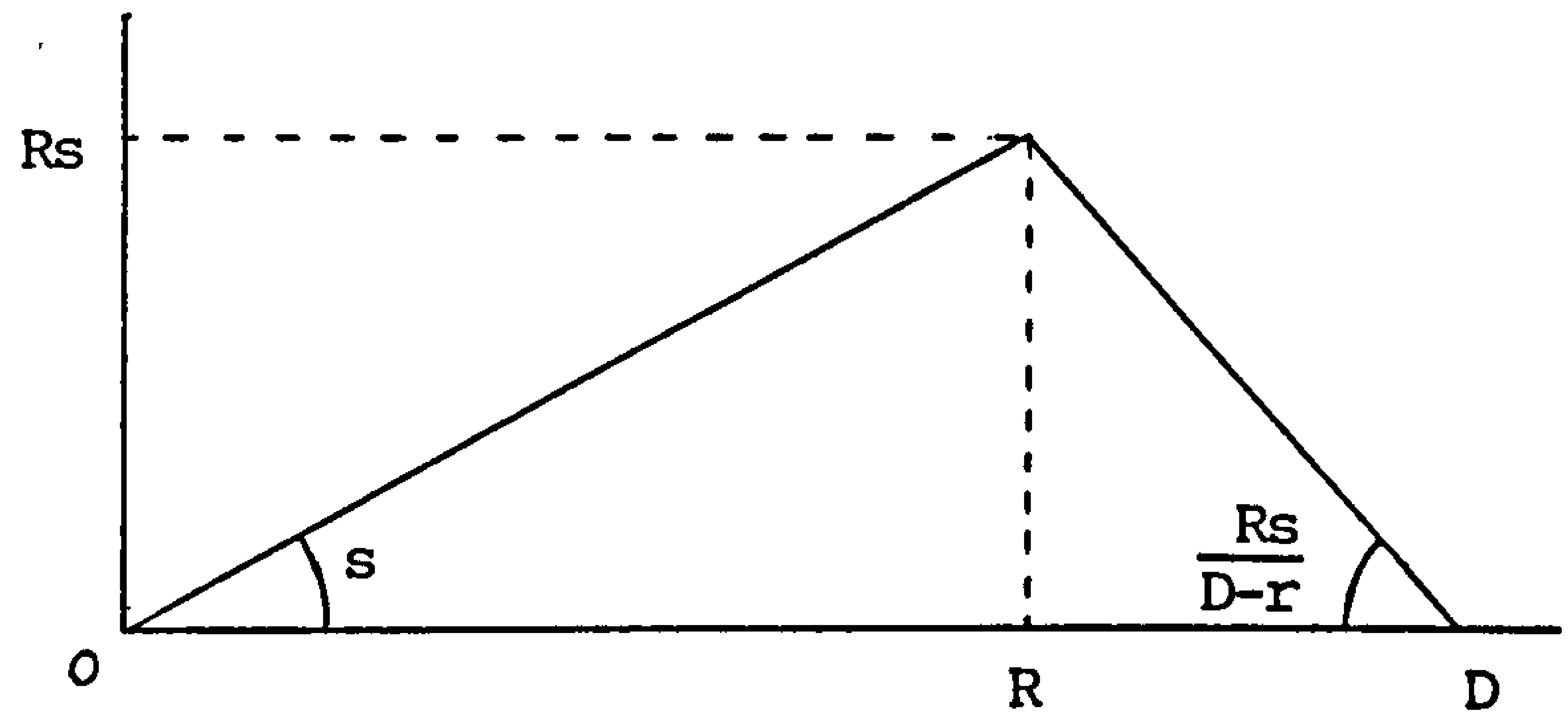
In practice it cannot really be expected that disincentives will be as strong as those illustrated above. Their presence depends on individual rationality, and if saving behaviour is not truly optimal (as may well be the case), the spreading of pension dependence need not arise at all. Other influences may also be relevant. In particular, a social stigma may be attached to the claiming of social security, and some individuals may be reluctant to declare themselves impoverished in old age in order to obtain benefits. If this is true, then benefits are limited to the small number of willing claimants, who receive a net income transfer from the rest of the population. The state can also choose to base its income support on lifetime income, and refuse to maintain individuals who have squandered a high income in their working years. To do this requires the ability to ignore those whose current (but not lifetime) income is genuinely low. The effects depicted here can therefore potentially be avoided in reality. Nevertheless, statistics do show an expansion of social security over time, and especially state pensions (Peacock and Wiseman (1967), Klein and O'Higgins (1985)). An additional feature of U.K. experience has been the breakdown of the insurance/social security division envisaged in the Beveridge Report; income support for pensioners is now dependent on social security payments not originally intended for that purpose. These facts are broadly consistent with the preceding discussion, although they do not perforce result from saving disincentives.

A possible rationale for state intervention now emerges, despite the optimal saving assumption and the absence of redistributive ambitions. It goes on the following lines. Firstly, some degree of income maintenance is always necessary (a fact agreed on by almost all economists), and some of the aged are almost certain to be among the recipients. Given optimal saving and non-redistributive taxation, this leads to saving disincentives in a group much larger than the original targets for income maintenance. The result is that a significant part of the population becomes reliant on a 'pension' which was intended only as minimum income support. One is left with a haphazard system of state pensions, which is unplanned, inadequate and produces an inferior social outcome. It is preferable to accept the need for state intervention and design a more extensive set of redistributive policies (which, at a first best, would not necessarily involve pensions). Although such an argument may seem a little fanciful, it does show that optimal individual saving can be consistent with a case for state intervention and that strong saving disincentives can be used to argue for, as well as against, social security measures.

(b) Retirement Dates and Aggregate Output

The discussion so far has been restricted to individual choice, with no mention of the productive impact of pensions and retirement. It may be that observed policies are determined by productivity considerations, which (justifiably or not) dominate individual preferences; analyses based on these lines are Lazear (1979) and Lapp (1985). In the present context, the relevant issue is whether the government wishes to use pension policies to influence aggregate production. A binding statutory retirement age provides a direct control over total labour supply; if individuals can choose their retirement date, their decision can be influenced by linking pension receipts to retirement. For instance, the Beveridge Report (Beveridge (1942)) insisted on the 'retirement condition', requiring individuals to cease full-time work before they could claim the state pension. The intention was to encourage people to continue work beyond the statutory retirement age, on the assumption that some individuals would prefer an extended working life to early receipt of the pension (though in practice it seems the opposite has occurred, with a large majority of people now retiring at the statutory age). To place an extra constraint on the pension scheme is never desirable from the viewpoint of individual choice. For the government to do so there must be some additional relevant feature, the most likely being a wish to increase the aggregate output level by inducing higher labour force participation.

The position can be illustrated by a simple model which allows individuals to choose their retirement date. It is assumed that people have a known continuous life span from 0 to D, and are able to decide on any retirement date, R, such that $R \varepsilon [0,D]$. Weekly working hours are fixed at T-l (where T is total time available and l is leisure), the wage W is constant over time, and future utility is discounted at a rate ρ. Saving is assumed to be a constant, s, over the working period (which may not be a totally unrealistic view of actual saving behaviour) and accumulated wealth at retirement is consumed at a constant rate. Interest returns on wealth are assumed to be zero in order to simplify the outcome. Individuals therefore adopt a 'pyramidal' form of wealth holding,



which approximates to the usual $\cap$-shape pattern of life-cycle saving. Total utility is given by

$$V = (1-e^{-\rho R})\ U(W(T-1)-s,\ 1) + (e^{-\rho R}-e^{-\rho D})\ \bar{U}(\frac{Rs}{D-R},\ T)$$

where U, $\bar{U}$ are the utility functions pertaining to the employed and retired (not necessarily identical). The individual chooses R and s to maximise V, yielding

$$\frac{\partial V}{\partial R} = \rho e^{-\rho R}(U-\bar{U}) + (e^{-\rho R}-e^{-\rho D})\ \frac{\partial \bar{U}}{\partial x}\ \frac{Ds}{(D-R)^2} = 0$$

$$\frac{\partial V}{\partial s} = -(1-e^{-\rho R})\ \frac{\partial U}{\partial x} + (e^{-\rho R}-e^{-\rho D})\ \frac{R}{D-R}\ \frac{\partial \bar{U}}{\partial x} = 0$$

where x denotes consumption of 'goods' (price normalised to unity). The condition for retirement requires that the utility return from the higher income gained in delaying retirement be equated with the loss from shortening the retirement period. It is noteworthy that voluntary retirement before death occurs only if $\bar{U} > U$, so that the retired have a higher utility than workers. Retirement in practice does not always seem to have this characteristic, suggesting difficulties in picturing actual retirement behaviour as resulting from voluntary choice of this nature. The saving condition implies that the utility loss from a marginal increase in saving during the working period is equated with the marginal gain from increased consumption during retirement.

Suppose now that the government pays out a pension B during the retirement period, financed by a tax of b on the employed. If the pension scheme operates as actuarially fair insurance, then net pension receipts are equal to total tax payments; B and b are linked by the relation

$$B = \frac{Rb}{D-R}$$

Since this has no net effect on lifetime income, it does not influence the total saving level derived from the optimality conditions above. The sole impact is to displace an amount of private saving equal to b, leaving the retirement date unchanged. As was remarked in part (i), there is no resason to implement policies of this type.

In a more general case the tax and pension can be introduced without being linked by a budget constraint, so that B and b each take a fixed level. Individuals under these conditions no longer receive a higher pension when retiring later, and face a loss of pension income when deciding to extend their working life. The retirement condition now takes the form

$$\rho e^{-\rho R}\left[U(W(T-1)-s-b,\ 1) - \bar{U}\left(\frac{Rs}{D-R} + B,\ T\right)\right] + (e^{-\rho R}-e^{-\rho D})\ \frac{\partial \bar{U}}{\partial x}\ \frac{Ds}{(D-R)^2} = 0$$

and is therefore no longer independent of b and B. In general a rise in either b or B will tend to induce earlier retirement and a reduced level of private saving. The government is now in a position to use pensions to influence individual behaviour if it wishes to do so, giving a possible rationale for a state pension scheme. The crucial feature is that net pension receipts are made to be dependent on the individual's retirement decision, and appropriate schemes could take several different forms; a 'retirement condition' fulfils this function if there is a loss of expected income when retirement is delayed.

As was mentioned above, the most likely reason for a divergence between government and individual attitudes is the level of aggregate output. Individual consumption must ultimately be related to production, and it is not assured that individual decisions on retirement (and other issues) must always produce the best possible aggregate output level. In particular, individuals do not take account of the physical link between work and output when making decisions, even though this may be of relevance to them in social welfare terms. It could be that the government has a better knowledge of the economy's aggregate characteristics, allowing it to guide individual decisions towards an outcome superior in social welfare terms. Otherwise the society might, for example, find itself facing restrictions on its consumption that it would prefer not to have and which could be alleviated by an alternative working arrangement. In practice it is likely that some considerations of this type are relevant, since the determinants of national prosperity are not generally known to individuals and do not feature in their decision making. It is not always clear that governments have this knowledge either, but they ought to be in a better position to judge the condition of the national economy. Such notions go well beyond the individualistic framework used here, and would in any case be difficult to depict theoretically in a plausible way. Their conceivable relevance should nevertheless be borne in mind.

In conclusion, Section (3) has identified only three theoretical situations in which pensions are beneficial. The first is where individuals cannot organise their own savings optimally, whether this be due to lack of information or a mere inability to handle finances. A 'paternalistic' state pension is then justifiable. Secondly, it may so happen that a suboptimal pension scheme arises spontaneously out of a limited social security system, because of strong saving disincentives. A purposely designed state pension could improve on this position. The third case is where a state pension is used to influence aggregate labour supply and output, in the interests of attaining a higher social welfare level.

Apart from these cases, the outcome is that rational individuals are capable of looking after their own intertemporal finances, and that pensions are superfluous. The one other reason for having pensions is as a redistributive tool, and this is considered in the following section.

(4) Pensions and Redistribution

Unlike taxation, pension policies are seldom seen as redistributive in purpose. Any pension system must nonetheless entail redistribution unless individual pensions are deliberately designed to preserve lifetime incomes : the general situation is for pensions to be redistributive, with insurance-based non-redistributive pensions as a special case. A truly optimal pension policy would unavoidably encounter redistributive questions, so it is relevant to broaden the discussion sufficiently to include them.

Redistribution through pensions is a more complex issue than in the static models considered in other chapters. In particular, the treatment of pensions requires intertemporal modelling, so that redistribution can occur between different dates and generations as well as between individuals at a single date. With a varying population and uncertainty, the position in reality is far from straightforward. The discussion below separates the 'static' problem of redistributing between members of a given generation from the 'intertemporal' problem of redistributing between individuals of different ages; Section (4) considers the former case and Section (5) the latter. It should be noted that such a distinction does not arise naturally, and that an optimal pension scheme would simultaneously involve both types of redistribution.

In the present section intertemporal redistribution is removed by assuming a constant steady-state population. Each generation is then identical to all other generations, and can be treated as representative of the population at large. Policy measures aim to redistribute income within a particular generation, which need not be dated since all generations are identical. The problem therefore resembles the static redistributions considered in other chapters. Several cases are discussed below under varying sets of assumptions.

(i) Optimal Saving/First-best Income Redistribution

Suppose firstly that individuals follow an optimal savings pattern, and that the government possesses full information on individual characteristics. These are the ideal conditions for undertaking redistribution, providing no obstacle to a first-best outcome. Intervention is necessary only at a single date, involving lump-sum income/wealth transfers on the pattern of Chapter 4, Section (3). Rational individual saving ensures social optimality at all dates, as long as individual preferences coincide with the government's (which can be assumed here). It follows that there is no need for formal pensions, and any retirement period is covered by the individual's private savings.

(ii) Optimal Saving without First-best Income Redistribution

Suppose now that a first-best is not attainable, so that pensions may have some redistributive role. The case considered below involves the linear income and commodity taxes of Diamond (1975), used in Chapter 2, Section (3). Individuals are assumed to live for two periods, working in the first and consuming from savings or pension income in the second. With H individuals and N commodities, individual utilities take the form

$$U = \hat{U}(\hat{x}_1^h, \ldots, \hat{x}_N^h, 1) + \frac{\tilde{U}}{1+\rho}(\tilde{x}_1^h, \ldots, \tilde{x}_N^h, T) \qquad h = 1, \ldots, H$$

where $\hat{U}$, $\tilde{U}$ denote preferences during the working and retirement periods, and consumption levels $\hat{x}_i^h$, $\tilde{x}_i^h$ are differentiated in a similar way. $\tilde{U}$ incorporates a fixed number T of leisure hours per week, and need not imply a utility function identical to $\hat{U}$ if retirement alters the nature of individual preferences; the leisure of the employed, l, is variable and chosen in the conventional way. Government policy is based on linear direct/indirect taxes, $t_1, \ldots, t_N$ (under the usual assumption of fixed producer prices normalised to unity and a tax on employment income normalised to zero), and two uniform lump-sum payments/taxes b,B applying to the employed and retired respectively. A positive payment B can be regarded as a state pension paid to retired individuals. Under optimal saving the timing of the payments b and B is immaterial, and they can be unified as a single payment; in the discussion below this is interpreted as a second-period pension payment, α, such that $\alpha = b(1+r) + B$. The policy problem is to choose t_i, $i=1, \ldots, N$ and α to maximise social welfare subject to a revenue constraint. Such an optimisation is virtually identical in form to the standard case, except for the presence of dated commodities and discounting; the optimality conditions are analogous to those for the single period model. The main feature of interest here is the presence of a redistributive 'pension' payment. Optimal taxation models are most often used to show that commodity/income taxes can serve redistributive ends, given that a lump-sum tax is the better revenue raising tool if redistribution is an issue. The argument works equally well in the reverse direction, justifying government income support either as a pension or in some other guise. Indeed, there is a good chance that it is the uniform lump-sum payment that is responsible for redistribution at the optimum, not the taxes. Commodity taxation is then valuable not as a direct redistributive tool, but as a means of financing a progressive uniform payment (albeit inefficiently), which is the main agent of redistribution. The part played by the uniform lump sum in this case should not therefore be neglected, and it is misleading to view the model as illustrating the effects of taxation alone.

Redistribution can be taken further if the government also taxes retirement incomes. Suppose, for example, that a linear tax at the rate β is imposed on savings transferred into the retirement period (noting that the question of observing the wage rate does not occur among the retired). This can be interpreted in a number of ways: as a direct tax on observed savings; as an indirect interest rate tax, leaving a net return to saving of $r-\beta$; or as a pension inversely related to private means. The model below follows the last of the interpretations, so that individuals receive a uniform lump-sum 'pension', α, in the second period, from which an amount is deducted equal to β times their level of private savings (with α, β free to take negative values). Individual budget constraints take the form

$$\sum_{i=1}^{N} (1+t_i)\left(\hat{x}_i^h + \frac{\tilde{x}_i^h}{1+r-\beta}\right) = W_h(T-l_h) + \frac{\alpha}{1+r-\beta} \qquad h = 1, \ldots, H$$

and the level of savings is equal to first-period income minus consumption, that is

$$W_h(T-l_h) - \sum_{i=1}^{N} \hat{x}_i^h(1+t_i) \qquad i = 1, \ldots, N$$

The government's revenue constraint is therefore

$$\sum_{i=1}^{N} t_i \sum_{h=i}^{H} \left(\hat{x}_i^h + \frac{\tilde{x}_i^h}{1+r}\right) + \frac{\beta}{1+r} \sum_{h=1}^{H} \left(W_h(T-l_h) - \sum_{i=1}^{N} \hat{x}_i^h(1+t_i)\right) = \frac{H\alpha}{1+r} + R$$

In order to undertake policy optimisation it is convenient to write individual utilities as an indirect utility function dependent on prices, wage levels, the rate of interest and the rate of time preference, such that

$$U = U(p, w, r, \rho)$$

where $U = \hat{U} + \tilde{U}/_{1+\rho}$, and $\partial U/_{\partial r} < 0$ will generally hold. Policy optimisation goes ahead as usual, choosing α, β and t_i, $i = 1, \ldots, N$ to maximise social welfare, V, subject to the revenue constraint above.

The Lagrangian and first-order conditions are:

$$L = V + \mu\left[\sum_{i=1}^{N} t_i\left(\sum_{h=1}^{H}\left(\hat{x}_i^h + \frac{\tilde{x}_i^h}{1+r}\right)\right) + \frac{\beta}{1+r}\sum_{h=1}^{H}\left(W_h(T-l_h) - \sum_{i=1}^{N}\hat{x}_i^h(1+t_i)\right) - \frac{H\alpha}{1+r} - R\right]$$

$$\frac{\partial L}{\partial t_k} = \sum_{h=1}^{H}\frac{\partial V}{\partial U_h}\cdot\frac{\partial U_h}{\partial t_k} + \mu\left[\sum_{h=1}^{H}\left(\hat{x}_i^h + \frac{\tilde{x}_i^h}{1+r}\right) + \sum_{i=1}^{N} t_i \sum_{h=1}^{H}\left(\frac{\partial \hat{x}_i^h}{\partial t_k} + \frac{1}{1+r}\frac{\partial \tilde{x}_i^h}{\partial t_k}\right) + \frac{\beta}{1+r}\sum_{h=1}^{H}\left(W_h\frac{\partial l_h}{\partial t_k} + \hat{x}_k^h + \sum_{i=1}^{N}(1+t_i)\frac{\partial \hat{x}_i^h}{\partial t_k}\right)\right] = 0$$

$$k = 1, \ldots, N$$

$$\frac{\partial L}{\partial \alpha} = \sum_{h=1}^{H}\frac{\partial V}{\partial U_h}\cdot\frac{\partial U_h}{\partial \alpha} + \mu\left[\sum_{i=1}^{N} t_i \sum_{h=1}^{H}\left(\frac{\partial \hat{x}_i^h}{\partial \alpha} + \frac{1}{1+r}\frac{\partial \tilde{x}_i^h}{\partial \alpha}\right) + \frac{\beta}{1+r}\sum_{h=1}^{H}\left(W_h\frac{\partial l_h}{\partial \alpha} - \sum_{i=1}^{N}(1+t_i)\frac{\partial \hat{x}_i^h}{\partial \alpha}\right) - \frac{H}{1+r}\right] = 0$$

$$\frac{\partial L}{\partial \beta} = \sum_{h=1}^{H}\frac{\partial V}{\partial U_h}\cdot\frac{\partial U_h}{\partial \beta} + \mu\left[\sum_{i=1}^{N} t_i \sum_{h=1}^{H}\left(\frac{\partial \hat{x}_i^h}{\partial \beta} + \frac{1}{1+r}\frac{\partial \tilde{x}_i^h}{\partial \beta}\right) + \frac{1}{1+r}\sum_{h=1}^{H}\left(W_h(T-l_h) - \sum_{i=1}^{N}\hat{x}_i^h(1+t_i)\right) - \frac{\beta}{1+r}\sum_{h=1}^{H}\left(W_h\frac{\partial l_h}{\partial \beta} + \sum_{i=1}^{N}\frac{\partial \hat{x}_i^h}{\partial \beta}(1+t_i)\right)\right] = 0$$

Let the MSU of income to individual h be denoted by

$$\lambda_h = \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial M_h} + \mu \left[\sum_{i=1}^{N} t_i \left(\frac{\partial \hat{x}_i^h}{\partial M_h} + \frac{1}{1+r} \frac{\partial \tilde{x}_i^h}{\partial M_i} \right) + \frac{\beta}{1+r} \left(W_h \frac{\partial l_h}{\partial M_h} - \sum_{i=1}^{N} (1+t_i) \frac{\partial \hat{x}_i^h}{\partial M_h} \right) \right]$$

which resembles the usual case, except for the presence of dated commodities and the final term reflecting revenue from the interest/retirement income tax (∂M_h denotes a change in first-period income)

From the condition $\partial L / \partial \alpha = 0$,

$$\frac{\mu}{1+r} = \sum_{h=1}^{H} \frac{\lambda_h}{H} = \bar{\lambda}$$

allowing for the fact that α is a second-period income payment.

Modifying Roy's Identity to the present circumstances yields

$$\frac{\partial U_h}{\partial t_k} = \frac{\partial U_h}{\partial M_h} \left(\hat{x}_k^h + \frac{\tilde{x}_k^h}{1+r-\beta} \right) \qquad \begin{matrix} h = 1, \ldots, H \\ k = 1, \ldots, N \end{matrix}$$

and the Slutsky equations take the form

$$\frac{\partial \hat{x}_i^h}{\partial t_k} = \hat{S}_{ik}^h - \left(\hat{x}_k^h + \frac{\tilde{x}_k^h}{1+r-\beta} \right) \frac{\partial \hat{x}_i^h}{\partial M_h}, \quad \frac{\tilde{x}_i^h}{\partial t_k} = \tilde{S}_{ik}^h - \left(\hat{x}_k^h + \frac{\tilde{x}_k^h}{1+r-\beta} \right) \frac{\partial \tilde{x}_i^h}{\partial M_h}$$

Substituting these into the tax equation, rearranging and dividing by H gives

$$\underbrace{\sum_{i=1}^{N} t_i \bar{\hat{S}}_{ik} + \frac{1}{1+r} \sum_{i=1}^{N} t_i \bar{\tilde{S}}_{ik}}_{\text{Compensated reduction in demand from taxing } k} = \underbrace{\frac{1}{H} \sum_{h=1}^{H} \left(\frac{\lambda_h}{\bar{\lambda}} \right) \left(\hat{x}_k^h + \frac{\tilde{x}_k^h}{1+r-\beta} \right) - \left(\frac{1+r-\beta}{1+r} \right) \bar{\hat{x}}_k - \frac{\bar{\tilde{x}}_k}{1+r}}_{\text{Distributional characteristic of } k}$$

Compensated reduction in demand from taxing k

Distributional characteristic of k

$$+ \frac{\beta}{1+r}\left(\sum_{i=1}^{N} (1+t_i)\ \bar{\tilde{S}}_{ik}\right) \qquad k = 1, \ldots, N$$

$$\underbrace{\qquad\qquad\qquad\qquad}_{\text{Savings effect of taxing } k}$$

where a bar denotes averages over h.

This has a broadly similar form to the standard tax equation, except for the final expression. The term on the left-hand side is the compensated reduction in demand resulting from taxing good k; the first term on the right-hand side is the 'distributional characteristic' of the good, that is, the extent to which it is consumed by individuals with a high MSU of income (note, however, that the expression is now influenced by β). The final savings effect indicates the extent to which taxing good k influences compensated final-period consumption expenditure : the more negative it is, the more the tax discourages first-period spending and encourages saving. The compensated reduction in demand is therefore greater for goods which are highly substitutable between time periods, and vice versa. No obvious intuitive label can be placed on such commodities. In general, the fixed leisure hours of the retired should reduce their substitutability in consumption relative to a work force with flexible labour supply, since rationing lessens substitution between commodities (according to the 'le Chatelier Principle' of Samuelson (1947)). From the government's viewpoint the reason for acknowledging the savings effect is entirely in terms of revenue raising, on the grounds that forcing later consumption allows a 'double taxation' of both consumption expenditure and savings.

Let ν_h denote the MSU of a rise in the interest rate as it affects the h^{th} individual, so that

$$\nu_h = \frac{\partial V}{\partial U_h} \cdot \frac{\partial U_h}{\partial r} + \mu\left[\sum_{i=1}^{N} t_i \left(\frac{\partial \hat{x}_i^h}{\partial r} + \frac{1}{1+r}\frac{\partial \tilde{x}_i^h}{\partial r}\right) - \frac{\beta}{1+r}\left(W_h \frac{\partial l}{\partial r} + \sum_{i=1}^{N} (1+t_i)\frac{\partial \hat{x}_i^h}{\partial r}\right)\right]$$

The optimality condition for the interest/retirement income tax then implies that

$$\mu = \bar{\lambda} = \frac{\sum_{h=1}^{H} \nu_h}{\text{Total Savings}}$$

so that the MSU of government revenue and the mean MSU of income to individuals are equated with the sum of the ν_h's divided by total first-period savings. This is intuitively appealing, as the last term describes the MSU of increasing β per unit of savings; at the optimum the MSU of government revenue therefore matches the impact of the equivalent relaxation of the interest tax or rise in the uniform lump-sum payment.

The main importance of this case is that it allows for redistributive measures imposed during the retirement period. A positive α and β produce a progressive redistributive pension, where state pension receipts decline with the extent of private savings. Consequently, in second-best situations optimal pensions may be paid on redistributive principles.

In more complex models there is a clearer distinction between interest taxation and direct taxation of retirement income. Consider, for example, a situation with work and retirement periods spanning more than a single date (as is true in practice). An interest tax applies at the same rate at all dates, regardless of whether individuals are working or retired. A retirement income tax/redistributive pension, on the other hand, depends partly on an individual's age, applying only to the retired (though it may still be beneficial to tax interest or wealth at earlier periods). The tax could also be imposed at a rate varying continuously with age, instead of remaining constant throughout retirement. The informational requirements for a direct tax are greater than for a tax on interest, and if knowledge of retirement incomes is not available an interest tax may be the best feasible policy tool. When full information on incomes is present, it is desirable to use it in imposing a direct tax. The linear structure assumed above is not the ideal form, and an unrestricted non-linear tax structure varying over time would achieve better results. The (already quite complex) structure of the model is capable of substantial extension, to include continuous time, non-linear pensions/taxes in retirement and various different forms of income, wealth and commodity taxation during the working period. This would not alter the main point, however, that a failure to reach a first-best income redistribution among the working population opens the way for redistributive pension schemes.

(iii) Pensions as a Redistributive Tax on Age

A rather different view of pensions emerges when a person's age is seen as an observable characteristic, open to taxation like any other. There is no reason to tax age on its own (under optimal saving), but it may be a proxy for some unobservable characteristic which has distributional relevance. This is a possibility whenever the population varies in a systematic way with age, as a result of different death rates in different groups. If people with a certain characteristic have a lower death rate than those without, then that characteristic occurs in a higher proportion of the population as age increases. It is then possible to tax the characteristic indirectly by imposing age-related payments which increase over time. In models with differing death rates (unlike the fixed life spans above), it is possible to justify pensions as a form of age tax/benefit, where age is correlated with some other unobservable characteristic. Even when the main aim of pensions is not seen in these terms, the effect must influence the policy optimum of any model with varying death rates. Since death rates in practice do vary quite considerably between population groups (Kelsall (1979)), the impact of population 'sorting' over time is almost certain to be relevant to the design of pension schemes.

To illustrate this consider a model with two groups in the population, differentiated by their initial wealth and their death rates. Individuals live for either one or two periods and are assumed to save optimally, maximising expected lifetime utility. The 'wealthy' group has a higher initial capital level $\bar{K}_o^W$ and a mortality parameter θ_W, where θ_W is the chance of a wealthy individual living for two periods. The corresponding values for the 'poor' group are $\bar{K}_o^p$ and θ_p, such that $\bar{K}_o^p < K_o^W$ and $0 < \theta_p < \theta_W < 1$. Expected lifetime utilities are

$$U_i^* = U(C_o^i) + \theta_i U(C_1^i) \qquad i = W,p$$

with all individuals aware of the relevant θ value. The government knows the patterns of wealth and mortality in the population, but for some reason (informational or otherwise) does not tax initial wealth directly. It remains possible to place a uniform lump-sum tax/payment on a person's age, and, given the disparity between θ_W and θ_p, this is a potential tool of redistributive policy.

Let b_t denote the tax payable by individuals of age t, t = 0, 1, so that a negative value represents a 'pension' payment. The revenue constraint is then

$$H\left(b_o + \frac{(\alpha_W\theta_W + \alpha_p\theta_p)}{1+r}\,b_1\right) = R$$

where H is the size of the population, α_W, α_p are the proportions of wealthy and poor individuals ($\alpha_W + \alpha_p = 1$), and R is the revenue requirement. With a utilitarian social objective there is no problem of saving disincentives, since individuals save optimally and taxes take a lump-sum form. Treating C_o^W, C_1^W, C_o^p, C_1^p as policy instruments, the government maximises social welfare subject to the individual budget constraints and the revenue constraint above, yielding the Lagrangian

$$L = H \sum_{i=W,p} \alpha_i \left(U_i^* + \lambda_i\left(\bar{K}_o^i - b_o - \frac{\theta_i b_1}{1+r} - C_o^i - \frac{\theta_i C_1^i}{1+r}\right)\right)$$

$$+ \mu\left(H\left(b_o + \sum_{i=W,p} \frac{\alpha_i \theta_i b_1}{1+r}\right) - R\right)$$

From the first-order conditions for the consumption levels it follows that

$$\frac{\partial U}{\partial c_o^i} = (1+r) \frac{\partial U}{\partial c_1^i} \qquad i = W,p$$

coinciding with the individual optimal saving conditions. Choice of b_o and b_1 gives

$$\mu = \alpha_W \lambda_W + \alpha_p \lambda_p = \frac{\alpha_W \theta_W \lambda_W + \alpha_p \theta_p \lambda_p}{\alpha_W \theta_W + \alpha_p \theta_p}$$

$$\Rightarrow \mu = \lambda_W = \lambda_p$$

The optimum is consequently a first best, equating the MSU of income to the wealthy and poor groups. If preferences are uniform this means an equating of expected lifetime incomes, with tax rates

$$\hat{b}_o = \frac{R}{H} - \left((\alpha_W \theta_W + \alpha_p \theta_p) \frac{(\bar{K}_o^W - \bar{K}_o^p)}{\theta_W - \theta_p}\right)$$

$$\hat{b}_1 = (1+r) \frac{(\bar{K}_o^W - \bar{K}_o^p)}{\theta_W - \theta_p}$$

The second-period tax $\hat{b}_1$ permits differential treatment of the wealthy and poor, owing to their varying survival rates, and is levied at a rate which equates expected incomes. The first period tax $\hat{b}_o$ may well be negative, implying a uniform benefit, unless the revenue requirement is sufficiently high to require taxation in both periods. Redistribution thus involves a tax payment increasing with age, exploiting the greater longevity of the wealthy.

In practice the correlation between wealth and life span is less than perfect, preventing the achievement of a first best : the position then resembles that in Section (2), with longevity as a proxy for income or wealth. It is also likely that mortality is at least partly endogenous, such that life spans converge as wealth becomes more equal. Although changing the details of the model, these effects would, if anything, strengthen the case for redistributive policies, based either on age or initial wealth.

A redistributive age-related tax is potentially justified whenever death rates differ systematically with income or wealth and more direct measures are not feasible. The outcome has little in common with the usual notion of pensions, having no link with retirement and (probably) implying a tax increasing with age. In a more comprehensive model these redistributive considerations would have to be superimposed on the conventional case for benefit payments increasing with age. Even if pensions are not seen as primarily redistributive, some consideration of changing population composition is almost certainly relevant to the design of optimal policy.

(iv) Sub-optimal Savings

With suboptimal saving there is already a possible rationale for a paternalistic pension scheme. Redistributive concerns must stand alongside any paternalism, and may or may not conflict with it.

Introducing suboptimal behaviour can potentially weaken the conclusions of (ii) and (iii) above. The drift of the second-best policy optima derived there is that, in the presence of a failure to achieve a first-best income redistribution, a set of redistributive taxes and payments in the retirement period may be desirable. This gives the impression of being harsh on the retired : in (ii) they may have to pay a tax on their accumulated saving, and in (iii) they face payments increasing with age. Under optimal savings the timing of payments and receipts has little importance, and individuals can readjust their behaviour in response to any net changes in their lifetime income. If a group of individuals follow suboptimal saving, however, they may be more vulnerable to the timing of any taxation. For example, individuals who save in an excessive and inflexible manner are hard hit by taxation of savings; conversely, individuals who do not save sufficiently suffer under an increasing tax on age. Making allowance for such cases in social preferences discourages the imposition of such high redistributive taxes and thus tempers the conclusions obtained. The severity of the effects depends on the initial wealth of the suboptimal groups, attaining a maximum where they are the poorest individuals.

When many varieties of suboptimal saving are present, the ability to redistribute income depends on the relation between income levels and saving behaviour. If the rich save in one way and the poor in another, it should be possible to organise redistributive payments so as to benefit the poor (and in the extreme case observed saving behaviour would fully identify the rich, allowing direct lump-sum taxes to be imposed). Where no such relationship occurs, the effects of policy on particular types of saving behaviour is not concentrated among rich or poor, and it is more difficult to achieve income redistribution. One can say in conclusion that, although redistribution (if anything) strengthens the case for formal pensions under suboptimal saving, the degree of redistribution attainable relies on the forms of saving behaviour present and their relation with the initial income distribution. Possible outcomes range from the first best (or something close to it) to an inability to secure any significant redistribution.

The discussion above suggests that in many cases pensions or age-related payments can be used to redistribute income. The main exception is in the presence of a first-best income distribution and optimal saving, when pensions are not justified on redistributive or any other grounds. Otherwise, under rational saving, three particular possibilities arise:

(a) In combination with optimal commodity and income taxation, a uniform lump-sum payment can have a progressive impact on the income distribution. If paid in the retirement period this may be interpreted as a pension payment.

(b) The scheme in (a) can be extended by allowing a tax to be imposed on retirement incomes or savings (assuming these are observable). If incorporated in a pension scheme, such a tax produces genuinely redistributive pensions, in which receipts decline with the extent of other income. When savings are not directly observed, it may still be possible to achieve similar effects by taxing the interest rate.

(c) Under differing death rates, age-related payments can redistribute lifetime income, irrespective of work and retirement practices. Since death rates usually decline with wealth, the chance is that optimal payments are decreasing.

Reality is inevitably more complex than the models above, with population out of steady state. This leads to the issue of intergenerational transfers, considered in Section (5).

(5) Pensions and Retirement Practices under Population Ageing

Observed populations normally rise or fall over time, experiencing variations in the relative sizes of different age groups. Countries with fast growing populations face 'population youthening', where the average age of the population decreases and the elderly are a falling proportion of the total. Conversely, countries with stagnant or falling populations often face 'population ageing', with a rising average age and an elderly community of growing relative importance. Changes of this nature are relevant to pension policies, since pension payments are frequently financed from the incomes of the working population. An ageing population suggests that the elderly may become an increasing 'burden' on those currently employed.

The U.K. has faced population ageing for almost a century; its implications for pensions were considered in the Beveridge Report, and by several authors in the 1950's (Paish and Peacock (1954), Hopkin (1953), Titmuss (1955)). More recent debate has centred on the 1975 commitment to the 'State Earnings Related Pension Scheme'(SERPS) and its effect on pension payments in the twenty-first century (Ermisch (1981, 1983), Creedy (1982), Hemming and Kay (1982)). This has led to proposals for reducing the provision of SERPS or for abandoning it altogether.

Against this background of policy discussion, it is interesting to consider formal policy optimisation under population ageing. The discussion below sets up a simple theoretical model that permits population ageing, and looks at the retirement and pension practices that arise from social welfare maximisation. Before proceeding it is convenient to distinguish three possible methods of financing.

Insurance Principle (no transfer between generations)

Under conventional insurance principles individuals finance their own pension receipts, with no transfers either from their own generation or from people of different ages (assumed in Sections (2) and (3) above).

This can be relaxed slightly to permit redistribution within a generation (as in Section (4)), but not between individuals of different ages.

'Pay-as-you-go' (limited transfer between generations)

In practice pensions are often financed from current tax revenue, on 'pay-as-you-go' principles. If the government budget approximately balances at each date, there will be a limited income transfer between generations, restricted to those working or retired at the particular date in question. Transfers between individuals with non-overlapping adult life spans do not occur under this system.

'Social Funding' (unlimited transfer between generations)

With 'socially funded'pensions, the government possesses an accumulated stock of pension contributions from which current pension payments are financed. At any date receipts do not have to equal outgoings, allowing transfers to occur between any two generations; the only requirement is that over some long horizon (perhaps infinite) the present value of receipts is equal to that of payments. In this case a 'social fund' is held over many generations, as distinct from the 'individual funding' of insurance based pensions.

By definition the insurance principle rules out intergenerational transfers, and thereby avoids the need to adjust in response. Pensions financed in this way were considered in previous sections. The following discussion concentrates on the implications of 'pay-as-you-go' or 'socially funded' financing.

(i) Pensions Financed by 'Pay-as-you-go' (PAYG)

As Sauvy (1969) points out, there are three main options in adjusting pensions to demographic change : to raise taxes, reduce pensions or delay retirement. Sauvy favours later retirement on the grounds that it minimises the social costs involved. This three way choice is an accurate view of the situation when pensions must be financed from the current incomes of the population, though not necessarily otherwise.

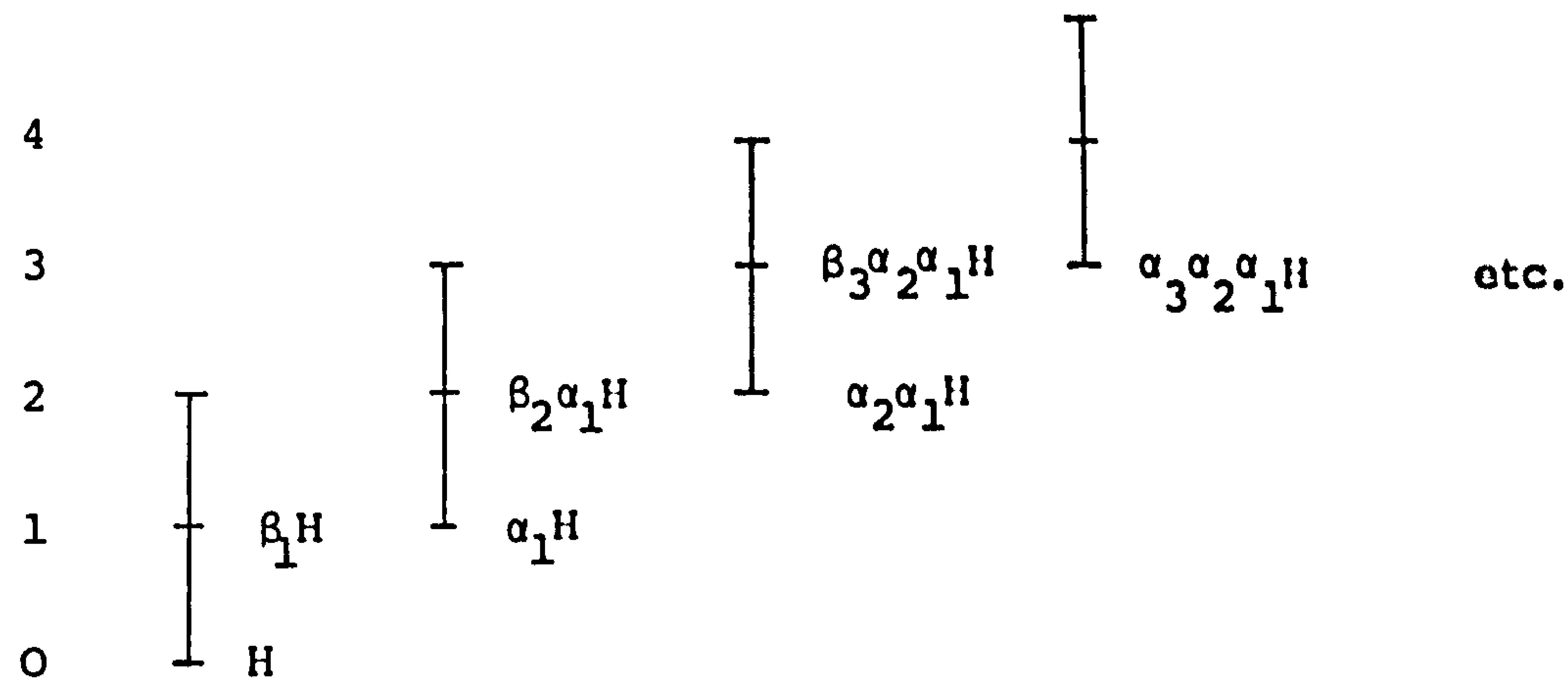
The simplest way to represent this is to have overlapping generations (as initially used by Samuelson (1958)), in which life spans have a 'young' and 'old' period and the old age of one generation coincides with the youth of the next. The discussion below adopts this method, while making a number of additions in order to accommodate population ageing and variable retirement dates.

Individuals are assumed to be identical (so there is no need for redistribution within a generation, as discussed in Section (4)), and to live for up to two discrete periods of equal length. The periods are termed 'youth' and 'old age', and are such that the old age of one generation is simultaneous with the youth of the next. For population ageing to be possible, some assumption must be made about the relative sizes of the different generations. Let fertility and mortality at date t be represented by the parameters α_t, β_t defined such that

α_t = births at date t as a proportion of births at date t-1

β_t = proportion of those born at date t-1 surviving to old age at date t.

As a given generation passes from youth to old age two things happen: a new generation is born which is α_t times the size of the present generation, and a proportion $1-\beta_t$ of the middle-aged generation dies and fails to reach old age. α_t, β_t are assumed to be exogenous and govern the age structure of the population at any date. The position can be depicted as below:



H is the number of births at date 0, so that at date 1 there are $\beta_1 H$ in the old aged group and $\alpha_1 H$ in the young group. Generally, at any date t, the sizes of the two groups are such that

$$\text{Young population} = \left(\prod_{i=1}^{t} \alpha_i\right) H$$

$$\text{Old population} = \beta_t \left(\prod_{i=1}^{t-1} \alpha_i\right) H$$

$$\text{Ratio of old to young at t} = \frac{\beta_t}{\alpha_t}$$

At any given date the age structure of the population is described by the current values of α_t and β_t; the bias towards the elderly is greater the lower the birth rate, α_t, and the higher the survival rate, β_t. It makes no difference which of them is responsible for a changing population structure, so they can be grouped together as a single exogenous expression, $(^{\beta}/\alpha)_t$. In a steady-state population α and β are constants implying an invariant rate of population growth (or decline) and a fixed age structure at each date. This requires no policy adjustments, allowing the same pension to be maintained indefinitely (as, for example, in Sheshinski (1978)). The present discussion depends on the model being out of steady state, and $^{\beta_t}/_{\alpha_t}$ is allowed to vary exogenously over time to produce a changing population structure. The changes are observed by the government, which adjusts its policy measures accordingly.

It is desirable to depict retirement practices in some way, and especially variations in the retirement date. To do so, it is assumed that the population work throughout their period of 'youth' and also for a fixed proportion, γ_t, of their 'old age'. The parameter γ_t is chosen by the government and can be altered to allow for changes in the population structure; this corresponds to having a fixed statutory retirement age, as is indeed usually the case in practice. At any period t the aged population is divided into two groups, with a proportion γ_t working and $1-\gamma_t$ retired. γ_t can be seen as the position of the retirement age within the old age period, a rise in γ representing a later retirement date for all the elderly (although, strictly speaking, it does not have to be interpreted this way). Under these conditions the statistic that matters is not the ratio of the old to the young β/α, but the ratio of the retired to the working population, which can be termed the 'dependence ratio'. This can be expressed as follows

$$\text{Dependence ratio at time t} = \frac{(1-\gamma_t)\beta_t}{\alpha_t + \gamma_t\beta_t} = \frac{1-\gamma_t}{\left(\frac{\alpha}{\beta}\right)_t + \gamma_t}$$

so that the ratio rises with $\left(\frac{\beta}{\alpha}\right)_t$ (but in a lesser proportion) and falls with γ_t. The government's policy problem is to make adjustments to its pension and retirement practices in response to changes in the dependence ratio caused by shifts in the population ageing parameter, β_t/α_t.

In each period a uniform state pension, B_t, is paid to the retired, financed by a uniform tax, b_t, levied on the employed (regardless of their age). Individuals are assumed not to undertake their own saving, relying entirely on the state pension in their retirement; the effects of dropping this assumption are considered below, although it can be defended as a reasonable picture of reality. In the face of changes in β_t/α_t, the government is able to adjust its policy variables B_t, b_t, γ_t at each date t in such a way as to maximise current social welfare. This does not necessarily mean that pension policies are being rapidly altered, and each period can be viewed as a quite substantial stretch of time (say, thirty to forty years). The policy adjustments are thus compatible with a certain stability, as one would expect with pension and retirement conditions. PAYG implies no connection between policy finance at different dates, and policy optimisation at each date is on the basis of the currently living population. Utilitarian social welfare is expressible as

$$V_t = \left[\left(\frac{\alpha}{\beta}\right)_t + \gamma_t \right] U(M-b_t) + (1-\gamma_t)\, \tilde{U}(B_t) \qquad \forall_t$$

where M is the income of the employed, and U, $\tilde{U}$ are the utilities of the employed and retired respectively. Preferences of the employed and retired are different in form, and this can be interpreted as due to the greater leisure of the retired or alternatively caused by a shift in preferences resulting from the different conditions experienced in retirement. The supposition is that for a given income or utility level the retired have a higher marginal utility of income. Under PAYG the government balances its budget at each separate date, yielding a revenue constraint of the form

$$\left[\left(\frac{\alpha}{\beta}\right)_t + \gamma_t \right] b_t = (1-\gamma_t)\, B_t$$

The policy problem is to maximise V_t subject to the revenue constraint above, producing the following Lagrangian and first-order conditions:

$$L_t = V_t + \mu\left[\left(\left(\frac{\alpha}{\beta}\right)_t + \gamma_t\right) b_t - (1-\gamma_t) B_t\right]$$

$$\frac{\partial L}{\partial B_t} = (1-\gamma_t)\frac{\partial \tilde{U}}{\partial M_t} - (1-\gamma_t)\mu_t = 0 \quad \Rightarrow \quad \frac{\partial \tilde{U}}{\partial M_t} = \mu_t$$

$$\frac{\partial L}{\partial b_t} = -\left(\left(\frac{\alpha}{\beta}\right)_t + \gamma_t\right)\frac{\partial U}{\partial M_t} + \left(\left(\frac{\alpha}{\beta}\right)_t + \gamma_t\right)\mu_t = 0 \Rightarrow \frac{\partial U}{\partial M_t} = \mu_t$$

$$\frac{\partial L}{\partial \gamma_t} = U_t - \tilde{U}_t + \mu_t(b_t + B_t) = 0 \quad \Rightarrow \quad \mu_t = \frac{\tilde{U}_t - U_t}{b_t + B_t}$$

There are no behavioural disincentives, so the optimum successfully equates the marginal utilities of income for the employed and retired.

Since $\mu_t > 0$, the γ_t condition can be satisfied only if $\tilde{U}_t > U_t$, that is, if the utility of the retired exceeds that of the employed. Otherwise there would be no incentive to have retirement, and the optimum would be a corner solution in which $\gamma = 1$ and everyone works for their full lifetime. This utility gain is to be expected, as retirement involves a sacrifice of income (both for the individual and society) and will be undertaken voluntarily only if it brings some non-monetary utility return. Retirement therefore has to be a pleasant experience, superior to the working alternative for it to be justified as the outcome of social or individual choice.

It is assumed here that there is some benefit derived from retiring, and that the government does implement a retirement date before the end of the old age period. The fact that $\tilde{U}_t > U_t$ means that individuals cannot be allowed to choose their own retirement date for given B_t, b_t values since all would choose the earliest possible retirement and the system would collapse. Individual choice of retirement dates is feasible only if pension receipts are more closely linked to contributions so that retirement automatically means a drop in the pension received. Adjustments in the retirement date can then be achieved by manipulating individual choices through the pension scheme, rather than by varying a statutory retirement date.

The main item of interest is the optimal response to population ageing, in other words, the comparative static effects on B, b and γ of a shift in α/β. Let α/β fall marginally, implying that society faces an ageing population structure. Inspection of the first-order conditions shows that both α/β and γ appear only in the revenue constraint, and that the three marginal conditions are soluble for B, b, μ. The optimality conditions can be preserved by holding B_t, b_t constant and making a compensating adjustment in γ_t to hold revenue constant. Specifically, it must be true that

$$d\left(\frac{\alpha}{\beta}\right) b + d\gamma\,(B+b) = 0 \qquad \Rightarrow \qquad \frac{d\gamma}{d(\alpha/\beta)} = \frac{-b}{B+b} < 0$$

so that a fall in α/β produces a rise in γ and vice versa. The best policy to adopt under population ageing is therefore to increase the retirement age, while maintaining the same level of pensions (and vice versa for population youthening). This has a fairly clear intuitive rationale, since at any given date changes in the ratio of contributors to retired in the budget constraint are automatically balanced by the same changes in the social welfare function, irrespective of the values of α/β and γ; the only adjustments needed are to satisfy the revenue constraint, and these can be limited to compensating movements in α/β and γ. Society consequently fixes a constant real level of pension provision and makes any appropriate movements in the retirement age needed to finance pensions at each date.

The outcome confirms Sauvy's conclusion on the desirability of raising the retirement age under population ageing, but it is at odds with much recent commentary on pensions, which often argues for cuts in pension provision. In the present model the latter views can be optimal only under the assumption of a fixed retirement date, leaving a fall in pensions and rise in taxes as the only means of preserving the revenue constraint. They might also arise if the commentator sees current pensions as suboptimal and too generous, requiring a cut in pensions as a movement towards an optimal position. Much of the concern over pension finance is based on the possible disincentive effects of raising future taxation levels. Paish and Peacock (1954) discuss the 'transfer problem' of having to shift increasing resources from the working to the retired population; among their suggested solutions are more flexible forms of retirement, encouraging later retirement dates. Such conclusions accord with the analysis above, but for the wrong reason. In the present model later retirement is intrinsically desirable, and the question of tax disincentives is never raised

The redistributive effect of PAYG is to benefit those generations which are small relative to the generation following; it is consequently in a given generation's interest to more than reproduce itself and to have high mortality at early ages (for those who survive, at any rate). Under the optimal policy this allows them to retire earlier than other generations, and, as retirement brings a utility return in this model, they have higher lifetime utility. The position is in line with the 'relative generation size' hypothesis used in the forecasting and explanation of population trends. This assumes that small generations always benefit throughout their life cycle from smaller family size, better employment prospects, faster promotion, etc. (Easterlin (1980)). Without the transference of wealth over many generations, it is difficult to see how these effects can be prevented from spilling over into pensions policies, even at the policy optimum.

(ii) Pensions Financed by 'Social Funding'

An alternative approach is to set aside a special fund to cover all pension payments. At any given time the employed contribute to the fund and pensioners draw from it, subject to the sole requirement that the fund breaks even over some long future horizon (possibly infinite). The government is then acting as if it were a single long-lived individual, making rational saving decisions to maximise social welfare. Where such a scheme is feasible, the outcome must always be preferable to PAYG, which is an arbitrarily restrictive special case.

The model assumes a uniform population, so the contribution of social funding turns on redistribution between generations. This can only be achieved when it is possible to forecast population changes over some future period, the longer the better. Unlike (i), the discussion below assumes that the government can forecast accurately the relevant population structure over some given period. Let T denote the planning horizon over which policy is formulated; both social welfare and the revenue constraint are now defined over the full T periods, rather than for each separate date. A single formal retirement date, γ, is chosen optimally for the whole planning period, along with variable tax and pension levels. With no discounting of future utility (as seems reasonable), social welfare is the sum of current utilities in the population for all relevant dates, that is

$$V = \sum_{t=1}^{T} \left(\prod_{j=1}^{t-1} \alpha_j\right) \left((\alpha_t+\beta_t\gamma)(U(M-b_t) + \beta_t(1-\gamma)\tilde{U}(B_t)\right) H$$

where H is the population at date zero and M is assumed constant (an assumption dropped in Section (4)). The government is content to weight each generation by population size and does not seek to give more equal weightings to generations by using average utility measures. The corresponding revenue constraint takes a similar form, such that

$$\sum_{t=1}^{T} \frac{\left(\prod_{j=1}^{t-1} \alpha_j\right)}{(1+r)^t} \left((\alpha_t+\beta_t\gamma)b_t - \beta_t(1-\gamma)B_t\right) H = 0$$

where r is the constant interest rate received by the government. Maximising V subject to the single constraint above yields the first-order conditions

$$\left(\prod_{j=1}^{t-1} \alpha_j\right) \beta_t(1-\gamma) \left(\frac{\partial\tilde{U}}{\partial B_t} - \frac{\mu}{(1+r)^t}\right) = 0 \quad \Rightarrow \quad \frac{\partial\tilde{U}}{\partial B_t} = \frac{\mu}{(1+r)^t}$$

$$\left(\prod_{j=1}^{t-1} \alpha_j\right) (\alpha_t+\beta_t\gamma) \left(- \frac{\partial U}{\partial b_t} + \frac{\mu}{(1+r)^t}\right) = 0 \quad \Rightarrow \quad - \frac{\partial U}{\partial b_t} = \frac{\mu}{(1+r)^t}$$

$$\sum_{t=1}^{T} \left(\prod_{j=1}^{t-1} \alpha_j\right) \beta_t \left((U_t-\tilde{U}_t) + \frac{\mu}{(1+r)^t} (B_t+b_t)\right) = 0$$

As under PAYG, γ drops out of the 2T+1 equations above, which are soluble for B_t, b_t, t=1, ..., T, and μ. The marginal utilities of income for the employed and retired are equated, implying a first-best income redistribution. The ability to transfer incomes freely between dates means that the only influence on the time structure of pension payments is the interest rate. If $r < 0$, then marginal utilities of income are falling and incomes are rising: consequently pensions increase over time and taxes fall. The government in these circumstances builds up a positive pension fund, which subsidies later generations at the expense of earlier ones. If $r \leq 0$, then transferring wealth between dates brings no net income gain (a loss if $r < 0$) and is counter-productive; the government then reverts to PAYG, as in (i). Optimal choice of γ means that the aggregate utility loss from retirement in the planning period is balanced by the revenue gain. The particular value of γ is derived from the revenue constraint, such that

$$\gamma = \frac{\sum_{t=1}^{T}\left(\left(\prod_{j=1}^{t=1}\alpha_j\right)/(1+r)^t\right)\left(\alpha_t\hat{b}_t-\beta_t\hat{B}_t\right)}{\sum_{t=1}^{T}\left(\left(\prod_{j=1}^{t-1}\alpha_j\right)/(1+r)^t\right)\beta_t\left(\hat{B}_t+\hat{b}_t\right)}$$

where $\hat{B}_t$, $\hat{b}_t$ are the optimal values derived from the first-order conditions.

Social funding has the usual characteristics of optimal saving, where events within the planning period are allowed for in the initial decision. Population ageing is not an active policy issue, except for its role in the initial planning of policies. At the extreme of perfect foresight over an infinite horizon there is no need to respond to population ageing at all, and the entire future course of policy is determined at the initial date. Otherwise there remains a need for recalculation of policy plans, as the current horizon is reached or new forecast information is received; this would necessitate policy adjustments, as in (i) above. Reality probably comes between the extremes, where the feasible planning horizon is neither zero (as in PAYG), nor certain and infinite (as in idealised social funding).

(iii) Extensions to the Basic Model

Two extensions are considered here: the impact of rising incomes and productivity growth, and the presence of private saving. The discussion below introduces these into PAYG and socially funded models.

(a) Technical progress and productivity growth

Since the time periods being considered are quite lengthy, it is possible that changes in technology and productivity are taking place alongside population change. All being well these will involve a secular rise in productivity, leading to rising real incomes among the employed. They can be represented straightforwardly by assuming an exogenous rise in the general level of incomes, M.

Consider firstly a PAYG system. The impact on the optimum of the changes dM and $d(\alpha/\beta)$ can be identified by totally differentiating the first-order conditions in (i) and solving out. This yields expressions

$$dB = \underbrace{\frac{-U_M}{\tilde{U}_{MM}(B+b)}}_{+\text{ ve}} dM \quad ; \qquad db = \left(\underbrace{\frac{U_M}{U_{MM}(B+b)}}_{-\text{ ve}} + 1 \right) dM$$

$$d\gamma = \frac{-b}{(B+b)} d(\alpha/\beta) + \frac{1}{(B+b)} \left(\underbrace{\frac{-U_M(1-\gamma)}{\tilde{U}_{MM}(B+b)}}_{+\text{ ve}} - ((\alpha/\beta)+\gamma) \left(\underbrace{\frac{U_M}{U_{MM}(B+b)}}_{-\text{ ve}} + 1 \right) \right) dM$$

Rising incomes unambiguously produce rising pensions, with $dB/dM > 0$. The change in the tax rate can be positive or negative, but it is guaranteed that taxes cannot rise sufficiently to offset the rise in income, so the employed still experience rising utility. The retirement date continues to respond negatively to α/β, implying that, ceteris paribus, population ageing encourages later retirement. If, however, M also rises with $^{\alpha}/_{\beta}$, then the net impact on the retirement date can go either way, depending on the relative size of the terms in the $d\gamma$ expression; comparison with the equation for db shows that γ can fall only when b is rising. It is thereby conceivable (though not necessarily likely) to have earlier retirement alongside population ageing at the social optimum when economic growth is present.

Under social funding known changes in future income can be allowed for within a given planning horizon. The optimality conditions take the same form as in (ii), the pension level must be rising, and the retirement age is constant within the period. The time path of b_t depends on individual preferences and the nature of income growth.

(b) Individual saving

When individuals undertake private saving their behaviour must be included in the government's policy formulation if a true social optimum is to be achieved. The central issue is whether or not individual behaviour responds to government policy, leading to saving disincentives. If so, the nature of policy adjustments to ageing may become considerably more complex.

To illustrate this, consider a PAYG financed model without the zero private saving assumption made above. Suppose instead that individuals make saving decisions in their 'youth' and 'old age' which react to current policy variables; in other words, $s_y = s_y(B, b, \gamma)$, $s_o = s_o(B, b, \gamma)$, where s_y, s_o are the saving rates from employment in the 'youth' and 'old age' periods. If savings are consumed at a constant rate in retirement, the utility levels of a person retired at date t are $U(M-b_{t-1}-s_y(B_{t-1},b_{t-1},\gamma_{t-1}))$ in youth, $(U(M-b_t-s_o(B_t,b_t,\gamma_t))$ when employed in old age, and $\tilde{U}(B_t+((1+r)\bar{s}_y+\gamma_t s_o)/(1-\gamma_t))$ when retired, where $\bar{s}_y$ is saving inherited from youth. The government is assumed to know the nature of individual savings, and as before aims to maximise the total utility of the current population. Policy is set to maximise

$$V_t = (\alpha/\beta)_t \, U_t^y + \gamma_t U_t^o + (1-\gamma_t)\tilde{U}_t$$

subject to the budget constraint of (i). The resulting first-order conditions imply that

$$\mu_t = \frac{\partial \tilde{U}}{\partial B_t} - \frac{1}{1-\gamma_t}\left(\gamma_t\left(\frac{\partial U^o}{\partial M_t} - \frac{\partial \tilde{U}}{\partial M_t}\right)\frac{\partial s_o}{\partial B_t} + \left(\frac{\alpha}{\beta}\right)_t \frac{\partial U^y}{\partial M_t}\frac{\partial s_y}{\partial B_t}\right)$$

$$= \frac{1}{(\alpha/\beta)_t + \gamma_t}\left(\left(\frac{\alpha}{\beta}\right)_t \frac{\partial U^y}{\partial M_t}\left(1 + \frac{\partial s_y}{\partial b_t}\right) + \gamma_t \frac{\partial U^o}{\partial M_t} + \gamma_t\left(\frac{\partial U^o}{\partial M_t} - \frac{\partial \tilde{U}}{\partial M_t}\right)\frac{\partial s_o}{\partial b_t}\right)$$

$$= \frac{1}{B_t + b_t}\left(\tilde{U}_t - U_t^o + \gamma_t\left(\frac{\partial U^o}{\partial M_t} - \frac{\partial \tilde{U}}{\partial M_t}\right)\frac{\partial s_o}{\partial \gamma_t} + \left(\frac{\alpha}{\beta}\right)_t \frac{\partial U^y}{\partial M_t}\frac{\partial s_y}{\partial \gamma_t}\right)$$

where μ_t is again the multiplier on the revenue constraint. Compared with (i), all these equations are complicated by the addition of extra terms representing movements in private saving. The main point to note is that the ageing parameter α/β and the retirement date γ now play an integral part in the determination of the tax and pension levels. Shifts in α/β can therefore be expected to produce changes in all three policy variables, in contrast with (i). Nothing can be said in general about the direction of the movements involved, which depend on the form and strength of private saving adjustments. The optimum could in principle differ substantially from that obtained previously, although it will not necessarily do so. Identical observations are applicable to socially funded pensions.

Whether or not the earlier outcomes still obtain depends on the stability of the s_y and s_o values. If private saving is inflexible or small in scale (as, say, in the 'classical savings postulate' that workers do not save), then saving behaviour becomes negligible and the conclusions of (i) will hold. Under stronger saving responses (such as optimal life-cycle saving) the position is less clear, although the pattern of (i) remains the 'desired' outcome to be overturned.

Consequently, economic growth or private saving can alter the results of (i) or (ii). No definite conclusions can be reached in these cases without specific information on preferences and individual behaviour.

In considering the effects of population ageing on optimal pension policies two main cases have been identified. When the pension scheme must balance at each separate date the preference is for an accommodating rise in the retirement date, keeping pensions constant. When the pension scheme needs only to balance over a much longer period, the impact of population ageing is absorbed into the society's total expected income for the period and there is no need to make any specific responses to it; policy can then take the form of rising pensions (given positive real interest rates) coupled with a fixed retirement age. Under ideal conditions it follows that the need for specific responses to population ageing can be avoided. But if pension finance is constrained to depend on the current population structure, then it should react to population ageing by raising the retirement age and not by cutting pensions or raising taxes.

Qualifications to these conclusions are manifold. As was noted above, the findings do not necessarily apply when there is secular income growth or when private saving disincentives are present. Other relevant features are the institutional aspect, the uncertainty attached to forecasting, and the question of social attitudes to retirement. With regard to actual policy recommendations, the position depends on whether current pension policies are viewed as an optimum resembling those in the models. If not, then the conclusions obtained cannot be translated directly into practice. Any policy recommendations require an analysis of the particular pension scheme under consideration.

(6) Conclusion

With a rational population and individualistic social welfare, the immediate case for formal pensions and retirement is by no means clear. Intervention is desirable only in two main sets of circumstances.

The first is where individual behaviour diverges from social optimality. Variants of this identified above are:

(i) Pure paternalism, where the government knows what is best for individuals and/or is better placed to implement it.

(ii) Cases where productive issues are important (in addition to individual utilities), and are not properly allowed for in individual decision making.

(iii) Use of pensions for redistribution, where a first-best income distribution is not attained.

(iv) Transfer of resources to future generations : there may be a case for a social pension fund, ensuring an indefinitely growing capital stock.

Policies of this type all have a positive intent, geared to some social objective not otherwise attainable.

The other possibility is faute de mieux, as in Section (3), part (a). If moral hazard leads to widespread social security claims in old age, then the outcome is state 'pensions' at a low level of provision. The government may instead prefer to introduce formal redistributive pensions at a higher level.

Little can be said in general about the nature of optimal pensions and retirement. On grounds of individual choice there seems to be a good case for flexible and gradual retirement, with pensions (if any) adjusted accordingly. This at least is what emerges from utility based models, as in Section (2) above. Some caution is needed in relating this to reality, however, given that work/leisure preferences do not cover all aspects of retirement decisions. It remains possible that constraints on individual choice are warranted by productive considerations excluded from the standard formulations of welfare economics. To make specific policy recommendations it would be necessary to look more closely at the particular population and economy in hand.

CHAPTER 8 : GENERAL CONCLUSION

The relevant concluding comments have already been appended to the separate chapters above; there is little to be added in the way of a general conclusion. It will suffice to reiterate a couple of the issues raised in the Introduction.

The first point arises from the scarcity of general properties derivable from theoretical modelling of optimal policy; outcomes nearly always depend on assumptions made about preferences or other features of the model (as, for example, in optimal tax theory). This emphasises the importance of personal judgement and subjective factors in assessing policy. Such factors appear at two levels in the present context. At one level they are built into the model, in the form of individual preferences. The conclusions reached depend on what preferences are like in the case being considered, or, rather, on what we think they are like. At the higher level there is a judgement to be made about the model itself. It must be felt to be providing an acceptable picture of issues which are significant in reality. If this is so, then the model may be of assistance in forming opinions about policies. But its role is always as an aid to personal judgement, not a substitute for it.

Exercising a little judgement leads on to the second point. It seems that there is a considerable gap between the policies observed in reality and those imaginable as in any circumstances 'optimal'. This observation, if accepted, can produce a multitude of alternative conclusions. It can lead to calls for policy reform, scepticism about the impartiality of government, doubts about the quality of the model, advocacy of a change of economic system, accedence to the result of a 'democratic' process, and so on. None of these is beyond question, and the conclusion selected is entirely a matter of opinion.

Bibliography

Addison, J.T. and W.S. Siebert (1979), 'The Market for Labour: an Analytical Treatment', Goodyear, Santa Monica, California.

Akerlof, G.A. (1978), 'The Economics of Tagging', American Economic Review, 68, 8-19.

Ando, A. and F. Modigliani (1963), 'The Life-Cycle Hypothesis of Saving: Aggregate Implications and Tests', American Economic Review, 53, 55-84.

Armstrong, H. and J. Taylor (1985), 'Regional Economics and Policy', Philip Allan, Oxford.

Arrow, K.J. (1951), 'Social Choice and Individual Values', Wiley, New York.

Atkinson, A.B. (1971a), 'Capital Taxes, the Redistribution of Wealth and Individual Savings', Review of Economic Studies, 38, 209-228.

Atkinson, A.B. (1971b) 'The Distribution of Wealth and the Individual Life Cycle', Oxford Economic Papers, 23, 239-254.

Atkinson, A.B. (1972), 'Unequal Shares', Allen Lane, London.

Atkinson, A.B. (1983), 'The Economics of Inequality', Oxford University Press, Oxford.

Atkinson, A.B. and A.J. Harrison (1978), 'The Distribution of Personal Wealth in Britain', Cambridge University Press, London.

Atkinson, A.B. and N.H. Stern (1981), 'On Labour Supply and Commodity Demands', in A.S. Deaton (ed.), 'Essays in the Theory and Measurement of Consumer Behaviour', Cambridge University Press, Cambridge.

Atkinson, A.B. and J.E. Stiglitz (1980), 'Lectures on Public Economics', McGraw-Hill, London.

Atkinson, A.B. et al. (1984), 'Unemployment Benefit, Duration and Incentives in Britain: How Robust is the Evidence?', Journal of Public Economics, 23, 3-26.

Aumann, R.J. and M. Kurz (1977), 'Power and Taxes', Econometrica, 45, 1137-1161.

Baily, M.N. (1978), 'Some Aspects of Optimal Unemployment Insurance', Journal of Public Economics, 10, 379-402.

Batchelor, R.A. and T.D. Sherriff(1980), 'Unemployment and Unanticipated Inflation in Postwar Britain', Economica, 47, 179-192.

de Beauvoir, S.(1970),'Old Age', Penguin, Harmondsworth.

Becker, G.S. (1965), 'A Theory of the Allocation of Time', Economic Journal, 75, 493-517.

Becker, J. (1972), 'Experience Rating in Unemployment Insurance', Johns Hopkins University Press, Baltimore.

Beckerman, W. (1979), 'Does Slow Growth Matter? Egalitarianism versus Humanitarianism', in W. Beckerman (ed.), 'Slow Growth in Britain; Causes and Consequences', Oxford University Press, Oxford.

Benjamin, D.K. and L.A. Kochin (1979), 'Searchingfor an Explanation of Unemployment in Interwar Britain', Journal of Political Economy, 87, 441-478.

Beveridge, W.H. (1942), 'Social Insurance and Allied Services: Report'.

Blinder, A.S. and H.Y. Esaki (1978), 'Macroeconomic Activity and Income Distribution in the Postwar United States', Review of Economics and Statistics, 60, 604-609.

Blinder, A.S. and Y. Weiss (1976), 'Human Capital and Labour Supply: a Synthesis', Journal of Political Economy, 84, 449-472.

Bosanquet, N. (1983), 'After the New Right', Heinemann, London.

Buchanan, J.H. (1968), 'Social Insurance in a Growing Economy: a Proposal for Radical Reform', National Tax Journal, 21, 386-395.

Buchanan, J.M. and G. Tullock (1962), 'The Calculus of Consent', University of Michigan Press, Ann Arbor, Michigan.

Cigno, A. (1983), 'On Optimal Family Allowances', Oxford Economic Papers, 35, 13-22.

Creedy, J. (1982), 'State Pensions in Britain', Cambridge University Press, Cambridge.

Cubbin, J.S. and K. Foley (1977), 'The Extent of Benefit Induced Unemployment in Great Britain: Some New Evidence', Oxford Economic Papers, 29, 128-140.

Dasgupta, P. and P.J. Hammond (1980), 'Fully Progressive Taxation', Journal of Public Economics, 13, 141-154.

Davidson, P. (1962), 'Income and Employment Multipliers and the Price Level', American Economic Review, 52, 738-752.

Davidson, P. and E. Smolensky, (1964), 'Aggregate Supply and Demand Analysis', Harper and Row, New York.

Deaton, A.S. (1981a), 'Optimal Taxes and the Structure of Preferences', Econometrica, 49, 1245-1260.

Deaton, A.S. (1981b), 'Theoretical and Empirical Approaches to Consumer Demand under Rationing', in A.S. Deaton (ed.), 'Essays in the Theory and Measurement of Consumer Behaviour', Cambridge University Press, Cambridge.

Deaton, A.S. and J. Muellbauer (1980), 'Economics and Consumer Behaviour', Cambridge University Press, Cambridge.

Deaton, A.S. and N.H. Stern (1986), 'Optimally Uniform Commodity Taxes, Taste Differences and Lump-sum Grants', Economics Letters, 20, 263-266.

Diamond, P.A. (1975), 'A Many-person Ramsey Tax Rule', Journal of Public Economics, 4, 335-342.

Diamond, P.A. (1981), 'Mobility Costs, Frictional Unemployment and Efficiency', Journal of Political Economy, 89, 798-812.

Diamond, P.A. and J.A. Mirrlees (1978), 'A Model of Social Insurance with Variable Retirement', Journal of Public Economics, 16, 295-336.

Disney, R. (1979), 'Recurrent Spells and the Concentration of Unemployment in Great Britain', Economic Journal, 89, 109-119.

Disney, R. (1981), 'Unemployment Insurance in Britain', in J. Creedy (ed.) 'The Economics of Unemployment in Britain', Butterworths, London.

Doeringer, P.B. and M. Piore (1971), 'Internal Labour Markets and Manpower Analysis', D.C. Heath, Lexington, Massachusetts.

Duesenberry, J.S. (1949), 'Income, Saving and the Theory of Consumer Behaviour', Harvard University Press, Cambridge, Massaschusetts.

Dunlop, J.T. (1938), 'The Movement of Real and Money Wage Rates', Economic Journal, 48, 413-434.

Easterlin, R.A. (1980), 'Birth and Fortune, the Impact of Numbers on Personal Welfare', Grant McIntyre, London.

Eichner, A.S. (1979), 'A Guide to Post-Keynesian Economics', Macmillan, London.

Ermisch, J.F. (1981), 'Paying the Piper: Demographic Change and Pension Contributions', Policy Studies, 1, 213-220.

Ermisch, J.F. (1983), 'The Political Econmy of Demographic Change', Heinemann, London.

Feldstein, M.S. (1974), 'Social Security, Induced Retirement and Aggregate Capital Accumulation', Journal of Political Economy, 82, 551-573.

Feldstein, M.S. (1977), 'Social Security and Private Savings: International Evidence in an Extended Life-Cycle Model', in M.S. Feldstein and R.P. Inman (eds.), 'The Economics of Public Services', Macmillan, London.

Flemming, J.S. (1978), 'Aspects of Optimal Unemployment Insurance: Leisure, Savings and Capital Market Imperfections', Journal of Public Economics, 10, 403-425.

Flemming, J.S. and I.M.D. Little (1974), 'Why We Need a Wealth Tax', Methuen, London.

Friedman, M. (1957), 'A Theory of the Consumption Function', Princeton University Press, Princeton, New Jersey.

Friedman, M. (1962), 'Capitalism and Freedom', University of Chicago Press, Chicago.

Galbraith, J.K. (1958), 'The Affluent Society', Houghton Mifflin, Boston.

Galbraith, J.K. (1967), 'The New Industrial State', Houghton Mifflin, Boston.

Ghez, G.R. and G.S. Becker (1975), 'The Allocation of Time and Goods over the Life Cycle', NBER, New York.

Gordon, D.M. (1972), 'The Theory of Poverty and Underemployment', D.C. Heath, Lexington, Massachusetts.

Gordon, R.A. and A.S. Blinder (1980), 'Market Wages, Reservation Wages and Retirement Decisions', Journal of Public Economics, 14, 227-308.

Gorman, W.M. (1976), 'Tricks with Utility functions', in M. Artis and A.R. Nobay (eds.), 'Essays in Economic Analysis', Cambridge University Press, Cambridge.

Gramlich, E.M. (1974), 'The Distributional Effects of Higher Unemployment', Brooking Papers on Economic Activity.

Green, G.F. (1981), 'The Effect of Occupational Pension Schemes on Saving in the United Kingdom: a Test of the Life Cycle Hypothesis', Economic Journal, 91, 130-144.

Grossman, S.J. and O.D. Hart (1983), An Analysis of the Principal-Agent Problem', Econometrica, 51, 7-45.

Guesnerie, R. and K.W.S. Roberts (1984), 'Effective Policy Tools and Quantity Controls', Econometrica, 52, 59-86.

Gujarati, D. (1972), 'The Behaviour of Unemployment and Unfilled Vacancies: Great Britain 1958-71', Economic Journal, 82, 195-204.

Hahn, F.H. (1982), 'On Some Difficulties of the Utilitarian Economist', in A.K. Sen and B. Williams (eds.), 'Utilitarianism and Beyond', Cambridge University Press, Cambridge.

Hamermesh, D.S. and A. Rees (1984), 'The Economics of Work and Pay', Harper and Row, New York.

Hammond, P.J. (1976), 'Endogenous Tastes and Stable Long Run Choice', Journal of Economic Theory, 13, 329-340.

Harbury, C.D. and Hitchens, D.M.W.N. (1979), 'Inheritance and Wealth Inequality in Britain', George Allen and Unwin, London.

Harrison, A. (1979), 'The Distribution of Wealth in Ten Countries', Royal Commission on the Distribution of Income and Wealth.

Harrod, R.F. (1948), 'Towards a Dynamic Economics', Macmillan, London.

Hayek, F.A. (1960), 'The Constitution of Liberty', Routledge and Kegan Paul, London.

Hayes, J. and P. Nutman (1981), 'Understanding the Unemployed: the Psychological Effects of Unemployment', Tavistock, London.

Heckman, J.J. (1976), 'A Life Cycle Model of Earnings, Learning and Consumption', Journal of Political Economy, 84, 511-549.

Hemming, R. and J.A. Kay (1982), 'The Costs of the State Earnings Related Pension Scheme', Economic Journal, 92, 300-319.

Hey, J.D. and K.G. Mavromaras (1981), 'The Effect of Unemployment Insurance on the Riskiness of Occupational Choice', Journal of Public Economics, 16, 317-341.

Hirsch, F. (1977), 'The Social Limits to Growth', Routledge and Kegan Paul, Henley-on-Thames.

Hopkin, W.A.B. (1953), 'The Economics of an Ageing Population', Lloyds Bank Review, 27, 25-36.

Jahoda, M. (1982), 'Employment and Unemployment: a Social-Psychological Analysis', Cambridge University Press, London.

Joseph, G. (1983), 'Women at Work', Philip Allen, Oxford.

Kalecki, M. (1971), 'Selected Essays on the Dynamics of the Capitalist Economy', Cambridge University Press, Cambridge.

Kay, J.A. (1985), 'Incomes of the Elderly: the Future of State Provision', in R. Berthoud (ed.), 'Challenges to Social Policy', Gower, Aldershot.

Kay, J.A. and M.A. King (1983), 'The British Tax System', Oxford University Press, Oxford.

Kelsall, R.K. (1979), 'Population', Longman, London.

Keynes, J.M. (1936), 'The General Theory of Employment, Interest and Money', Macmillan, London.

Keynes, J.M. (1939), 'Relative Movements of Real Wages and Output', Economic Journal, 49, 34-51.

Killingsworth, M.R. (1983), 'Labour Supply', Cambridge University Press, Cambridge.

Klein, R.E. and M. O'Higgins (1985), 'The Future of Welfare', Basil Blackwell, Oxford.

Knight, F. (1921), 'Risk, Uncertainty and Profit', Little Brown, New York.

Kregel, J.A. (1973), 'The Reconstruction of Political Economy: an Introduction to Post-Keynesian Economics', Macmillan, London.

Kuznets, S. (1946), 'National Income: a Summary of Findings', NBER, New York.

Lancaster, K. (1971), 'Consumer Demand: a New Approach', Columbia University Press, New York.

Lapp, J.S. (1985), 'Mandatory Retirement as a Clause in an Employment Insurance Contract', Economic Inquiry, 23, 69-91.

Lazear, E.P. (1979), 'Why is there Mandatory Retirement?', Journal of Political Economy, 87, 1261-1284.

Le Grand, J. (1982), 'The Strategy of Equality: Redistribution and the Social Services', George Allen and Unwin, London.

Lippmann, S.A. and J.J. McCall (1976), 'The Economics of Job Search: a Survey', Economic Inquiry, 14, 155-189.

Lydall, M.F. (1959), 'The Long-term Trend in the Size Distribution of Income', Journal of the Royal Statistical Society, Series A, 122.

Lydall, H.F. and D.G. Tipping (1961), 'The Distribution of Personal Wealth in Britain', Bulletin of the Oxford Institute of Economics and Statistics, 23.

Lynch, L.M. (1983), 'Job Search and Youth Unemployment', Oxford Economic Papers, 35(sup.), 271-282.

Maki, D.R. and Z.A. Spindler (1975), 'The Effect of Unemployment Compensation on the Rate of Unemployment in Great Britain', Oxford Economic Papers, 27, 440-454.

McGregor, A. (1978), 'Unemployment Duration and Re-employment Probability', Economic Journal, 88, 693-706.

Meade, J.E. (1964), 'Efficiency, Equality and the Ownership of Property', George Allen and Unwin, London.

Meade, J.E. (1973), 'The Inheritance of Inequalities', Proceedings of the British Academy, 59.

Meade, J.E. (1978), 'The Structure and Reform of Direct Taxation', IFS, London.

Mincer, J. (1958), 'Investment in Human Capital and Personal Income Distribution', Journal of Political Economy, 66, 281-302.

Mincer, J. (1974), 'Schooling, Experience and Earnings', NBER, New York.

Mirrlees, J.A. (1971), 'An Exploration in the Theory of Optimum Income Taxation', Review of Economic Studies, 38, 175-208.

Mirrlees, J.A. (1972), 'Population Policy and the Taxation of Family Size', Journal of Public Economics, 1, 169-198.

Modigliani, F. and R. Brumberg (1954), 'Utility Analysis and the Consumption Function: an Interpretation of Cross-section Data', in K. Kurihara (ed.), 'Post Keynesian Economics', George Allen and Unwin, London.

Mowat, C.L. (1955), 'Britain between the Wars 1918-40', Methuen, London.

Muellbauer, J. (1977), 'Testing the Barten Model of Household Composition Effects and the Cost of Children', Economic Journal, 87, 460-487

Muellbauer, J. (1980), 'The Estimation of the Prais-Houthakker Model of Equivalence Scales', Econometrica, 48, 153-176.

Munnell, A.H. (1974), 'The Effect of Social Security on Personal Savings', Ballinger, Cambridge, Massachusetts.

Narendranathan, W., S.J. Nickell, and J. Stern (1985), 'Unemployment Benefits Revisited', Economic Journal, 95, 307-329.

Neary, J.P. and K.W.S. Roberts (1980), 'The Theory of Household Behaviour under Rationing', European Economic Review, 13, 25-42.

Nickell, S.J. (1979a) 'The Effect of Unemployment Benefit and Related Benefits on the Duration of Unemployment', Economic Journal, 89, 34-49.

Nickell, S.J. (1979b), 'Estimating the Probability of Leaving Unemployment', Econometrica, 47, 1249-1266.

Nickell, S.J. (1980), 'A Picture of Male Unemployment in Britain', Economic Journal, 90, 776-794.

Nozick, R. (1974), 'Anarchy, State and Utopia', Basil Blackwell, Oxford.

Oi, W.Y. (1962), 'Labour as a Quasi-fixed Factor', Journal of Political Economy, 20, 538-555.

Paish, F. and A.T. Peacock (1954), 'The Economics of Dependence 1952-82', Economica, 21, 279-299.

Palmore, E.B. (1972), 'Compulsory versus Flexible Retirement: Issues and Factors', The Gerontologist, 12, 343-348.

Parker, S. (1982), 'Work and Retirement', George Allen and Unwin, London.

Peacock, A.T. and J. Wiseman (1967), 'The Growth of Public Expenditure in the United Kingdom', George Allen and Unwin, London.

Pissarides, C.A. (1983), 'Efficiency Aspects of the Financing of Unemployment Insurance and Other Government Expenditure', Review of Economic Studies, 50, 57-69.

Pollak, R.A. and T.J. Wales (1979), 'Welfare Comparisons and Equivalence Scales', American Economic Review, 69, 216-221.

Prais, S.J. and H.S. Houthakker (1955), 'The Analysis of Family Budgets', Cambridge University Press, Cambridge.

Ramsey, F.P. (1927), 'A Contribution to the Theory of Taxation', Economic Journal, 37, 47-61.

Ramsey, F.P. (1928), 'A Mathematical Theory of Saving', Economic Journal, 38, 543-559.

Rawls, J. (1971), 'A Theory of Justice', Harvard University Press, Cambridge, Massachusetts.

Roberts, K.W.S. (1984), 'The Theoretical Limits to Redistribution', Review of Economic Studies, 51, 177-195.

Rothbarth, E. (1940), 'The Measurement of Changes in Real Income under Conditions of Rationing', Review of Economic Studies, 8, 100-107.

Rothschild, M. and J.E. Stiglitz (1976), 'Equilibrium in Competitive Insurance Markets', Quarterly Journal of Economics, 90, 629-650.

Samuelson, P.A. (1947), 'Foundations of Economic Analysis', Harvard University Press, Cambridge, Massachusetts.

Samuelson, P.A. (1958), 'An Exact Consumption-Loan Model with or without the Social Contrivance of Money', Journal of Political Economy, 66, 467-82.

Sampson, A.A. (1978), 'Optimal Redundancy Compensation', Review of Economic Studies, 45, 177-195.

Sandford, C.T. (1983), 'Capital Taxes - Past, Present and Future', Lloyds Bank Review, 150, 34-49.

Sauvy, A. (1969), 'General Theory of Population', Weidenfeld and Nicholson, London.

Sawyer, M.C. (1979), 'The Effects of Unemployment Compensation on the Rate of Unemployment in Great Britain: a Comment', Oxford Economic Papers, 31, 135-146.

Seabrook, J. (1982), 'Unemployment', Quartet Books, London.

Seade, J.K. (1977), 'On the Shape of Optimal Tax Schedules', Journal of Public Economics, 7, 203-236.

Sen, A.K. (1970), 'Collective Choice and Social Welfare', Oliver and Boyd, Edinburgh.

Sen, A.K. (1973), 'On Economic Inequality', Oxford University Press, Oxford.

Shavell, S. (1979), 'Risk Sharing and Incentives in the Principal and Agent Relationship', Bell Journal of Economics, 10, 55-73.

Shavell, S. and L. Weiss (1979), 'The Optimal Payment of Unemployment Insurance Benefits over Time', Journal of Political Economy, 87, 1347-1362.

Sheshiniski, E. (1978), 'A Model of Social Security and Retirement Decisions', Journal of Public Economics, 10, 337-360.

Simon, H.A. (1957), 'Administrative Behaviour', Macmillan, New York.

Sinfield, A. (1981), 'What Unemployment Means', Martin Robertson, Oxford.

Smart, J.J.C. and B. Williams (1973), 'Utilitarianism, For and Against', Cambridge University Press, Cambridge.

Stiglitz, J.E. (1969), 'Distribution of Income and Wealth among Individuals', Econometrica, 37, 382-397.

Tarshis, L. (1939), 'Changes in Real and Money Wage Rates', Economic Journal, 49, 150-154.

Titmuss, R.M. (1955), 'Pension Systems and Population Change', in 'Essays on the 'Welfare State'', George Allen and Unwin, London, 1963.

Tobin, J. (1970), 'On Limiting the Domain of Inequality', Journal of Law and Economics, 13, 263-277.

Tobin, J. and H.S. Houthakker (1950), 'The Effect of Rationing on Demand Elasticities', Review of Economic Studies, 18, 140-153.

Todaro, M.P. (1969), 'A Model of Labor Migration and Urban Unemployment in Less Developed Countries', American Economic Review, 59, 138-148.

Townsend, P. (1981), 'The Structured Dependency of the Elderly: a Creation of Social Policy in the Twentieth Century', Ageing and Society, 1, 5-28.

Walker, A. (1980), 'The Social Creation of Poverty and Dependency in Old Age', Journal of Social Policy, 9, 49-75.

Weintraub, S. (1958), 'An Approach to the Theory of Income Distribution', Chilton, Philadelphia.

Weiss, Y. (1972). 'On the Optimal Pattern of Labour Supply', Economic Journal, 82, 1293-1315.

Welford, A.T. (1958), 'Ageing and Human Skill', Oxford University Press, London.

Williamson, O.E. (1975), 'Markets and Hierarchies', The Free Press, New York.