

LOANS, BEQUESTS AND TAXES WHERE ABILITIES DIFFER:
A THEORETICAL ANALYSIS USING A TWO-ABILITY MODEL

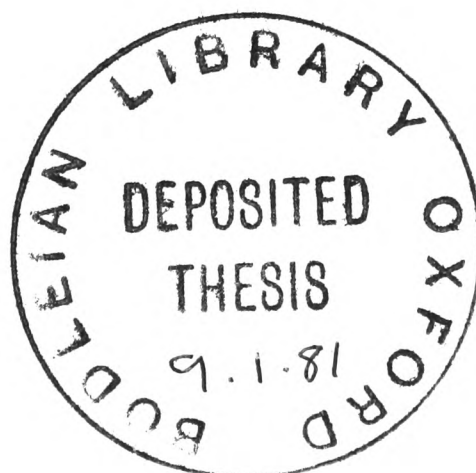
by

H. F. Allen

D.Phil. Thesis

July 1980^{MF}

Nuffield College, Oxford



LOANS, BEQUESTS AND TAXES WHERE ABILITIES DIFFER:
A THEORETICAL ANALYSIS USING A TWO-ABILITY MODEL

H. F. Allen, Nuffield College
D.Phil. Trinity Term, 1980

ABSTRACT

Three problems that are associated with differences in abilities are investigated using a model with two ability groups.

The first concerns the capital market. With differing abilities which cannot be identified, lenders are unable to predict borrowers' outputs and hence whether they will default or not. The penalty for default is assumed to be that borrowers are in future unable to obtain loans. To enforce contracts it is therefore necessary for the current payment to be less than the value of future access to the capital market. It is shown that this implies lenders should make the payment for borrowing a share in the output since this enables them to receive the maximum amount obtainable from each borrower without precipitating default.

The second concerns the distribution of wealth and inheritance when there is an imperfect capital market. The only capital people can use is their own savings and inheritance; with differing abilities the productive capacity of the economy then depends on the distribution of wealth. One of the ability groups is much better than the other at organising production; they found, expand or maintain fortunes whereas the low ability individuals dissipate any inheritance they receive. The various types of equilibrium are categorised and the effect of inheritance taxation is considered. The latter is then compared with a wage tax. Either can plausibly be superior on the basis of the change in Rawlsian social welfare for a given yield.

The final problem concerns the redistribution of income by taxation. It is argued that if differences in abilities lead to people supplying different types of labour which are combined together to produce output, the endogeneity of wages may make a considerable difference to optimal linear tax rates. In this case there is not only the usual redistribution via the fiscal system but also redistribution via the production function. These may be reinforcing or opposing. It is shown that in plausible cases the production effect may outweigh the fiscal so that net redistribution to the poor involves a wage subsidy financed by a lump sum tax.

PREFACE

I have incurred many debts during the preparation of the thesis. My profound thanks are due to James Mirrlees, who supervised this work and whose advice and encouragement have been invaluable, and to Joseph Stiglitz, my M.Phil. supervisor, who stimulated my initial interest in the problems considered here. I have learnt an immense amount from them and regard myself as extremely fortunate to have been their student.

I am also very grateful to Christopher Bliss and John Flemming for all that they have taught me, and for their comments on various parts of the thesis.

Many helpful remarks were received from an anonymous referee and from participants of seminars held at Oxford in May, 1978 and February, 1980, at Princeton in September, 1978 and February, 1980 and at the University of Pennsylvania in February, 1980.

Finally, I would like to thank my wife for typing the thesis and my baby son for allowing her to do so!

TABLE OF CONTENTS

	<u>Page</u>
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 SHARE CONTRACTS AND DEFAULT	9
2.1. A survey of theories of share contracts	9
2.2. A theory of share contracts and default	29
2.2.1. Introduction	29
2.2.2. The model	30
2.2.3. The analysis	34
2.2.4. Some extensions	45
2.2.5. Concluding remarks	53
CHAPTER 3 INHERITANCE, SAVINGS AND TAXATION WITH AN IMPERFECT CAPITAL MARKET	56
3.1. A survey of theories of wealth distribution	56
3.2. A theory of inheritance, savings and taxation with an imperfect capital market	76
3.2.1. Introduction	76
3.2.2. The model	77
3.2.3. Inheritance taxation	95
3.2.4. Wage taxation	103
3.2.5. A Rawlsian comparison	105
3.2.6. Concluding remarks	108
CHAPTER 4 OPTIMAL LINEAR INCOME TAXATION WITH GENERAL EQUILIBRIUM EFFECTS ON WAGES	109
4.1. Introduction and a survey of theories of income taxation	109
4.2. A theory of optimal linear income taxation with general equilibrium effects on wages	121
4.2.1. The model	121
4.2.2. The incidence of a linear tax	125
4.2.3. The normative analysis	130
4.2.4. Some possible cases	133
4.2.5. Concluding remarks	140
CHAPTER 5 CONCLUSIONS	146
REFERENCES	152

CHAPTER 1

INTRODUCTION

Differences in people's achievements in similar situations are widely observed. This variation is usually attributed to two factors. The first is that people differ in their ability. The second is that although circumstances may appear similar they are actually not, and there is some stochastic external influence which produces the diversity of outcomes. The essential distinction between these two is that in the first case people's accomplishments are consistent over time whereas in the second they are unrelated.

This thesis investigates three problems that are associated with differences in achievements, on the assumption that these are due to variations in ability whether acquired or inherent. They are all considered using a model with two ability groups.

The first concerns the capital market. With differing abilities which cannot be observed, lenders do not know how much borrowers will produce and hence whether they will default or not. It could be argued that if criminal penalties for breaking contracts were sufficiently high then this would not be a problem. However, it may be difficult to effectively enforce these; humanitarian considerations may be the overriding factor in determining punishments, and if people are at all uncertain of their ability to pay, such measures may prevent everybody from borrowing. For these types of reason, the case where economic incentives are

required to enforce contracts is of interest. This is the situation that is considered in Chapter 2.

It is assumed that the structure of the two-ability model which is used to illustrate the various possibilities is not known by the transactors, which implies that direct screening is ruled out. Although slightly unreasonable in such a simple case this can be justified on the grounds that it is plausible if there are a large number of types and the two group assumption is just for clarity of exposition.

Output depends only on the amount of capital used and the ability of the person in control of production; the quantity produced is observable. Capital is fixed and immobile so that it is only the rental payment which can be defaulted on.

Apart from the difference in ability everybody is the same. They are all taken to be risk neutral, and to have the same time horizon and rate of discount. Their utility is determined by the total discounted value of their consumption.

The main result of the chapter concerns the situation where for people without collateral, the criterion for being able to borrow is that they have not previously defaulted. This can be justified on the grounds that if somebody has already refused, or been unable to pay the rental of capital they have borrowed, then it is reasonable for lenders to suppose that this will happen again. It is then possible to enforce contracts by ensuring that the borrowers' current payment is less than the value of future access to the capital market.

Although ability cannot be observed by other people, individuals have some idea of their own production possibilities. However, lenders cannot acquire this information because of the conflict of interests that arises with default. It may only be marginal for borrowers whether to default or not since in both cases they receive something: if they make the current payment they can borrow and earn profits in future periods; if they do not pay they keep the rental but cannot subsequently earn profits. In the same situation it will make a greater difference for lenders since in the first case they receive a payment and in the second they do not. For this reason lenders specify the contract.

Given this it is argued that the restriction on current payments implies that in order to maximise their return lenders should make the payment for borrowing the capital a share in the output produced since this enables them to receive the maximum amount obtainable from each borrower without precipitating default.

This contrasts with the usual theories of share contracts which are mainly concerned with incentives, risk sharing and screening. These are surveyed in Section 1 of Chapter 2. Section 2 is concerned with the theory of share contracts outlined above. Subsection 2.1. is an introduction, 2.2. describes the model, 2.3. gives the basic analysis and 2.4. some extensions to this. 2.5. contains a few concluding remarks.

In Chapter 2 the distribution of wealth is taken as given. The analysis stresses the problems involved in borrowing

capital to produce with, and this suggests that with differing abilities the distribution of wealth may be an important determinant of the output of an economy and of the other variables of interest. This is the question considered in Chapter 3.

In recent years there has been a certain amount of discussion of the determination of wealth distributions, and in particular of the role of inheritance and life cycle savings. Empirical work tends to have shown that wealth is very unequally distributed and many authors have argued that this is undesirable and inheritance and other forms of taxation should be used to reduce it. This literature is surveyed in Section 1 of Chapter 3.

The focus of most of the theoretical work has been on the consumer and there has been little discussion of the relationship between the ownership of wealth and the production possibilities of the economy. With perfect capital markets this is understandable since producers can borrow the capital they require. However, this is a very strong assumption. Capital markets are notoriously imperfect; the analysis of Chapter 2 provides just one of many reasons why this should be so.

In Section 2 of Chapter 3 the extreme case where the only capital that people can use is their own savings and inheritance is considered in a model which again has two ability types. One group is taken to be much better at organising production than the other. The high ability people either found fortunes, expand their inheritances or maintain them. The low ability individuals born with an

inheritance dissipate it; those born without one supply the manual labour needed for production. The transmission of ability between generations is taken to be Markovian in form. The model thus describes an economy of capitalists and workers in which firms are associated with families. The history of firms may involve a number of phases of expansion and contraction depending on the ability of the generation in control at any particular moment.

After an introduction in Subsection 2.1. the basic model is described in 2.2. and the various types of equilibrium are categorised. In order to analyse the effects of taxation a number of further simplifying assumptions are made and these are described at the end of the subsection.

In 2.3. the incidence of an inheritance tax is considered. The crucial variable is the capital labour ratio. Having found the change in this it is then a relatively simple matter to determine the changes in the other variables of interest. The effect of the tax is complex and the various direct and indirect general equilibrium effects are identified. The formulae are such that no definite results can be obtained; the possible signs and magnitudes of the relevant elasticities imply that changes can plausibly be in either direction.

As an example of the analysis of another tax in the model, Subsection 2.4. considers a tax on wage income and in 2.5. this is compared with an inheritance tax. It is shown that either can plausibly be superior on the basis of the change in Rawlsian social welfare for an equal yield of revenue. Finally Subsection 2.6. contains some concluding

remarks.

Differences in ability lead not only to inequality of wealth but also to inequality of income. Chapter 4 considers the redistribution of income by taxation. This raises a number of different incentive issues depending on the source of the earnings which are taxed. The focus of the chapter is on the taxation of income from labour and the distortion of the choice between labour and leisure.

In Section 1 previous theories are surveyed. The usual assumption concerning production is that there is one type of labour; differences in earnings arise from differing abilities which determine the amount of labour which can be supplied within a given time period. The results tend to confirm the intuition that with utilitarian, Rawlsian and similar social welfare functions, taxes should redistribute from rich to poor and the higher the weighting of low utilities the greater this should be.

In Section 2 redistribution using a linear income tax consisting of a proportionate component and a lump sum grant is considered analytically, in a model with different assumptions concerning production. This is outlined in Subsection 2.1. Instead of there being one type of labour with abilities referring to the amount of this that can be supplied within a given time period, there are taken to be two types of labour which are combined together in a constant returns to scale production function to produce the output. These are referred to as entrepreneurial and raw labour. There are two ability groups, one of which is sufficiently better at performing entrepreneurial tasks to ensure that they

supply all of this type of labour; the other group supplies the raw labour. Wages for each type are equated to their marginal products and so are determined by relative labour supplies.

This endogeneity of wages significantly changes the problem of redistribution. Now a tax not only alters net incomes through the fiscal system but also changes relative labour supplies and hence wages so that there is redistribution through the production function as well. This is shown in Subsection 2.2. where expressions are derived for the incidence of the tax. The special case where a linear tax is introduced at the market solution is also considered.

In 2.3. these results are used to consider the tax in a normative context. It is shown that provided the marginal utility of income is higher for the better paid the condition for the introduction of a small tax to improve social welfare is the same in both Rawlsian and utilitarian cases. The first order conditions a linear tax must satisfy in order to maximise social welfare are algebraically complex; since the analysis is again similar for both social welfare functions, only the Rawlsian case is given.

Subsection 2.4. contains a number of plausible special cases. These show it is quite possible for the introduction of a small wage tax to finance a lump sum grant so that there is redistribution through the fiscal system from rich to poor, to result in the latter being worse off. This occurs because the redistribution through the production function is offsetting and dominates. In such cases Rawlsian and utilitarian social welfare are reduced by the tax. If the government is

interested in maximising Rawlsian social welfare it is shown that a negative linear tax consisting of a wage subsidy financed by a uniform lump sum tax can be optimal. This possibility again arises because of the inclusion of a production function; similar examples can be given in the utilitarian case. Subsection 2.5. consists of concluding remarks.

Finally, Chapter 5 summarises the main findings of the thesis.

CHAPTER 2

SHARE CONTRACTS AND DEFAULT

2.1. A survey of theories of share contracts

There are three main types of component in contracts concerning payment between owners of capital or land and those who use it: wages, rent and a share of the output. In neoclassical theory wages and rent are the prices of labour and capital services, which can be thought of in the same way as any other commodities. In such a framework the purpose of specifying payments which depend on the level of output is not immediately obvious.

However, contracts with share components are widely observed. For example, in agriculture sharecropping is a common form of land tenure in many countries. In industry joint stock companies are one of the most usual types of organisation and the incomes of many managers, salesmen and other types of employee are partly derived from a share in output or profits. In service industries as diverse as retailing and the theatre the payment for the premises on which the business is conducted is often at least partly in the form of a share of the profits.

The properties of share contracts have been fairly extensively considered in an attempt to understand this widespread occurrence. This analysis has usually been in an agricultural context. Among the theories put forward three main aspects can be discerned: incentives, risk sharing and screening.

The incentive properties of share contracts have been a

subject of dispute for many years. The arguments that have been put forward can be illustrated here with the following simple model.

Each person has a production function expressing output as a function of land K and labour L .

$$(1) \quad Y = Y(K, L)$$

$$(2) \quad Y_K, Y_L > 0; \quad Y_{KK}, Y_{LL} < 0$$

where a subscript denotes a partial derivative. There are assumed to be constant or decreasing returns to scale.

Utility depends on consumption C and the amount of labour supplied.

$$(3) \quad U = U(C, L)$$

$$(4) \quad U_C > 0; \quad U_L, U_{CC}, U_{LL} < 0$$

Among classical writers such as Smith (1776) it was generally believed that the system of renting was the most desirable form of land tenure (see Cheung (1969)(Ch. 3A) for an account of classical views). The basis for this belief was that such a system provided the correct incentives and is therefore efficient. The argument can be illustrated in the simple model given above where labour is the only nonland input.

For the moment it is assumed that all contracts can be costlessly enforced.

There are competitive markets for the services of K

and for L, and the prices of these are denoted by r and w respectively. In this model each person is not only a consumer but also a producer. In the latter role they all consider setting up a firm or in this context a farm. K and L are chosen to maximise profits.

$$(5) \quad \text{Max}_{K, L} Y(K, L) - w L - r K$$

This gives the usual first order conditions which equate marginal products with prices.

$$(6) \quad Y_K = r$$

$$(7) \quad Y_L = w$$

As consumers, owners of land supply capital services and each person chooses his labour supply and consumption to maximise his utility subject to his budget constraint.

$$(8) \quad \text{Max}_{C, L} U(C, L)$$

subject to

$$(9) \quad C = w L + r K^* + \pi$$

where K^* is the amount of land owned and π is the pure profit which results from being a producer. The first order condition in this case equates the wage with the ratio of the marginal disutility of labour to the marginal utility

of consumption

$$(10) \quad w = - \frac{U_L}{U_C}$$

It is important to make clear the relationship between each person's choice of labour input as a producer and labour supply as a consumer. It is implicit here that the quantity demanded for production need not equal the quantity supplied as a consumer; any difference between these amounts is hired or sold on the labour market.

The model above is simply a special case of the standard neoclassical model, and its equilibrium has the usual property of being Pareto efficient (see, for example, Debreu (1959)).

The illustration above is in terms of the input labour but can clearly be extended to other nonland inputs.

Smith (1776) likened the share contract to a tax and argued it would have a similar effect in reducing the incentive to invest and improve the land. Marshall (1920) gave a more formal analysis of the tax-equivalent argument which can be applied to any nonland input. It can be illustrated in the simple model above in the following way. Instead of paying a rent r to the landlord the tenant pays a share σ of his output. (5) is therefore changed to

$$(11) \quad \text{Max}_{K, L} \quad Y(K, L) - w L - \sigma Y(K, L)$$

This gives the first order conditions

$$(12) \quad (1 - \sigma) Y_K = 0$$

$$(13) \quad (1 - \sigma) Y_L = w$$

It is usually assumed that $Y_K > 0$ for all K . If this is the case then it follows from (12) that unless there are other constraints (see Bliss and Stern (1980)), there must be excess demand and some sort of rationing of land. The inefficiency argument results from (13) which implies the marginal product of labour is less than the wage.

It follows from (12) that if the tenant chooses K he will demand an infinite amount and so the landlord must therefore specify the quantity of land. Marshall (1920) realised that if the landlord also specified other parts of the contract then the solution would be equivalent to that given by (6) and (7) rather than (12) and (13). This argument was later extended by Johnson (1950) and Cheung (1969).

Whereas with renting the landlord simply supplied the land and the tenant made the decisions, the process is now reversed. The landlord makes the decisions and the tenant has to supply the specified amount of labour. If this exceeds the amount he himself wishes to supply then he can hire the surplus on the labour market; if it is less he can sell his excess there.

In the model used here, there are a number of ways that the landlord can specify the contract to ensure that resource allocation is equivalent to (6) and (7) and therefore efficient. Johnson (1950) and Cheung (1969) pointed out that one way is for him to specify the share σ as well as the

inputs. To determine the conditions of the contract the landlord chooses σ , K and L to maximise his average return subject to the constraint that the tenant must earn at least as much as in his alternative employment. It is assumed here that there are also wage and rental markets. The tenant must therefore cover not only the opportunity cost of the labour the landlord specifies but also the pure profits π he would earn as a producer if he rented. The landlord's problem is therefore

$$(14) \quad \text{Max}_{\sigma, K, L} \quad \frac{\sigma Y(K, L)}{K}$$

subject to

$$(15) \quad (1 - \sigma) Y(K, L) \geq w L + \pi$$

Taking (15) as an equality and using it to substitute for $\sigma Y(K, L)$ simplifies (14) to

$$(16) \quad \text{Max}_{K, L} \quad \frac{Y(K, L) - w L - \pi}{K}$$

This gives the first order conditions

$$(17) \quad Y_K = \frac{\sigma Y}{K}$$

$$(18) \quad Y_L = w$$

where (17) is derived by substituting from (15) again.

Equation (17) implies that the marginal product of

capital is equal to the average return to the landlord which is equivalent to (6); (18) is the same as (7). It follows that in this form share contracts lead to an efficient allocation of resources and the distribution of income is the same as with rent and wage contracts.

Marshall (1920) took the share as given but realised that by means of a side payment the same result can be achieved. The formal derivation is due to Bliss and Stern (1980). Denoting the side payment from landlord to tenant by ω , the former determines the contract in the following way

$$(19) \quad \text{Max}_{\omega, K, L} \quad \frac{\sigma Y(K, L) - \omega}{K}$$

subject to

$$(20) \quad (1 - \sigma) Y(K, L) + \omega \geq w L + \pi$$

Taking (20) as an equality and substituting for $\sigma Y(K, L) - \omega$ it can be seen that (19) simplifies to the same form as (16) and so is equivalent.

Instead of the landlord specifying the nonland inputs it was pointed out by Heady (1947) that it is equivalent for the landlord to pay a share in the cost of inputs equal to his share in the output and allow the tenant to choose the levels. In the example, if the landlord pays his share of the wage costs and the tenant chooses the labour supply to maximise his return the $(1 - \sigma)$ term cancels and the tenant equates the marginal product of labour with the wage. Formally, the tenant's problem is to

$$(21) \quad \underset{L}{\text{Max}} \quad (1 - \sigma)(Y(K, L) - w L)$$

which gives (18) as the first order condition as required by the landlord. Thus if this method is adopted the landlord need not specify the input level although he still has to verify that the tenant uses the claimed amount of labour. The case where it is unobservable effort, rather than the observable number of hours worked which is relevant for the production function is discussed below. However, in the context of this method of achieving the optimal level of inputs it is often argued that because of the unobservability of effort it is only practical for inputs other than labour and land. It can be seen that with multiple inputs a combination of cost sharing for some inputs and the specification of others by the landlord is also optimal.

Summarising the incentive properties of share contracts: if tenants choose inputs independently then share contracts will lead to inefficiency. However, if enforcement costs are ignored, landlords choose the share and inputs subject to the constraint that tenants must earn at least as much as in their alternative employment then the allocation of resources and distribution of income is the same as under competitive rent and wage contracts and so is Pareto efficient.

The second aspect of share contracts which has received attention concerns the case where output is uncertain. With rent contracts the tenant bears all the risk and with wage contracts the landlord does. However, with share contracts the risk is spread between the two and it has been

argued that this is one of their major advantages (see, for example, Cheung (1969)).

Stiglitz (1974) provided a rigorous analysis of the risk spreading properties of share contracts. He considered linear schedules relating remuneration to output, consisting of a share and a fixed component which is either a wage or a rent depending on the direction of payment, in a model with multiplicative risk, no enforcement costs and constant returns to scale. Among other things, he showed that with a pure share contract both factors are paid their mean marginal product. If the fixed component is a rent then the worker bears proportionately more risk and receives more than his mean marginal product for doing so. Similarly if it is a wage the landlord receives more than the marginal product of the land. There will be a pure rental system if and only if workers are risk neutral and a pure wage system if and only if landlords are risk neutral. With the further assumption of representative landlords and workers it can be shown that the fixed component is a rent or a wage depending on whether the worker is less or more risk averse than the landlord. For crops with a greater variance which are therefore more risky the share component paid to the landlord increases provided he is less risk averse than the worker and vice versa.

However, the result which has perhaps received the most attention was that a mixture of wage and rent contracts could achieve the same apportioning of risk as a pure share contract. Newbery (1977) generalised this to the case of non-multiplicative risk. The demonstration of this result given below is based on that in Newbery and Stiglitz (1978).

Output is now dependent on K , L and a stochastic variable θ , which can be thought of as something like the weather. The tenant's production function becomes

$$(22) \quad Y = Y(K, L, \theta)$$

Constant returns to scale in K and L are again assumed.

Consider a share contract in which the landlord provides land K and receives σ of the output and the tenant provides labour L and receives $(1 - \sigma)$ of the output. Newbery and Stiglitz (1978) show that if the land is divided into two parts one of which is rented by the tenant at a fixed rate r and the other of which is cultivated by the landlord hiring the worker at fixed wage w , then there exist values of r and w such that the tenant and landlord will be indifferent to the original share contract.

If the tenant supplies σL of his labour on a wage basis, rents $(1 - \sigma) K$ of the landlord's land and the wage/rent ratio is set so that

$$(23) \quad \frac{w}{r} = \frac{(1 - \sigma) K}{\sigma L}$$

then by this and the constant returns to scale of (22) he receives

$$(24) \quad \{Y((1 - \sigma) K, (1 - \sigma) L, \theta) - (1 - \sigma) r K\} + \sigma w L \\ = (1 - \sigma) Y(K, L, \theta)$$

Similarly the landlord receives

$$\begin{aligned}
 (25) \quad & \{Y(\sigma K, \sigma L, \theta) - \sigma w L\} + (1 - \sigma) r K \\
 & = \sigma Y(K, L, \theta)
 \end{aligned}$$

Hence the share contract is equivalent to a mixed wage/rent contract.

Furthermore, it is argued that (23) must be satisfied for an equilibrium to be attained in the wage, rent and share markets since if for example the left hand side were greater than the right then landlords would prefer a mixed wage/rent agreement and tenants would prefer the share contract. Similarly if the inequality is reversed.

The result can be generalised even further. All that is really required is that there be no enforcement and transaction costs. This can be illustrated in the simple model above but without constant returns to scale and initially assuming the average output of all subplots is the same.

An equivalent mixed wage/rent contract can be obtained by reinterpreting a share agreement. Assuming the techniques of production used remain the same then given the uniformity of output it is clearly equivalent to give a share σ of the final output to the landlord and $(1 - \sigma)$ to the worker or assign the total output from σK of the land to the landlord and that from the other $(1 - \sigma) K$ to the tenant. Since the landlord owns all the land and the tenant provides all the labour, the latter assignment implicitly involves an exchange of σL of labour for the use of $(1 - \sigma) K$ of land. There is an implicit relative price w/r in this transaction which is the same as in (23). Given the no enforcement and transaction costs assumption, the form of the exchange makes

no difference, provided production techniques again remain the same; it is equivalent if the landlord hires $(1 - \sigma) L$ at w in the labour market and the tenant hires σK at rent r in the land services market. As before equilibrium in the wage, rent and share markets requires that (23) is satisfied with an equality. The landlord and tenant are therefore indifferent to the original share contract and the mixture of wage and rent contracts.

The difference between the argument here and that of Newbery and Stiglitz (1978) is that they assume the two lots of inputs are applied in two production functions which makes the constant returns to scale assumption important. The above shows this is not necessary provided the same techniques can be applied as in a share contract; one subplot is simply called the tenants private plot and the other the landlord's. However, it should be noted that with wage and rent agreements, the contracts may need to specify the techniques to be used. For example, if production requires the performance of a number of different tasks there may be increasing returns to scale because of the time spent switching from one to another. In such a case the equivalent wage and rent contracts will need to be formulated in such a way that the tenant will be able to cultivate both subplots as though they were one, so that having completed a particular task on the landlord's part of the land he can continue the same task on his own private plot.

It can also be seen that the assumption of a uniform output and the adoption of the simple model are not critical; the division of the inputs is just more complicated in other cases. For example, if for a given K and L the effect of

the weather was such that the best method of production used two techniques so that one part of the land was cultivated in a different way from the other, output would not be uniform. The appropriate equivalent contract would involve dividing both parts using the techniques into lots. A similar method can be used in more complex situations.

Thus pure share contracts may occur in a number of forms which in the absence of enforcement and transaction costs are essentially the same. Landlord and tenant may share the output after it has been produced; inputs may be split into two lots, the outputs of which are assigned to the landlord and tenant respectively so that there is a direct exchange of the use of land for labour, or finally tenants and landlords may use a mixture of explicit wage and rent contracts. The same conclusion applies to contracts with a share and a fixed component; however, whereas with a pure share agreement the tenant's rent payment for the land equals the landlord's wage payment, in this case the fixed component represents the excess of one payment over the other.

As in the case of incentives rent and wage systems have the same risk sharing properties as share contracts provided there are no transaction and enforcement costs. The comparison of rent and wage with share contracts has therefore tended to focus on the combination of incentives and risk sharing properties possessed when these costs are not absent. Rent contracts provide the correct incentives but force the tenant to bear all the risk. With wages the worker doesn't face any risk but has no incentives and so may have to be supervised; if it is effort that is relevant for production then this may be very difficult and costly. A share contract

represents a compromise in which risk is spread and the worker has some incentives.

Stiglitz (1974) considered formally the case where effort cannot be observed and the payment schedule is linear with a share and fixed component. He showed that if workers are risk neutral a rent contract is optimal; if they are risk averse there is a share component which is larger the greater is the responsiveness of effort to an increase in the share and which is smaller the higher is the risk aversion of the worker. If the elasticity of substitution of the production function is one, or workers are risk neutral, they receive their mean marginal product. If they are risk averse they receive more or less than the latter as the elasticity is greater or less than unity.

One of the things that theories of share contracts have tried to explain is the coexistence of wage, rent and share arrangements. The analysis of incentives and risk sharing have shown the systems give the same opportunities provided there are no transaction costs. Bliss and Stern (1980) have argued that equivalence provides a justification for coexistence. The converse argument is that transaction costs will differ. Cheung (1969) for example suggests that the sum of negotiation and enforcement costs is higher for share than either wage or rent contracts. As Newbery and Stiglitz (1978) point out given both the incentive and risk sharing equivalence results, this suggests only rent and wage contracts will be observed.

The third view of share contracts argues that the coexistence of the three types generates information on abilities. Tenants are assumed to know their own abilities but landlords

cannot distinguish between them. The use of several contracts allows tenants to identify themselves. Those with high ability will choose rent contracts so that they gain all the return to their ability; those with low ability choose the wage contract since in this case payment does not depend on their ability and those in between will choose the share contract. The coexistence of contracts thus generates valuable information which allows resources to be allocated more efficiently.

Two versions of the argument have been put forward independently. Hallagan (1978) suggested the case in which endowments of entrepreneurial ability differ. Newbery and Stiglitz (1978) used a model in which workers differed in the amount of labour they supplied within a given time period. However, there are problems with both these approaches.

Hallagan (1978) does not specify his model rigorously but it appears that he assumes that landlords always specify the contract without justifying why this should be the case. However, the asymmetry of information and hence the problem solved by the coexistence of the contracts arises because of this feature. To see this consider a slight variant of the simple neoclassical model in which the efficiency arguments were considered. Each person chose inputs of land and labour to maximise his profits taking the wage as given. In this case everybody has a production function of the form

$$(26) \quad Y = Y(A, K, L, \theta)$$

where A is ability, K and L are observable inputs and θ is a

stochastic variable. Provided each person knows his own ability and is risk neutral the efficiency argument is valid. It does not matter that other people's ability cannot be identified because each person specifies his own demands for inputs. Whereas Hallagan assumes only the suppliers of land organise production, here everybody has the opportunity to become a producer and do this.

In the model of Newbery and Stiglitz (1978) the information problem arises because labour cannot be quantified directly and buyers are unsure of the amount they are purchasing. With a wage system based on the number of hours worked this may lead to inefficiency, but their model is such that with risk neutrality it will not. They assume everybody has an identical constant returns to scale production function of the form

$$(27) \quad Y = \theta A L y(k)$$

where $k = K / A L$, L is the number of hours and $A L$ the number of labour units supplied. Efficiency requires that land labour ratios, k , are equated. Each person knows his own ability. If everybody chooses the amount of land to demand taking the rent r and their endowment of labour as given then this is sufficient to equate land labour ratios so that whether a labour market exists or not, or is imperfect is unimportant.

Also in both models there seems no reason why share rather than a mixture of wage and rent contracts should not be used to screen.

Perhaps the screening potential of the contracts can

best be illustrated by considering a simple combination of the models in which there are two dimensions to ability. People are taken to differ both in their ability to combine factors of production and in the amount of labour units that they can supply in a given time period and there is no systematic relationship between the two. In this case there is a real problem if labour productivity cannot be observed, and screening in both dimensions cannot be achieved by using mixtures of wage and rent contracts. It is explained below how it can be achieved by also using share contracts.

If entrepreneurs could identify labour productivity directly and chose inputs in the usual way to maximise profits taking prices as given, then production would be efficient. As above, even though entrepreneurial abilities differ and cannot be observed, provided contracts are enforceable there will be no inefficiency because entrepreneurs adjust their demands to suit their ability. The rental contract thus ensures an allocation such that marginal products of capital are equated between firms. If labour productivity is observable then a wage contract ensures the marginal products of labour are also equated. If it is not observable a combination of share and wage contracts may still enable this to be achieved.

To demonstrate this consider the simple case where there are two levels of productivity. Workers of type 1, who are the more able supply A_1 units of labour per time period and workers of type 2 supply A_2 . Under the supervision of the entrepreneur each worker is assumed to produce an output which depends on the amount of labour he supplies and the land he is given to work with and is

given by a constant returns to scale production function as in (27). The agents are taken to be risk neutral and $E\theta = 1$. It is therefore possible to use the average production function. The production uncertainty is important however in that it prevents output being used as a reliable indicator of inputs which would enable productivity to be directly observed.

When workers are paid on a wage system they are assumed to choose the number of hours they work to maximise their utility taking the wage as given. This gives a labour supply function

$$(28) \quad L = L(w)$$

If the two types of worker could be identified then they would be paid a wage w for each effective labour unit supplied. A_i 's would therefore receive a wage per unit time period given by

$$(29) \quad w_i = A_i w \quad i = 1, 2$$

Entrepreneurs would choose the total effective labour supply of the firm as described before. However, they must also allocate land efficiently to each worker within the firm. This combined with constant returns to scale implies that the land labour ratio of each worker must be equated.

When productivities are known, A_i workers thus receive in equilibrium a wage w_i , they work for $L(w_i)$ hours and are allocated K_i of land which is such that land labour ratios are equated. It is clearly equivalent if entrepreneurs specify

share contracts $(\sigma_i, L(w_i), K_i)$ such that

$$(30) \quad (1 - \sigma_i) A_i L(w_i) y(k) = w_i L(w_i) \quad i = 1, 2$$

Substituting from (29) into this it can be seen that

$$(31) \quad \sigma_1 = \sigma_2 = \sigma$$

Provided enforcement costs are neglected both workers and entrepreneurs will be indifferent between the share and wage contracts for their ability group.

Now consider what happens if it is assumed that workers cannot be distinguished by entrepreneurs but that they know their own ability. If entrepreneurs offer a share contract $(\sigma, L(w_1), K_1)$ and a wage contract w_2 then provided $L(w_1) \neq L(w_2)$, A_1 workers will choose the former and A_2 's the latter. For A_1 's this occurs because using (30)

$$(32) \quad \begin{aligned} U((1 - \sigma) A_1 L(w_1) y(k), L(w_1)) &= U(w_1 L(w_1), L(w_1)) \\ &> U(w_2 L(w_2), L(w_2)) \end{aligned}$$

with the inequality following from the fact that consumers prefer a higher wage and $w_1 > w_2$. Similarly for A_2 's

$$(33) \quad \begin{aligned} U((1 - \sigma) A_2 L(w_1) y(k), L(w_1)) &= U(w_2 L(w_1), L(w_1)) \\ &< U(w_2 L(w_2), L(w_2)) \end{aligned}$$

provided $L(w_1) \neq L(w_2)$ since otherwise $L(w_1)$ would have been

chosen at wage w_2 . If $L(w_2) = L(w_1)$, then A_2 workers are indifferent between the contracts and it will be necessary to alter the wage contract to $(w_2 - \epsilon)$ to ensure a positive incentive to choose it.

Thus a judicious use of share and wage contracts by entrepreneurs may allow the first best solution to be achieved even though they cannot directly observe productivity. This example is clearly a special case but it does illustrate the principle. In general it may not be possible to achieve the first best solution and there may be problems of existence similar to those given by Rothschild and Stiglitz (1976) in the context of insurance markets. However, the assumption of two groups is not important; with more than two it will just be necessary to use contracts with both wage and share components.

The survey above has concentrated on three of the most important theoretical views of share contracts. It is not exhaustive though; a number of other references should perhaps be briefly mentioned. Newbery (1977) shows that with risky labour markets share contracts have advantages over fixed rent and risky wage contracts. Bell and Zusman (1976) discuss the share contracts that are likely to emerge in a game-theoretic equilibrium. Since the concentration of this chapter is theoretical, empirical work has been hardly mentioned. The interested reader could perhaps consult as a starting point Cheung (1969), Rao (1971), Bardhan and Srinivasan (1971) and Bell (1977).

The theory of share contracts presented in the next section is rather different from those described above. It concerns the relationship between the incentives to default

and the form of the contract when there are different types of entrepreneur who cannot be distinguished. It is shown that share agreements are the profit maximising contracts which simultaneously provide incentives for different groups not to default.

2.2. A theory of share contracts and default

2.2.1. Introduction

One of the assumptions which is usually made and indeed which was implicit in the theories considered in Section 1 was that payments that arose from contracts were always made. It could be argued that provided criminal penalties for default are sufficiently high then this will be the case. However, a number of objections can be made. It may not be possible to enforce criminal penalties effectively enough for them to act as a deterrent; for humanitarian reasons a society may reject the use of severe penalties; in an uncertain world they may discourage economic activity and so on. For reasons such as these the assumption that there is no problem in ensuring payment would appear to be a strong one. In this section the effect of relaxing it in the context of the capital market is considered.

Collateral is one way in which settlement of debts can be ensured in capital markets where there are no legal means of enforcing payment. However, people's wealth may be insufficient for them to cover the payments on the amount of capital they would like to borrow. In this investigation the effect of adopting the convention that anybody who has defaulted or broken a contract is unable to borrow is considered. This implies that for a contract to be enforceable

when collateral is insufficient, current payments must be less than the future discounted value to the borrower of access to the capital market. The main result of this section is to show that such a restriction implies a share contract is the profit maximising agreement which prevents default.

In Subsection 2.2., the basic model used to illustrate the theory is described. In 2.3. the main results are derived and in 2.4. some extensions are considered. Subsection 2.5. contains some concluding remarks.

2.2.2. The model

For ease of exposition there are taken to be two types of people who are denoted A_1 and A_2 and who initially all appear alike. It is also assumed that none of the transactors knows the structure of the model. With only two groups this is rather unrealistic. However, it can be justified on the grounds that it is plausible for more complex models and the two group simplification is just for clarity.

There is one factor of production called capital and denoted K . The model is static in the sense that there is no saving to augment a fixed capital stock and the distribution of the latter is given. The capital is taken to be infinitely durable and immobile so that when rented it cannot be removed or damaged. It can be thought of as something like land or a factory which has already been built. The consequence of this assumption is that it is only possible to default on the rent.

There is one type of homogeneous output which is used for consumption. Output depends on the amount of the factor used and the ability of the person in control of production.

$$(1) \quad Y = Y(A_i, K) \quad i = 1, 2$$

Time is divided into discrete production periods with inputs being supplied at the beginning and outputs at the end. The production functions are assumed to be strictly increasing and concave in K . If $\partial Y(A_i, K)/\partial K$ is denoted $r(A_i, K)$ then

$$(2) \quad r(A_i, K) > 0 ; \frac{\partial r(A_i, K)}{\partial K} < 0 \quad \text{for all } K; \\ i = 1, 2$$

Apart from (2) no restrictions are placed on the production functions, they may or may not cross. Two examples of the types envisaged are shown in Figure 1.

Since production depends on the ability or type of the person in charge, it cannot be supervised and takes place independently from the lender so that he has no control over it and cannot observe what is being done. If it were to be supervised it would be the ability of the lender which was relevant and would be equivalent to using the capital himself. It is because of this independence of operation when lending, that the possibility of default arises. Although production takes place independently it is nevertheless assumed that the total amount of capital used and the amount of output are observable.

People are assumed to know their own production functions, but not with certainty and it is generally realised that this is the case. The actual production functions are deterministic in the sense that a person will produce the same amount in every period provided the quantity of capital used remains

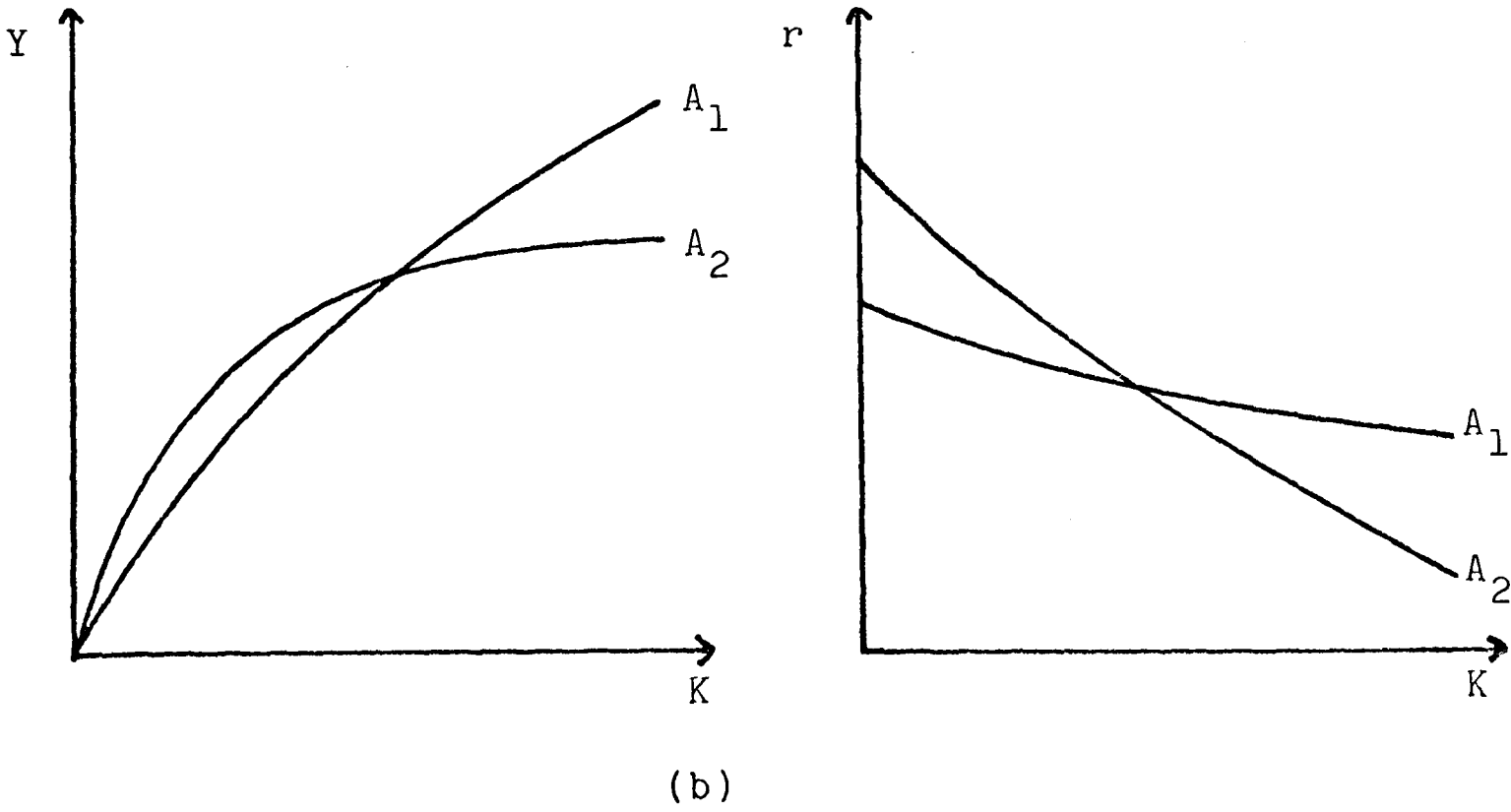
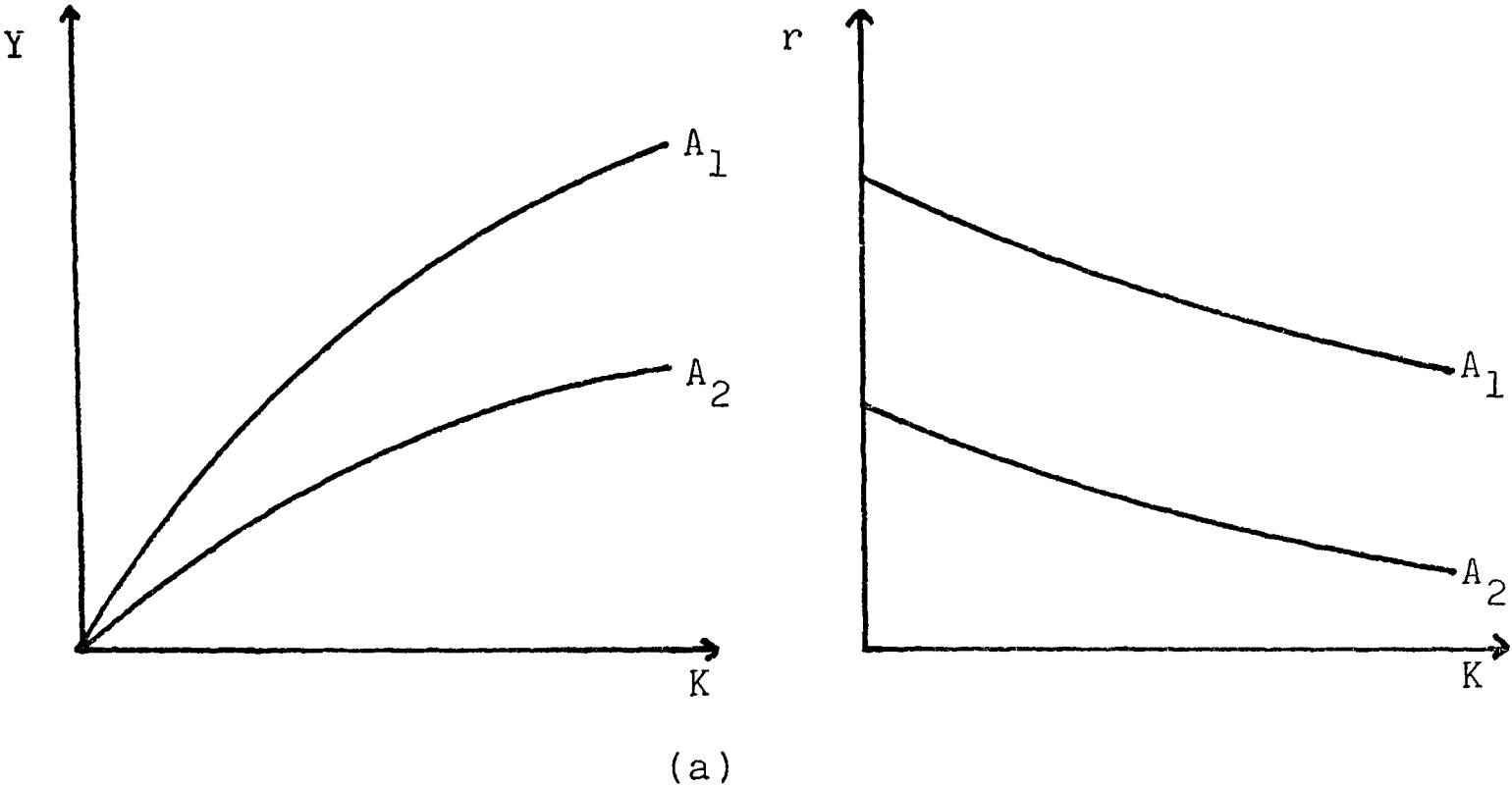


FIGURE 1

constant. For simplicity the expected value of output for given inputs is assumed to coincide with the actual production function. This ensures that when renting with collateral the amount borrowers demand turns out to be correct and avoids the consideration of subsequent adjustments in their borrowing. Once people have produced at a particular point on the production function then this point is taken to be known with certainty.

If owners of capital wish to borrow they can deposit their deeds of ownership with the lender who is thereby assured of receiving payment. This also prevents borrowers using their collateral more than once.

For borrowers who default or break contracts the penalty is that they are in future unable to obtain loans. The justification for this is that if somebody has already defaulted it is reasonable for lenders to suppose he will do so again and refuse to lend to him. It follows from this that for a contract to be enforceable the payments due in the current period must be less than or equal to the value of access to the capital market in future periods.

There is an infinitesimal transaction cost to default so that when the gains from defaulting are equal to the losses, transactors do not default.

There is one generation, the members of which have infinitely long lives, and a time horizon of H periods. Their utility depends only on their discounted consumption. There is no disutility from controlling wealth. Everybody is also taken to be risk neutral so that

$$(3) \quad U_j = I_j \quad \text{for all } j$$

where I_j is the net discounted income of the j^{th} person. The

rate of discount is denoted d and is taken to be strictly positive. Hence the market value V in the current period, denoted period 1, of a constant stream of returns r from time t to period H is given by

$$(4) \quad V = r D(t)$$

where

$$(5) \quad D(t) = \sum_{m=t-1}^H 1 / (1 + d)^m$$

2.2.3. The analysis

Owners of wealth face a number of possibilities. They can either use their own capital and borrow themselves, lend it to those with collateral, or lend it to those without, or a combination of these.

Consider an equilibrium with a rental rate of r' to those who can guarantee payment. The amount that the wealthy can borrow depends on the amount of collateral they have. Given the market rate r' , the value V' of one unit of capital used as collateral is

$$(6) \quad V' = r' D(2)$$

since if the borrower defaults the lender is only able to start obtaining a return to the capital in the period after the current one. It thus follows that for every one unit owned a further $D(2)$ units can be rented no matter what the rate r' is. Hence given ownership of capital K_w , it is possible to control

$$(7) \quad K_W + K_W D(2) = K_W + K_B$$

where K_B is the maximum amount that can be borrowed.

Owners choose the amount of capital to demand and supply to maximise their expected income. Since they are risk neutral and their expected production function coincides with their actual one, given r' they demand K_D to use themselves such that

$$(8) \quad r' = r(A_i, K_D) \quad \text{for } i = 1, 2$$

if there exists K_D satisfying this which is also such that

$$(9) \quad 0 < K_D < K_W + K_B$$

For $r' \geq r(A_i, 0)$ then $K_D = 0$ and for $r' \leq r(A_i, K_W + K_B)$ then $K_D = K_W + K_B$. Their net demand is $K_D - K_W$.

In order for those without collateral to borrow, current payments for capital must not exceed the future value of access to the capital market. It is assumed in this subsection that contracts are permanent, the case where they are not is considered in the next. If people could be identified then in order for the possibility to arise for somebody of ability A_i to rent, there must exist $r_i \geq r'$ and K_i such that

$$(10) \quad r_i K_i \leq \{Y(A_i, K_i) - r_i K_i\} D(2)$$

There are three main possibilities:

(i) There are no r_i and K_i such that (10) is satisfied.

- (ii) There are r_i and K_i such that (10) is satisfied for either $i = 1$ or 2 . Without loss of generality it is assumed in the discussion of this case that (10) is satisfied for $i = 1$.
- (iii) There are r_i and K_i such that (10) is satisfied for both $i = 1$ and 2 .

In case (i) there is no possibility of achieving a return equal to renting to those with collateral and at the same time preventing default. In case (ii) there exist profitable renting contracts for A_1 's but A_2 's will always default on any contract yielding r' . In (iii) there are contracts such that both A_1 's and A_2 's can rent.

In fact this categorisation is not strictly correct since it assumes a constant rate r' for renting to those with collateral. r' will actually depend on the amount of capital used by each group in each period. Taking this into account there may for example be equilibria in which a proportion of A_1 's can rent without defaulting but if all do then r' will be such that no contract is viable. Such a general equilibrium approach adds greatly to the complexity of the analysis without in this case adding significantly to the content. It is for this reason that a partial equilibrium approach has been adopted.

For the lenders in the model there is no knowledge of the structure and it is not even known that possibilities (i) to (iii) exist. One approach they could use is to try to determine the structure of the model, evaluate the possibilities and then screen if the situation corresponds to something like (ii) or (iii). In the two group case such a procedure will be extremely costly and this will be greatly exacerbated

as complexity increases. Even if it turns out that renting to those without collateral is possible it is unlikely that these costs can be recouped by charging a rent higher than r' since this is again limited by the possibility of default. For these reasons this approach is likely to be infeasible for all practical purposes.

An alternative is to offer a contract which makes use of the fact that output can be observed ex post to ensure that nobody defaults, no matter what their output turns out to be. The contract would be secured on the lender's part by the collateral of his capital and on the part of the borrower by the prospect of future earnings. Such an agreement turns out to have a particularly simple form.

Let $\phi(Y)$ denote the schedule of payments to the lender. For the moment K is taken to be the same in every period. The more general case where K alters is considered in the next subsection. In order to prevent default the current payment must be less than the discounted value of future payments so that ϕ must satisfy the following inequality

$$(11) \quad \phi(Y) \leq \{Y - \phi(Y)\} D(2)$$

If lenders specify the contract and are profit maximisers ϕ is found from (11) with an equality. Rearranging

$$(12) \quad \phi(Y) = \frac{D(2)}{D(1)} Y$$

since $1 + D(2) = D(1)$.

The lender can therefore prevent default no matter what the level of output subsequently turns out to be, by specifying

a share contract with the lender's share being given by

$$(13) \quad \sigma' = \frac{D(2)}{D(1)}$$

However, for low levels of output the total value of payments from a share contract may be less than if the borrower defaults and the capital is lent out in subsequent periods to somebody with collateral. Hence if K is the amount of capital lent, then for Y such that

$$(14) \quad \frac{\sigma' Y}{K} D(1) < r' D(2)$$

it is better for lenders to allow the borrowers to default. The payment schedule should therefore specify that the borrower must produce at least Y' , where this is defined by (14) with an equality, in order for the contract to continue. The optimal schedule is then

$$(15) \quad \phi(Y) = \sigma' Y \quad \text{for } Y \geq Y'$$

In equilibrium the amount of capital lent must be such that the return to the share contract is equal to the return from renting with collateral. Let K' denote this equilibrium value.

In case (i) no share contract will be as profitable as r' for any K . In case (ii) there are a number of possibilities. If $Y(A_2, K) < Y'$ for all K equilibrium requires

$$(16) \quad \alpha \frac{\sigma' Y(A_1, K')}{K'} D(1) + (1 - \alpha) r' D(2) = r' D(1)$$

where α is the proportion of those taking the contract who are A_1 . The first term on the left hand side represents the return from A_1 's who do not default and the second the return to renting to those with collateral, the capital of those who defaulted in the first period of the share contract. The right hand side is the return to lending to those with collateral throughout. If for all K the value of the left hand side is less than the right no share contracts will be profitable; A_1 's are prevented from borrowing at all by the defaults of A_2 's.

Alternatively if $Y(A_2, K) \geq Y'$ it may be better to allow the A_2 's to continue even though they do not earn r' , rather than bear the cost of their default. In this case in equilibrium

$$(17) \quad \frac{\alpha \sigma' Y(A_1, K') + (1 - \alpha) \sigma' Y(A_2, K')}{K'} = r'$$

In case (iii) (17) will also be the equilibrium condition.

In order for a (σ', K') contract satisfying (16) or (17) to be viable, borrowers must not be able successfully to counter-offer contracts. No lender will accept any contract with capital $K' + \delta K$, no matter what the form of payment, since it will be worthwhile for all borrowers to do this and default rather than accept a (σ', K') contract.

If borrowers knew their production functions with certainty and this was generally recognised they could counter-offer a contract with capital $K' - \delta K$ to show that they did not intend to default, and offer a rent $r' + \delta r$. It is worthwhile for the borrower to do this provided

$$(18) \quad Y(A_i, K' - \delta K) - (r' + \delta r) (K' - \delta K) \geq (1 - \sigma') Y(A_i, K')$$

If

$$(19) \quad \frac{\sigma' Y(A_i, K')}{K'} > r'$$

a counter-offer which satisfies (18) exists. Lenders would always accept such offers.

However, provided it is thought borrowers do not know their production function with certainty lenders will not accept such counter-offers because of the conflict of interest which arises from the possibility of default. It may only be marginal for a borrower whether to default or not since in both cases he receives something: if he makes the current payment he can borrow and earn profits in future periods; if he defaults he keeps the payment but cannot subsequently gain access to the capital market. In the same situation there can be a much greater difference for the lender between defaulting and not defaulting since in one case he receives a full payment but in the other nothing at all.

In this case a reduction in capital to $K' - \delta K$ suggested by the borrower no longer acts as a definite signal of no-default since he is unsure of his production function. In choosing the counter-offer he attaches the wrong values as far as the lender is concerned to chances of default. Since the lender does not know the various possibilities and the probabilities attached to them by the borrower, and there is no way for the latter to convey them to him with manifest honesty, the counter-offer will not be accepted.

A similar argument as to why lenders do not attempt to partially screen by competing for high output borrowers with different contracts can also be made.

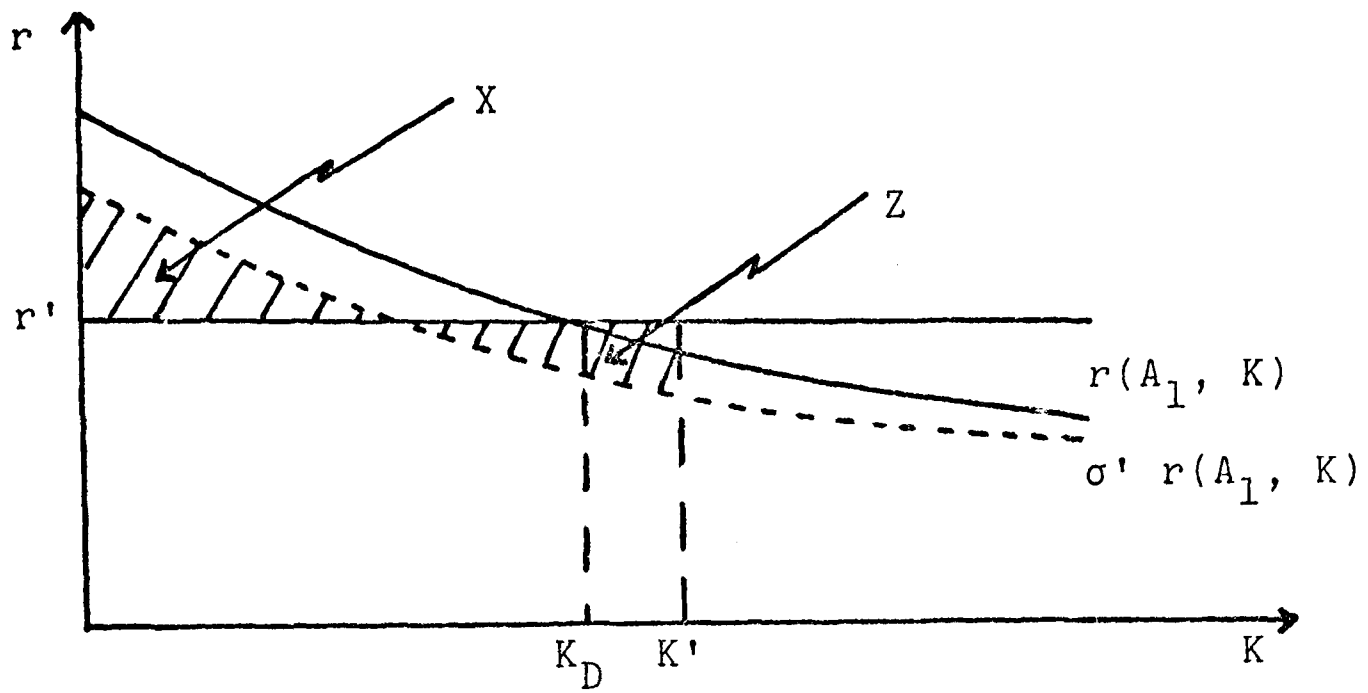
With the share contract all that has to be found are the appropriate σ' and K' . By starting the fairly low figures and gradually increasing them these should be relatively easy and cheap to find compared to a screening approach. Once the optimal σ' and K' have been found it follows from the above that there will be no competition for borrowers among lenders and they will simply offer the same contract to everybody.

As far as the allocation of capital is concerned share contracts result in lenders acquiring part of the pure profits and therefore may give more capital to borrowers than they would demand in an enforceable renting system with the same rate of return. For example, consider an equilibrium of the type defined by (16). Rearranging (16)

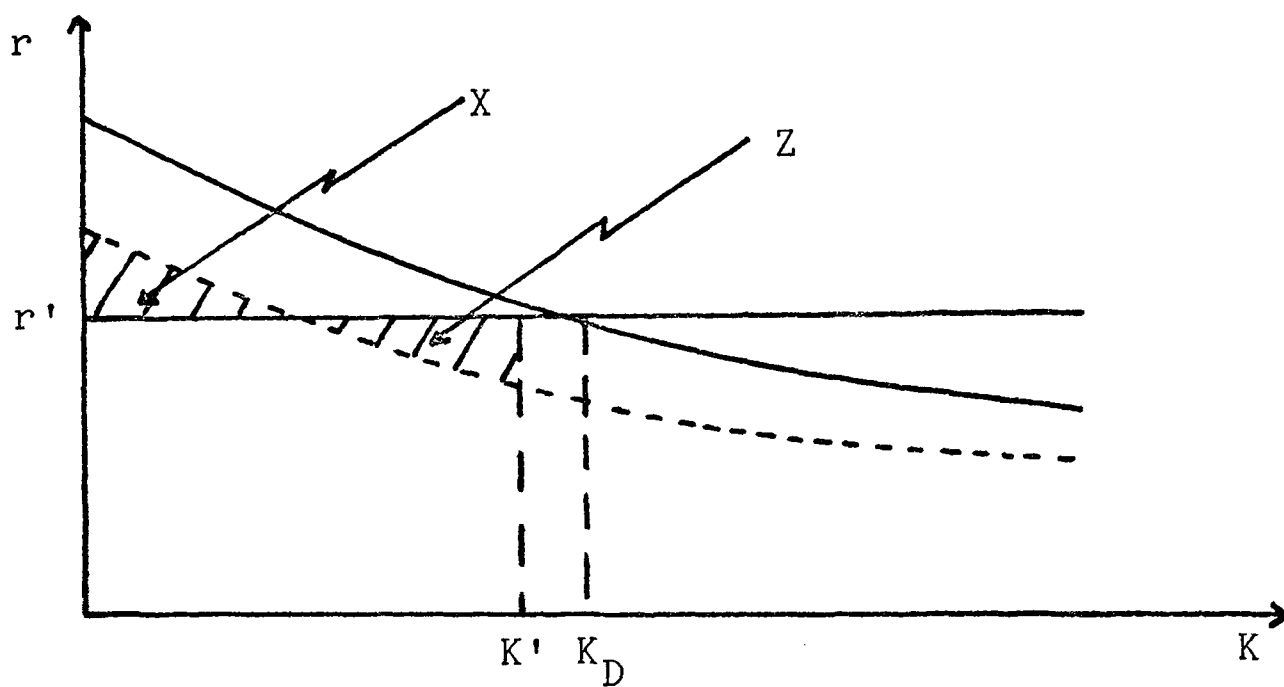
$$(20) \quad \frac{\sigma' Y(A_1, K')}{K'} = r' + \frac{(1 - \alpha) r'}{\alpha D(1)}$$

Three possibilities are illustrated in Figure 2. (20) implies area X is equal to the sum of Z and $(1 - \alpha) r' K' / \alpha D(1)$. In (a) it can be seen that $K_D < K'$. However, if σ' is small as in (b) or $(1 - \alpha) r' K' / \alpha D(1)$ is large as in (c) then $K_D > K'$.

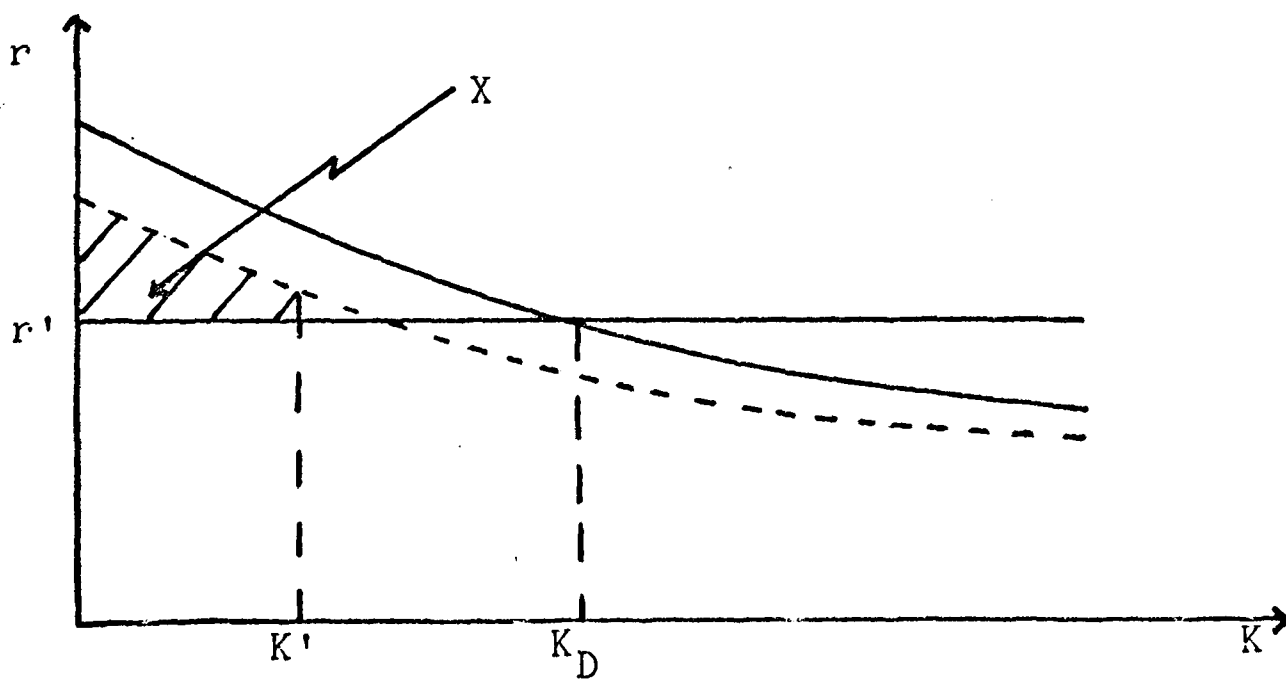
So far only the case where those without any wealth take share contracts has been considered. There are also circumstances where some of the wealthy would like to borrow more than their collateral allows them to rent. The assumption has been made that anybody who defaults will be unable to borrow again. The justification for this is that anybody who does so will be regarded as unable to produce enough to pay and a bad risk. Given this it seems reasonable that having defaulted it will still be possible to rent with collateral since in this case payment is guaranteed;



(a)



(b)



(c)

FIGURE 2

it will be renting without collateral that is ruled out.

The borrower has some idea of his output. If he expects the return on lending to him to be greater than r' he will not only wish to use his own capital but will also want to rent as much as possible at r' since this will increase his profits. Similarly if he expects the yield to be less than r' he would be better off renting his own capital to somebody with collateral and not borrowing any at r' . However, it will not be possible to do this since it is an implicit signal that the borrower expects a return less than r' and he will find it difficult to obtain loans. Given wealth K_w , people must use this and also rent K_B , which is defined in (7), with collateral. The payment schedule ϕ for the rest of the capital used must therefore satisfy

$$(21) \quad \phi(Y) \leq \{Y - r' K_B - \phi(Y)\} D(2)$$

The same argument concerning the conflict of interest applies so that lenders again specify the contract. Taking an equality as before and rearranging

$$(22) \quad \phi(Y) = \sigma'(Y - r' K_B)$$

A lower bound Y' is again placed on output; in this case it is given by

$$(23) \quad \frac{\sigma'(Y' - r' K_B)}{K - K_B} D(1) = r' D(2)$$

where K is the total capital used.

The analysis of the various equilibrium contracts is similarly altered. (16) and (17) are respectively changed to

$$(24) \quad \alpha \frac{\sigma'(Y(A_1, K') - r' K_B)}{K' - K_B} D(1) + (1 - \alpha) r' D(2) = r' D(1)$$

$$(25) \quad \frac{\alpha \sigma'(Y(A_1, K') - r' K_B) + (1 - \alpha) \sigma'(Y(A_2, K') - r' K_B)}{K' - K_B} = r'$$

so that K' depends on K_B and hence from (7) on K_w .

Implicitly differentiating (24) and (25) and using the facts that concavity implies average is greater than marginal product and $\sigma' < 1$ it can be seen that in both cases

$$(26) \quad 0 < \frac{\partial K'}{\partial K_B} < 1$$

If α is the same at all levels of wealth then as K_w and hence K_B increase the total amount a person can borrow also becomes larger but not by as much. The debt equity ratio is also greater the bigger is K_w . If A_1 's are taken to be the higher output group then if α increases with wealth these results remain valid. If α falls with wealth they may be reversed.

It may be worthwhile for those with wealth to borrow more than they can with collateral whether they are constrained or not. In the former case this is clear. In the latter it occurs because the group with the lower productivity is effectively subsidised by that with the higher, so that it may be worthwhile for them to borrow more than they would do if they were renting.

Formally, if people are constrained, borrowing more than

they can rent will be attractive to them if

$$(27) \quad \{\pi \sigma' + (1 - \sigma')\} \{Y(A_i, K') - r' K_B\} \\ > Y(A_i, K_W + K_B) - r' K_B$$

where $\pi = (K' - K_W) / K'$. If they are not constrained they prefer borrowing more than they could rent with collateral if

$$(28) \quad \{\pi \sigma' + (1 - \sigma')\} \{Y(A_i, K') - r' K_B\} \\ > Y(A_i, K_D) - r' (K_D - K_W)$$

Summarising briefly, in situations where information is scarce and costly so that screening is ruled out, an appropriately constructed share contract ensures that people borrowing without sufficient collateral only default when it is worthwhile as far as the lender is concerned to do so. The next subsection considers a number of extensions of the model.

2.2.4. Some extensions

It was argued in the introduction to the section that the case where criminal penalties would not be sufficient to prevent default was of interest. In Subsection 2.3. the extreme where there were no penalties to default was considered. Clearly the case where there is some punishment but this is not sufficient to prevent default should also be looked at.

When constructing payment schedules it is assumed for simplicity throughout this subsection that $K_B = 0$. Thus

if $P(M)$ denotes the expected penalty for defaulting on the amount M , taking into account the probability of being caught and so on, the no-default condition (11) becomes

$$(28) \quad \phi(Y) - P(\phi(Y)) \leq \{Y - \phi(Y)\} D(2)$$

To see the type of effect including penalties has on the payment schedule $P(M)$ is taken to be linear so that

$$(29) \quad P(M) = a + b M \quad b \geq 0$$

Substituting in (28), assuming equality as before and rearranging

$$(30) \quad \phi(Y) = \frac{D(2)}{D(1) - b} Y + \frac{a}{D(1) - b}$$

In this case the optimal contract has a share component with $\sigma' = D(2) / (D(1) - b)$ and a fixed component $a / (D(1) - b)$. If $a < 0$ then the latter is a payment from lender to borrower and so is a wage. Similarly if $a > 0$ the fixed component is a payment from borrower to lender and so is a rent.

One special case of the analysis with $b = 0$ and $a < 0$ is of particular interest. This is where, having defaulted, a person has alternative earning opportunities not previously open to him because for example he has spent his time administering his capital. If alternative earnings per period are denoted by E , the no-default constraint becomes

$$(31) \quad \phi(Y) + E D(2) \leq \{Y - \phi(Y)\} D(2)$$

which gives

$$(32) \quad \phi(Y) = \sigma' Y - \sigma' E$$

The lender pays a wage which is less than alternative earnings but also gives a share in output.

A further extension in this direction is to assume alternative earnings depend on the group to which the person belongs. One particularly simple case to consider is that where

$$(33) \quad E = e Y(A_i, K')$$

Here

$$(34) \quad \phi(Y) = (1 - e) \sigma' Y$$

and a pure share contract is optimal but the lender's share is lower than when there are no alternative earnings.

One of the other simplifying assumptions of the model was that there was no disutility from controlling wealth. The effect of including this is similar to that of alternative earnings. If E is reinterpreted as the compensation for the disutility of controlling wealth the no-default constraint is

$$(35) \quad \phi(Y) \leq \{Y - \phi(Y) - E\} D(2)$$

which again leads to (32).

In the previous subsection there was taken to be one generation with infinitely long lives. By relaxing the

latter the analysis becomes applicable to the many generation case. With finite lives where the date of death is known, the share of the lender steadily declines once the time to death of the borrower is less than H periods. In order for the rate of return to be equated to r' , the amount of capital lent must also decrease. In the last period it is always worthwhile for the borrower to default since he has no prospect of future earnings. The lender cannot combat this by ending a contract one period earlier since then the borrower simply defaults in the penultimate period and so on. Instead the lender must accept that he will have to make a payment to the borrower in the last period conditional on his not previously defaulting. Also the lender should not give the borrower any capital in the last period. The borrower thus retires before he dies and receives a pension in his last period.

The terms borrower and lender have been used in the previous analysis to denote the person who controls production and the person who supplies the capital respectively. It is important to point out that the model is not confined to the case where there are explicit loans. It applies equally well to situations where the loan is implicit as in the case of the employer and employee in which the latter does not work under the close supervision of the former. Similarly with managers and owners in a joint stock company. The important characteristic of applicability is the independence of operation of the party in control and the supply of factors of production by another party.

In applying the model to these various cases it is

important that the notion of output in the model is correctly reinterpreted. It corresponds to the maximum amount over which the entity referred to as the borrower has control at any one time.

Capital was taken to be immobile and perfectly durable. The assumption of immobility can be relaxed in certain circumstances. If the capital is initially in a form like money it is necessary for the lenders to ensure that borrowers do not abscond with the principal as well as the rent. Provided that in its eventual application it will consist of physical assets, one way to ensure that the capital is put to its proposed use is for the lender to maintain financial control during the initial stages while it is turned into an immobile form. If the borrower chooses the project his ability becomes incorporated in the capital. There will be no chance of reletting it at a higher rate to another person and so in this case there will be no lower bound Y' on Y which borrowers must achieve to remain in control.

If the final use of the capital is movable then it may not be possible to borrow without collateral or criminal penalties since restricting access to the capital market to non-defaulters will no longer be effective; those who default already have their capital and cannot be separated from it.

Growth and depreciation of capital both involve changes in K . In the derivation of ϕ it was assumed that K was constant through time. Since it is marginal not average output that is relevant and production functions can cross so that the two are not in general related, observation of output generates no new information to lenders on which

to base any decision to change K .

However, borrowers have some idea of their marginal product. It is not possible for lenders to ask them directly whether it is worthwhile to increase K since with a share contract and a positive marginal product of capital, all groups have an incentive to gain more capital. Nevertheless it is possible for lenders to extract the information.

At the end of the current period after output has been produced the price of capital p_K is

$$(36) \quad p_K = r' D(2)$$

Consider what happens if lenders suggest an expansion which is financed internally so that each shareholder including the borrower pays in proportion to their share. If $r(A_i, K') > r'$ the borrower will choose δK so that $r(A_i, K' + \delta K) = r'$ if allowed to do so, since if

$$(37) \quad (1 - \sigma') p_K \delta K < (1 - \sigma') r(A_i, K' + \delta K)$$

both he and the other shareholders have an incentive to increase δK . By using internal finance borrowers' and lenders' interests are made to coincide and the former's statements can be believed. Also the borrower will not default since

$$(38) \quad \sigma' Y(A_i, K') + (1 - \sigma') p_K \delta K \leq (1 - \sigma') Y(A_i, K' + \delta K) D(2)$$

However if

$$(39) \quad \sigma' Y(A_i, K') + (1 - \sigma') p_K \delta K > Y(A_i, K')$$

the expansion may need to take place over a number of periods.

Similarly if $r(A_i, K') < r'$ the borrower will refuse to expand. Provided K' is not reduced he will not default though since he is in the same position as before the lender made the offer. If there were no default it would be worthwhile for lenders to reduce the capital if $\sigma' r(A_i, K') < r'$ so that refusal to expand is not a definite signal that a reduction is desirable. However, with a lender's share of σ' it will not be possible to reduce K' even if this condition were satisfied since

$$(40) \quad \sigma' Y(A_i, K') > (1 - \sigma') Y(A_i, K' - \delta K) D(2)$$

and such an action would precipitate default.

If the structure of the model were known it would be possible to determine whether or not it is worthwhile lowering σ' in order to allow a reduction in K without default. However, it is not known and cannot be discovered except at great cost. Moreover, the information cannot be extracted from borrowers since interests are no longer shared. Thus although growth is feasible a reduction in the size of K' will be much more difficult to achieve.

The assumption that K was constant which was made in deriving ϕ is thus not strictly correct. However, a (σ', K') agreement will still usually be the best simple contract to enter into: it is not usually possible to reduce

K' and when it can be increased, although σ' is not as high as the share could be in expansionary periods without precipitating default, it allows the cost of the new capital to be shared and more importantly also makes it possible for the borrower and lender to share common interests so that the latter can believe the former's statements.

The problem which prevents the contraction of capital also has implications for the finance of depreciation. When capital is not completely durable it is gross output that is relevant when determining the payment of borrowers since this is the quantity over which they have independent control. Hence they receive $(1 - \sigma') Y$, where Y is gross output, and the total payment to lenders is $\sigma' Y$. However, lenders must use part of this to maintain the capital stock since otherwise the situation will be as in (40) and borrowers will default. If $\Delta(K)$ is the depreciation of capital K per period and p_K is again the price of capital, the return to the lender r_{si} , of a share contract with a person of ability A_i is

$$(41) \quad r_{si} = \frac{\sigma' Y(A_i, K) - p_K \Delta(K)}{K}$$

In the extreme case where $p_K \Delta(K) > \sigma' Y(A_i, K)$ for all A_i and K , then depreciation is effectively equivalent to mobile capital and again no profitable lending will be possible without collateral or criminal penalties.

Finally, one of the central assumptions of the model was that a person's default could be observed and this would prevent his future entry to the capital market. Although

this is fairly reasonable for a centralised market such as a stock exchange it may be rather strong in a spatial market such as that for land. To see the effect of the relaxation of this assumption the model can be changed in the following way. The borrower perceives that if he defaults once he can move and there is a probability γ that his default will be noticed; if he defaults a second time then he perceives that his notoriety will be widely observed so that he will be barred everywhere.

His decision on whether or not to default the first time then depends on whether

$$(42) \quad \phi(Y) \leq \gamma \{Y - \phi(Y)\} D(2)$$

As before assuming equality and rearranging

$$(43) \quad \phi(Y) = \frac{\gamma D(2)}{1 + \gamma D(2)} Y$$

The optimal contract is still a share contract but the proportionate payment to the lender is reduced. The theory can also be modified with more complicated sequences of probabilities of discovery with similar results.

2.2.5. Concluding remarks

This chapter has illustrated the variety of theories of share contracts. In Section 1 their efficiency, risk sharing and screening properties were recounted. In Section 2 it was argued that another property was that when screening was costly or impossible they are the profit maximising contracts which provide an incentive not to default.

These roles are not necessarily incompatible. For example the incentive effects and no-default properties of share contracts will both be desirable in situations where production takes place independently. Share contracts may thus fulfil a number of roles simultaneously.

It can perhaps be argued that theory has two main purposes. The first is to describe an advantageous way in which to conduct affairs and the second is to combine this with the hypothesis that individuals are maximisers to explain observed phenomena. This chapter has been mainly concerned with the former. However, it is perhaps worth briefly mentioning that the theory of default above is consistent with one of the main characteristics of share contracts in an agricultural context which other theories have found very difficult to explain.

The French word for sharecropping is 'metayer' which as Mill (1848) pointed out implies a fifty-fifty split. He viewed the distinguishing characteristic of share contracts as being the fact that they were determined by custom rather than competition. In support of this he quotes (pp. 348-349) Sismondi (1814) (pp. 41-42) who wrote of Tuscany where he was himself a metayer landlord:

'This connexion (of input commitments) is often the subject of a contract, to define certain services and certain occasional payments to which the metayer binds himself: nevertheless the differences in the obligations of one such contract and another are inconsiderable; usage governs alike all these engagements, and supplies the stipulations which have not been expressed: and the landlord who attempted to depart from usage, who exacted more than his neighbour, who took for the basis of the agreement anything but the equal division of the crops, would render himself so odious, he would be so sure of not obtaining a metayer who was an honest man, that the contract of all the

metayers may be considered as identical, at least in each province, and never gives rise to any competition among peasants in search of employment, or any offer to cultivate the soil on cheaper terms than one another.'

This view of the fixity of the share has also gained some support among modern writers as being relevant (see for example Newbery and Stiglitz (1978) and Bliss and Stern (1980)), but they provide no explanation for this. However, the phenomenon is consistent with the theory above concerning default. It was argued there that because of the conflict of interest there would be no competition for borrowers among lenders; they would simply all offer the same contract. The observation by Sismondi that lenders who tried to obtain a higher than customary share would obtain dishonest borrowers is also particularly interesting since this is predicted by the theory.

The problem in using the theory in this context concerns the magnitude of the lender's share being around one half. Reasonable assertions concerning the value of $D(2) / D(1)$ would suggest a much higher value than this. However, the assumption of only being able to default once and no alternative earnings are rather implausible in an agricultural context and provide an explanation of why the observed share should be lower than the analysis of Subsection 2.3. suggests.

In conclusion, share contracts have been seen to have a rich variety of properties. The main purpose of this chapter has been to point out and investigate their role in preventing default.

CHAPTER 3

INHERITANCE, SAVINGS AND TAXATION WITH
AN IMPERFECT CAPITAL MARKET3.1. A survey of theories of wealth distribution

In the previous chapter the distribution of wealth was taken as given. The analysis stressed the problems involved in borrowing capital to use for production and this suggests that with differing abilities the pattern of ownership of wealth may be an important determinant of the output of an economy and other variables of interest. If so, then attempts to redistribute wealth may be damaging because of this dependence. This is the subject the current chapter is concerned with: in the next section inheritance, savings and the effect of taxation, and in particular inheritance taxation are considered in the extreme case where borrowing is ruled out; the remainder of this section contains a survey of previous work on wealth distribution and taxation.

There appears to be some agreement that in many private ownership economies wealth is very unequally distributed. Atkinson and Harrison (1978)(p. 261), for example, estimate on a realisation basis of valuation that in the U.K. in the years 1966-72 the top 1 per cent of wealth owners owned 33 per cent of the wealth, the top 5 per cent 57 per cent, and the top 10 per cent 69 per cent.

It is usually suggested that there are two main factors leading to this inequality: inheritance and life cycle savings.

There have been a number of studies relating a sample of people's bequests to their inheritance. The technique involves observing large estates and then either finding the inheritances of the people who left them by tracing the bequests of the previous generation of the family or alternatively discovering the estates of their children. Wedgwood (1929) was the first to undertake the tracing back exercise for the U.K. in the years 1924-26. It has subsequently been repeated for different samples by Harbury (1962) (U.K.: 1956-57), Harbury and McMahon (1973) (U.K.: 1965) and Harbury and Hitchens (1976) (U.K.: 1973) who again traced back and Menchik (1979) (Connecticut, U.S.A.: 1931, 38, 44) who traced forward. These studies all showed that the chances of leaving a large bequest were greatly increased if the parents were wealthy and so suggest that inheritance may play a significant part in explaining the upper tail of the distribution.

It would appear that a large amount of saving takes place in order for consumption to be transferred from one period of a person's life to another; perhaps the most obvious example being that of saving for retirement. In an egalitarian society in which everybody had the same market opportunities and behaved in the same way, this type of saving would still produce inequality at any particular time since people of different ages would have different amounts of wealth.

Atkinson (1971) used a simple model in which everybody was the same to investigate how much inequality might arise from life cycle savings of this type. Flemming (1976)

extended the assumptions to allow, among other things, for earnings inequality. Taking into account this and certain other plausible factors such as random mortality, it would appear that this type of saving is capable of explaining a distribution of wealth similar to that observed in the U.K. in recent years, except for the top 1 per cent of wealth-holders.

These results are often interpreted as implying that inheritance is important in accounting for the upper tail of the wealth distribution and life cycle savings for the rest. It is necessary to point out though that this is a speculative conclusion. The life cycle calculations only give some idea of the upper bound of the importance of this factor; they are not a good substitute for detailed empirical work distinguishing the contribution of the two factors, which has yet to be undertaken. Nevertheless, it does seem reasonable that theoretical models should be concerned with both.

Models considering the development of wealth-holding over time and the determination of wealth distributions which are based on individual behaviour consist of some or all of three main components. Firstly there are customs of bequest and marriage, and a demographic structure. Secondly there are consumers who determine factors by choosing their labour supplies, life cycle savings and bequests. Finally there are firms and production functions which close the system and allow prices to be determined endogenously.

The consideration of these components, especially when combined, tends to be rather complicated and much of the

literature consists of numerical simulations. Although this allows some progress to be made it is not easy to understand why particular results occur, because of the difficulty in determining interactions. If some element turns out to be insignificant it is difficult to identify whether it really is unimportant or whether there is a counteracting influence special to the model. Authors often obtain differing conclusions and it is not clear which altered assumptions have caused this. The literature will be considered below by looking at the various sorts of component that have been used; the speculations concerning their importance will for these reasons necessarily be tentative.

The role of the customs or laws determining the way in which the wealth of parents is divided among children has long been recognised. Wedgwood (1929) (p. 72) for example quotes De Tocqueville (1850) (pp. 57-58) who wrote

'Framed in one way the law of succession combines and concentrates property and power in a few hands... In countries where the legislature has established an equal division of inheritances, property and particularly landed property must have a permanent tendency to decrease.'

Many authors have found that whether there is primogeniture with the family's entire wealth left to the eldest, or a more equal division among the children, is one of the most important factors in determining the inequality of wealth-holding (see, for example, Pryor (1973) and Blinder (1976)).

Historically it is a well known feature of many societies that the role of men and women as property-holders has differed. If it is the custom or the law that men hold all the wealth then this will have marked effects on its

distribution among individuals: there is the direct consequence that only half the population has the opportunity of holding wealth so that the distribution among households may be the appropriate measure. However, this is not the only means by which the comparative status of men and women affects the distribution. If both are wealth-holders or there is a dowry system, Meade (1964) and Blinder (1973) have stressed the importance of patterns of marriage. The rich marrying the poor and passing their wealth on to their children will lead other things being equal to less inequality than the rich marrying the rich. The simulations of Blinder (1976) suggest this may be an important factor.

If bequests are spread among children then population growth will tend to dissipate inherited wealth. The effect of this may be accentuated or reduced if there is differential fertility. Pryor's (1973) results suggest this can be important. Different sized age groups also alter the inequality of life cycle savings since the ratio of the young, who have little wealth, to the old, who have more, is changed. However, calculations by Atkinson (1971) suggest that for plausible growth rates this factor is probably not very significant.

An important determinant of wealth and income distributions would intuitively appear to be the correspondence between the intergenerational transmission of wealth and ability. If able parents earn large amounts which permits them to leave bequests to children who are also able, then inequality would be expected to be greater than if parents' and children's ability were unrelated.

One of the most actively publicised controversies of

recent years concerns the role of the environment in which a child is brought up as opposed to genetic factors in determining ability. This is a highly controversial area and the only definite conclusion appears to be that both have some importance but that it is very difficult to distinguish the contribution of each (see, for example, Goldberger (1979) for a survey). This debate is significant for the construction of economic models because presumably if genes are important then it is the earnings capacity of parents that is relevant in determining the ability of children, whereas if environment is, household consumption will be pertinent. The latter case would be difficult to analyse because of the endogeneity of ability and does not appear to have been considered. The more usual assumption is to base the earnings capacity of children on that of their parents.

A property of the relationship between the ability of parents and children which appears to be widely accepted is that of regression towards the mean. Bevan (1979), Flemming (1979) and Bevan and Stiglitz (1980) have incorporated this demographic aspect into their models. Earnings capacity Y is assumed to be related to a characteristic A . The hereditary relationship for A is taken to be Markovian in form.

$$(1) \quad A_{g+1} = m A_g + Z$$

where A_g is the characteristic of the g^{th} generation, m is

a parameter determining the degree of regression towards the mean and Z is an independent random variable.

The relationship between Y and A and the distribution of these and Z are taken to be such that the overall distributions remain constant while still allowing regression towards the mean within each family. For example Bevan (1979) outlines a method which allows any distribution of earnings to be considered. Y and A are taken to be perfectly rank correlated. A and Z are distributed $N(0, 1)$ and $N(0, \sqrt{1 - m^2})$ respectively. Thus given a particular Y_g in the q_g^{th} percentile of the earnings distribution there is a corresponding value of A_g in the q_g^{th} percentile of $N(0, 1)$. A_{g+1} is then determined by (1) so that it is distributed $N(m A_g, \sqrt{1 - m^2})$. Given any realisation of A_{g+1} in the q_{g+1}^{th} percentile of $N(0, 1)$ there is a corresponding value in the earnings distribution which gives Y_{g+1} . It follows from the properties of normal distributions that for each generation the distribution of A remains constant, even though within families there is regression towards the mean. Also it can be seen that as m approaches 1, children's earning capacity is determined almost completely by that of their parents whereas with $m = 0$ there is no link between generations.

In any model in which bequests depend on earnings a hereditary mechanism of the type described above might be expected to play a significant part in determining the wealth distribution. Bevan (1979) considered the case in which bequests are chosen to smooth consumption between generations; the way in which this is done is described below. Surprisingly,

his simulations indicated that m apparently played only a very small part in determining the wealth distribution although it was significant as far as the inequality of consumption was concerned. By making various drastic simplifying assumptions Bevan and Stiglitz (1980) were able to derive analytic expressions for the steady state variance of wealth. These suggest the degree of regression towards the mean may be more important than Bevan's (1979) results indicated.

The second component of models, consumers' behaviour, usually either involves some form of ad hoc savings and bequest rule or an extension of the life cycle model. Meade (1964) and Stiglitz (1969) who assume savings are a linear function of income, provide an example of the first of these.

The other approach is perhaps more theoretically satisfactory since it is based on maximising behaviour. In the context of the theory of the consumer under uncertainty, Yaari (1965) suggested an additively separable lifetime utility function depending on the path of consumption C_t and the amount bequeathed B . Atkinson (1971a) adopted a similar approach for an analytical investigation of wealth-holding. He assumed that the duration of life was known with certainty to be T years and used a utility function of the form

$$(2) \quad U = \int_0^T u(C_t) \rho_t dt + \psi(B)$$

where ρ_t is the discount factor. Transactors are assumed

to choose C_t and B to maximise (2) subject to a budget constraint. A model of consumers very similar to this is used as the basis for capital supply in the next section and Atkinson's main results are given there. This form has also been adopted by a number of other authors (see, for example, Blinder (1976)).

This approach has been criticised on a number of grounds. No justification is usually given for the additive form except convenience. It is not difficult to construct arguments as to why it should not be additive. For example if C refers to household consumption and parents are concerned with the welfare of their children, C and B may well be substitutes; offspring can either obtain a large lifetime utility by having the happy childhood associated with the high household consumption of parents, or from enhanced possibilities in later life if they receive an inheritance. Alternatively if the motivation for bequests is that of giving to others in order to assuage a conscience arising from high consumption, C and B may be complements.

This also illustrates the other main criticism of (2) which is that it does not specifically incorporate the motive for giving. A number of authors have argued that parents are concerned with the utility and hence consumption of their children. Depending on how concerned parents are and the expected earnings of children, they may wish to bequeath wealth in order to equalise consumption between generations (see, for example, Stiglitz (1978), Bevan (1979) and Flemming (1979)).

The utility functions in these models are usually taken

to be of the form

$$(3) \quad U_0 = u(C_0) + \sum_{g=1}^H (n \rho)^g \chi_g E U_g$$

where U_g is the lifetime utility and C_g the consumption of the g^{th} generation, H is the number of generations that are taken into account directly by the current one, n is the number of children in each generation, ρ is a discount factor, the χ_g 's are the weights given to subsequent generations and E is the expectations operator. Each generation chooses C_0 and their bequest to maximise U_0 subject to a budget constraint.

H is commonly taken to be 1, so that parents directly take into account only their own children; they are concerned with subsequent generations indirectly though, since the children's utility depends on that of the grandchildren and so on. Although it seems reasonable that parents should be less concerned with more distant descendants it is perhaps rather extreme to assume they directly value only their children's utility. The inclusion of grandchildren could make a considerable difference since there may be a multiplier effect: an increase in bequests to children which in turn allows them to increase their bequests to their children has a number of consequences for the utility of the original bequeather. The increased utility of the grandchild enhances the latter directly. However, there is also an indirect effect since the improved situation of the grandchild makes the child better off which in turn makes

the original parent better off. It may therefore be an important parameter in determining bequests and hence the wealth distribution. Bevan (1979) found that even with $H = 1$, so that the grandchildren and subsequent generations enter only indirectly, the wealth distribution was very sensitive to the number of generations that were considered when maximising (3).

The number of children is usually taken to be determined exogenously. With modern methods of birth control and a utility function such as (3) this seems rather unrealistic. With (3) the endogenous choice of n will, assuming u and $\partial u / \partial C > 0$ for all C , lead to people having the maximum number of children that is physically possible. This seems an undesirable feature but can be eliminated by assuming u depends on n as well as C_0 so that, for example, it is average consumption within the family that is relevant for utility.

χ_g is the parameter which determines the degree of altruism and is usually taken to be between 0 and 1. On the face of it there seems to be no reason why parents should not weight the utility of future generations more than their own, and it would appear desirable to allow this. However, it should be noted that such a configuration may lead to an unstable outcome in which capital per head continuously grows. Bevan's (1979) results indicate χ may play an important role.

In order to determine U_0 , people have to evaluate EU_g for each subsequent generation. There are two problems with this. Firstly people have to make estimates concerning

their successors' earning capacities and secondly assumptions about their behaviour. In the latter case it is commonly assumed that all generations behave in the same way. The determination of earnings capacity usually takes account of likely technological progress and where it is hereditary, regression towards the mean is also allowed for. Given that they know their own ability individuals are assumed to estimate the earnings capacity of future generations as though they were deterministically given by a linear approximation of the form

$$(4) \quad Y_g = \gamma^g \{ \hat{m}^g Y_0 + (1 - \hat{m}^g) \bar{Y}_0 \}$$

where γ is the rate of technological progress, $\hat{m}$ is analogous to m in (1), and $\bar{Y}_0$ is the mean income in the current generation. The significance of the degree of regression towards the mean has already been discussed. Flemming's (1979) results suggest γ may be important.

Two other main assumptions are implicit in (3). As with (2) there is an additive structure which can be similarly criticised. Also only the utility of subsequent generations is taken into account. Although some asymmetry is perhaps justified it would be more satisfactory if the utility of preceding generations were included in the utility function. Bevan's (1979) work indicates that the inclusion of negative bequests may significantly alter the results.

It can be seen that even when simplified, the optimisation of a utility function such as (3) is an extremely complex procedure. The use of one such as (2) may not only provide

a useful simplification but may also be a more appropriate representation of behaviour since the costs of going through the sort of calculation associated with (3) are so high.

The desire to equalise consumption between generations is certainly not the only, and may not even be the most important reason for leaving bequests. There are many examples of childless people who leave large fortunes, and rich people who do not bequeath wealth to their children. There are numerous explanations of bequest behaviour which are such that (2) is the appropriate formulation. For example, it could perhaps be argued that the longing for some form of immortality, to leave some manifestation of one's existence is a powerful human motivation and many legacies particularly of family firms, donations to museums and so on are consistent with this. Thus although (2) does have its disadvantages it is not clear that (3) is superior.

So far only voluntary bequests have been considered. Even if consumers are entirely selfish, with an uncertain timing of death there may be legacies because of capital market imperfections. Flemming (1976, 1979) has stressed two particular examples: imperfect annuity and housing rental markets. The first makes it difficult to consume all wealth without risking living at a low standard if an old age is reached. The second may make it worthwhile to hold an asset until death. Flemming's calculations suggest this may be an important factor in determining wealth distributions.

The final component of models concerns firms and production functions. Together with competitive assumptions these

close the system and allow prices to be determined endogenously. Where production functions are included, output Y is usually taken to depend on the amount of capital K and labour L used and there are constant returns to scale.

$$(5) \quad Y = L y(k)$$

where $k = K / L$, the capital labour ratio.

It seems possible that pure profits may play an important role in determining the distribution of wealth and income, and the assumption of constant returns to scale is rather undesirable in this context. However, it provides the great simplification that production can be aggregated and there is no need to consider individual firms and rights to pure profits.

Compared to the other two components the effect of including a production function and determining prices endogenously have been relatively neglected. Stiglitz (1969) considered the development of wealth-holding among individuals in a neoclassical growth model and the inclusion of a production function allowed steady states and their stability to be analysed. Pryor's (1973) simulations included one, but the effects of varying the parameters of this were particularly difficult to interpret. Laitner (1979) also incorporated a production function into a growth model and showed that the bequest behaviour associated with a utility function of the same type as (3), prevents the steady state interest rate rising too far above the golden rule level.

Other authors have tended to ignore the production function and have taken prices as exogenously determined. However, where variations in the interest rate are considered (see, for example, Bevan (1979)) bequests appear to be quite sensitive to the assumed rate of return and so the failure to include a production function may be a serious deficiency.

The approach considered above is based on individual behaviour. Another tradition of analysis focuses on the distribution of wealth without specifying the exact relationship with individual behaviour. Since the analysis of Section 2 is of the former type these distributional models are not considered further here. They are surveyed by Atkinson and Harrison (1978) (Ch. 8.3., pp. 219-226).

It has been argued by many authors that the degree of inequality of wealth which appears to prevail in countries such as the U.K. is undesirable and that measures should be taken to reduce it (see, for example, Wedgwood (1929), Meade (1964), Sandford (1971) and Atkinson (1972)). It has been described above how even in an egalitarian society differences in wealth may arise because of life cycle savings; such inequality is not usually opposed, because it arises from foregone consumption and there seems little justification for discriminating against people who enjoy consuming in the latter part of their lives. Essentially it is the large share of top wealth-holders which is regarded as objectionable, because of the greatly enhanced opportunities and power of these people. Since the evidence appears to indicate that such inequality arises to a large degree from inheritance it is this which has been the main focus of pressure for reform. More generally it is the

right to transfer wealth that is considered since a bequest which yields utility can be thought of as a gift at death and any attempt to reform inheritance may simply lead to a rescheduling of the gift.

Arguments for and against the modification of this institution can be divided into two categories: ethical and consequential.

The case for equality is usually taken as a basic premise; it is regarded as desirable in its own right (see, for example, Atkinson (1972) and Sandford (1971)).

Two main ethical justifications of inheritance have been put forward. The first concerns the concept of property which asserts that people have a natural right to do what they like with their possessions and any attempt to interfere with this is wrong. Wedgwood's (1929) perception of his contemporaries was that such a view had 'gone out of fashion' (p. 188). Atkinson (1972) is unsympathetic to the notion of a right in this context and argues 'the right to convey property is in no sense an absolute one, and it is subject to any restrictions that society chooses to impose' (p. 83). Current opinions on this question would appear to be a matter for speculation.

The second justification is that an individual has a duty to support his family and that this does not cease at death. Wedgwood (1929) (pp. 189-90) writes

'One is expected to conjure up a picture of a penniless widow, or of orphans in their 'teens. Their father's hard-earned savings have rightly been put by for their benefit. Have they not a moral claim to these savings?'

There appears to be widespread agreement that such an argument

is valid but is only relevant in a very small number of cases. In general it is taken to be a poor vindication of inheritance.

There remains the other category of consequential arguments. The most usual reason put forward in favour of inheritance concerns saving. It is often suggested that the desire to give bequests is a major motivation for accumulation and any attempt to restrict inheritance will lead to a reduction in the supply of capital. Atkinson (1971a) investigated this proposition using a linear tax based on the size of wealth transfer and a utility function of the same form as (2). He showed that although the effect of the tax was always to reduce net bequests, gross bequests and hence savings might in some circumstances be increased in order to offset the effects of the tax.

If utility is gained from the act of giving as in (2) then by changing the base of the tax it may be possible to create a more equal distribution without affecting savings. Meade (1964) categorises four possible systems of determining liability. The first is the size of bequest or total lifetime wealth transfer which is the basis for the current U.K. capital transfer tax. The second is the size of inheritance or gift received by any particular person. Making the schedule progressive will then provide an incentive to bequeath to a larger number of people which may result in reduced inequality. This will not necessarily be the case though since, for example, groups of rich people may simply divide all their wealth among each other's children so that the outcome can be the same as if each left to their own. To prevent this

sort of occurrence, the third proposal is that the schedule should also depend on the wealth of the inheritor. There would again be a disadvantage though in that those people who accumulated their own wealth would be affected by this. The final proposal is that the tax should be based on a person's lifetime capital receipts by inheritance and gift.

However, if utility depends on who the bequest is given to as in (3) there may be little difference between such systems. Also with this type of tax administration costs are usually high and may be a more important factor in determining the basis for liability.

Stiglitz (1978) has pointed out that even if inheritance taxes lower saving they may not reduce wealth inequality. This occurs because a change in the capital labour ratio alters the distribution of income which in turn affects the inequality of wealth. He also argues that even if this general equilibrium effect were offset for example by an appropriate debt policy, inequality of consumption could still be increased by a wealth transfer tax. In a static context transfers from rich to poor are clearly inequality reducing. Any tax on these gifts will thus increase inequality. If the interest rate is zero, transactors have a utility function as in (3) and there is regression towards the mean, inheritances are on average equivalent to transfers from rich to poor in a static situation. Provided the interest rate is sufficiently low inheritance taxes may therefore increase the inequality of consumption.

This raises the question of the appropriate focus of analysis. Is it wealth inequality per se which is of interest,

the associated distribution of income or consumption, or something else? Atkinson (1970) argued that conventional measures of inequality such as the coefficient of variation have little inherent economic rationale, are difficult to interpret and fail to make explicit the trade-off between inequality and average levels. He suggested that income distributions should be ranked on the basis of a social welfare function. A similar argument can be made for the consideration of wealth inequality. In this case there is the added advantage that the problem of distinguishing life cycle inequality from inequality due to inheritance and other factors which are regarded as less defensible can be avoided by defining social welfare in terms of the lifetime utility of individuals. If the possession of wealth involves advantages over and above the accruing income, such as security and independence, which is the usual justification for considering wealth inequality per se, then these can be included in the specification of social welfare. In this view the evaluation of inheritance and other taxes involves two stages. Firstly an analysis of incidence to determine the effects of a tax and secondly the use of a social welfare function to summarise these and give a basis for comparison. This is the approach which is adopted in the next section.

Although an inheritance tax might be expected to be one of the most effective as far as altering the wealth distribution is concerned it is not the only relevant one. In a general equilibrium model any tax could change the wealth distribution; however, attention has usually been restricted to other taxes which might also be expected to have a significant

impact namely a wealth tax and a capital levy.

Atkinson (1971a) considered these and compared them to a wealth transfer tax in the partial equilibrium framework referred to above. On the basis of their effect on a representative individual and a concern for a given redistributional impact with the least effect on savings, a capital levy was found to be the most desirable tax. However, it was shown this might have a merely temporary effect. The only definite ranking for the other two concerned the case where the elasticity of the marginal utility of consumption was greater than one; there was then a certain critical level of lifetime wealth above which the capital transfer tax dominated over the wealth tax. Other authors tend to have considered these taxes in a less formal way (see, for example, Sandford (1971)).

Taxation is not the only method of achieving greater equality of wealth; the permanent transfer of assets to the State rather than a capital levy whose effect may only be temporary will also have this effect. The question of public versus private ownership is a very complex one. It will not be considered further here but it should be noted that there may be cases where this provides a better alternative than taxation.

Finally, Atkinson (1972) suggested that the problem of gaining access to capital may also provide a consequential argument in favour of inheritance. This aspect seems to have received very little attention. The purpose of the next section is to develop a model to consider it.

3.2. A theory of inheritance, savings and taxation with an imperfect capital market

3.2.1. Introduction

The analysis of Chapter 2 suggests that capital markets may be very imperfect. The problem of default can make it difficult if not impossible to obtain loans. In this section the extreme case where people are unable to borrow at all is considered; if they wish to produce the only capital they can use is their own inheritance and savings. The productive capacity of the economy thus depends on the distribution of wealth as well as total factor supplies.

As was mentioned in the last section it is often suggested that the ability of parents is correlated with that of their children. Combining this with the assumption of imperfect capital markets it is sometimes argued that the institution of inheritance ensures that some of the good entrepreneurs start their careers with substantial amounts of capital. In this view inheritance taxation is undesirable since it reduces the amount these people start with.

Others argue that the correlation of abilities is small and the main result of inheritance is to permit heirs to lead a lavish lifestyle squandering the fortunes acquired by their ancestors. In this view taxation of bequests prevents consumption by those with a low marginal utility of income and has little effect on the production of the economy. On this basis the tax is regarded as desirable on a wide range of ethical criteria.

The purpose of this section is to develop a model to examine the role of inheritance and saving in a model in which there are two ability groups which cannot be distinguished and an imperfect capital market. The first group consists of able entrepreneurs who either create fortunes, or add to, or maintain their inheritance. The members of the second group are unable to organise factors of production as efficiently. Those born without an inheritance supply manual labour and those born with one squander it. Subsection 2.2. describes the model. 2.3. considers the effect of an inheritance tax and 2.4. the effect of a wage tax as an example of another tax. Subsection 2.5. compares the change in Rawlsian social welfare caused by equal yield taxes and 2.6. contains some concluding remarks.

3.2.2. The model

There are assumed to be two roles that people play in the productive process. The first is that of the entrepreneur who decides what to produce and how to produce it. The second is that of the worker who together with capital actually produces the output.

People of ability A_1 are taken to perform the entrepreneurial role well. They are good at organising factors of production and predicting future patterns of demand. Those of ability A_2 cannot do these tasks so well.

Output at time t , Y , is assumed to depend on the ability of the person who sets up and organises the firm and the amount of capital K and manual labour L used in production.

$$(1) \quad Y_i = Y(A_i, K, L) \quad i = 1, 2$$

$$(2) \quad Y_K, Y_L > 0 ; Y_{KK}, Y_{LL} < 0$$

where a subscript denotes a partial derivative.

These production functions are taken to display constant returns to scale in K and L.

$$(3) \quad Y_i = L y_i(k) \quad i = 1, 2$$

where $k = K / L$. The type of situation envisaged is shown in Figure 1.

There is an imperfect capital market such that it is impossible to borrow because, for example, of the possibility of default. Entrepreneurs thus supply all their own capital. For each firm K is determined by the savings decision of the entrepreneur who owns it.

The number of people employed in each business is chosen to maximise profits taking the wage w as given.

$$(4) \quad \max_L L y_i - w L \quad i = 1, 2$$

The first order condition for this gives

$$(5) \quad w = y_i - y_i' k \quad i = 1, 2$$

where y_i' denotes $\partial y_i / \partial K$. However, if y_2 is sufficiently small no solution to (5) for $i = 2$ may exist; output may

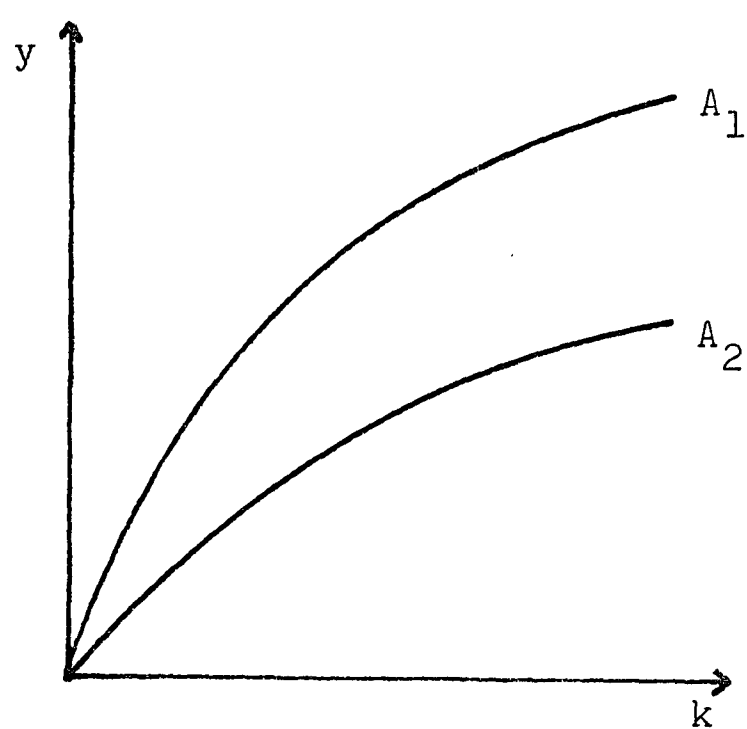


FIGURE 1

not even be enough to cover wages.

Since there are constant returns to scale it follows that the return on capital r_i is given by

$$(6) \quad r_i = y'_i \quad i = 1, 2$$

In addition to choosing productive projects there is taken to be another way of transferring goods from one period to another. This can be thought of as the holding of physical assets such as gold, or assets issued by the government such as money. Although it has been assumed that private borrowing is impossible it does not necessarily follow that government borrowing or the issuing of money is also infeasible. The return on this type of asset is d and its supply at this price is assumed to be perfectly elastic.

The theory of capital supply is very similar to that of Atkinson (1971a). People live T years. Their lifetime utility function is taken to be of the form

$$(7) \quad U = \int_0^T u(C_t, L_t) e^{-\rho t} dt + \psi((1 - \tau_B) B) e^{-\rho T}$$

where C_t is consumption, L_t is the amount of manual labour supplied, ρ is the rate of time discount and τ_B is the rate of tax on bequests B so that $(1 - \tau_B) B$ is the inheritance received by heirs. Entrepreneurial tasks are taken to be equivalent to leisure-time activities and so do not appear explicitly in the utility function. The properties of u and ψ are

$$(8) \quad u_C, \psi_B > 0 ; u_L < 0 ; u_{CC}, u_{LL}, \psi_{BB} < 0$$

The budget constraint is

$$(9) \quad \dot{K}_t = r_i K_t + w L_t - C_t$$

where K_t is the derivative with respect to time. Integrating this gives capital at time t

$$(10) \quad K_t e^{-r_i t} = I + \int_0^t (w L_a - C_a) e^{-r_i a} da$$

where I is the inheritance received.

The problem faced by people is therefore to choose C_t , L_t and $B (= K_T)$ to maximise their utility subject to (9). Choosing bequests gives the terminal condition

$$(11) \quad u_C(C_T, L_T) = \psi_B((1 - \tau_B) B) (1 - \tau_B)$$

The maximum principle can be used to find the timepath of consumption and labour supplied. The Hamiltonian for the problem is

$$(12) \quad H = u(C_t, L_t) e^{-\rho t} + \pi_t (w L_t + r_i K_t - C_t)$$

Choosing C_t and L_t to maximise this gives

$$(13) \quad u_C e^{-\rho t} = \pi_t$$

$$(14) \quad u_L e^{-\rho t} = -w \pi_t$$

Eliminating π_t gives the usual static condition for utility maximisation

$$(15) \quad -\frac{u_C}{u_L} = w$$

It follows from the maximum principle that

$$(16) \quad \dot{\pi}_t = -r_i \pi_t$$

Integrating

$$(17) \quad \pi_t = \pi_0 e^{-r_i t}$$

Substituting for π_t into (13) and (14) gives a pair of implicit functions in terms of C_t and L_t . Differentiating these with respect to time

$$(18) \quad u_{CC} \dot{C}_t + u_{CL} \dot{L}_t + (r_i - \rho) u_C = 0$$

$$(19) \quad u_{LC} \dot{C}_t + u_{LL} \dot{L}_t + (r_i - \rho) u_L = 0$$

Solving these simultaneously for $\dot{C}_t$ and $\dot{L}_t$

$$(20) \quad \dot{C}_t \left[1 - \frac{u_{LC}^2}{u_{CC} u_{LL}} \right] = (r_i - \rho) \left[\frac{C}{\eta_C} + \frac{u_{LC}}{u_{CC}} \frac{L}{\eta_L} \right]$$

$$(21) \quad \dot{L}_t \left[1 - \frac{u_{LC}^2}{u_{CC} u_{LL}} \right] = - (r_i - \rho) \left[\frac{C}{\eta_C} \frac{u_{LC}}{u_{LL}} + \frac{L}{\eta_L} \right]$$

where $\eta_C = - u_{CC} C / u_C$ and $\eta_L = u_{LL} L / u_L$.

From the static second order condition for a local maximum it follows that

$$(22) \quad \left[1 - \frac{u_{LC}^2}{u_{CC} u_{LL}} \right] > 0$$

It is assumed that

$$(23) \quad \left[\frac{C}{\eta_C} + \frac{u_{LC}}{u_{CC}} \frac{L}{\eta_L} \right], \quad \left[\frac{C}{\eta_C} \frac{u_{LC}}{u_{LL}} + \frac{L}{\eta_L} \right] > 0$$

The states that are considered involve the following assumptions concerning the magnitudes of the rates of return

$$(24) \quad r_2 < \rho < r_1$$

Also it is taken that

$$(25) \quad u_C(w L_2, L_2) > \psi_B(0) (1 - \tau_B)$$

where L_2 is given by

$$(26) \quad - \frac{u_C(w L_2, L_2)}{u_L(w L_2, L_2)} = w$$

It follows from (24) that A_1 's increase their consumption through their lifetimes, and their supply of manual labour falls. If r_1 or I is sufficiently high there will come a

point where consumption is such that

$$(27) \quad - \frac{u_C(C, 0)}{u_L(C, 0)} \geq w$$

and they cease to work for a wage and live off their earnings from capital.

Equations (24) and (25) together imply that A_2 's born without an inheritance will consume all their wages every period and acquire no assets. If ψ is defined for negative B they would like to die in debt but the imperfect capital market prevents this. A_2 's born with an inheritance decrease their consumption through their lifetime. They live lavishly while young but as they grow older they cut back. If at any time C falls below the value defined by (27) with an equality they may even start going out to work and as they grow older the amount they work will steadily increase.

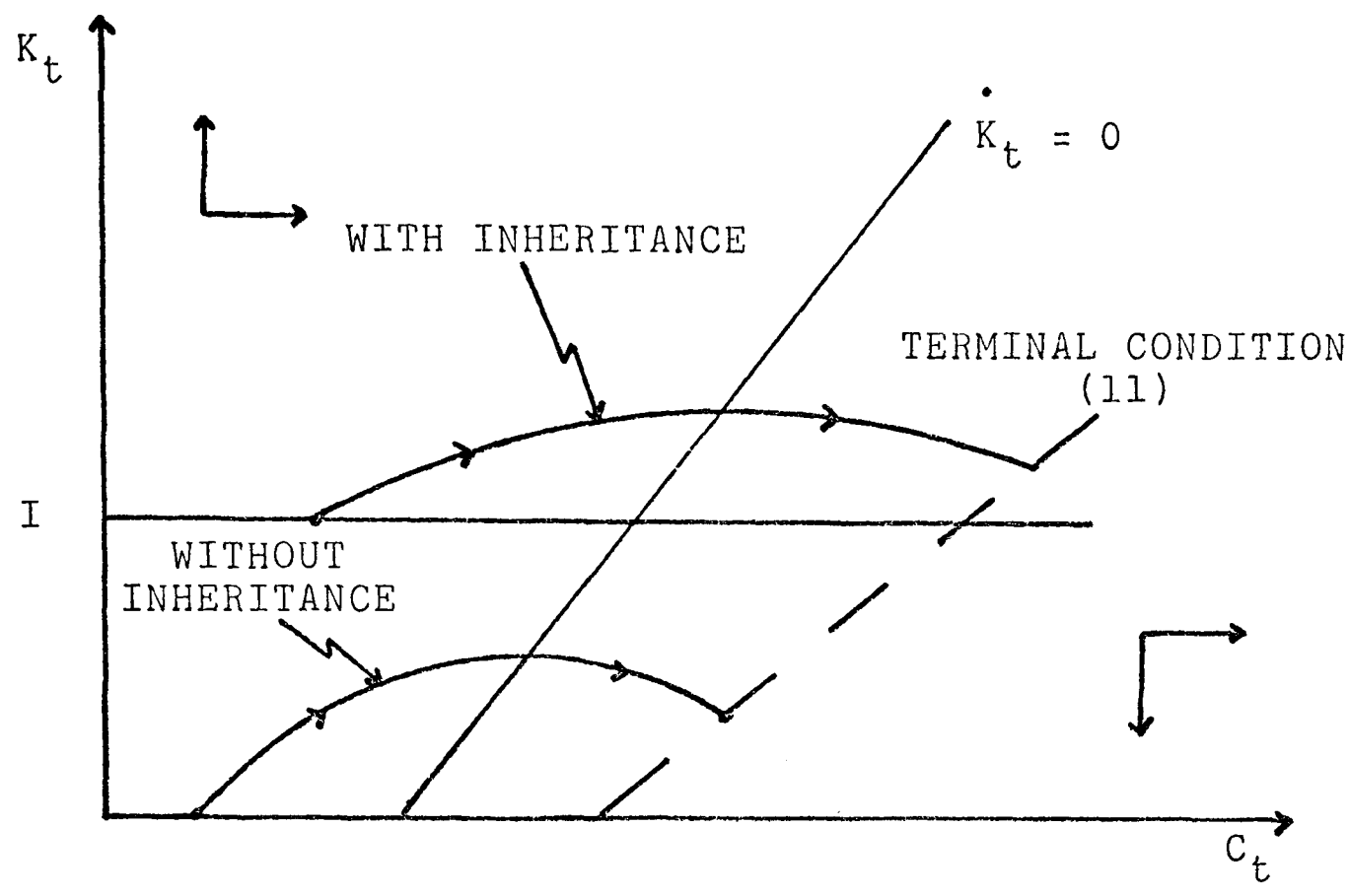
Figures 2(a) and (b) illustrate these possibilities for A_1 's and A_2 's respectively leaving aside the change in L_t .

The capital of a person at time t can be expressed as a function of his inheritance, the rate of return he can earn, the wage and the tax rate on his bequest. His average lifetime holding of capital is thus given by

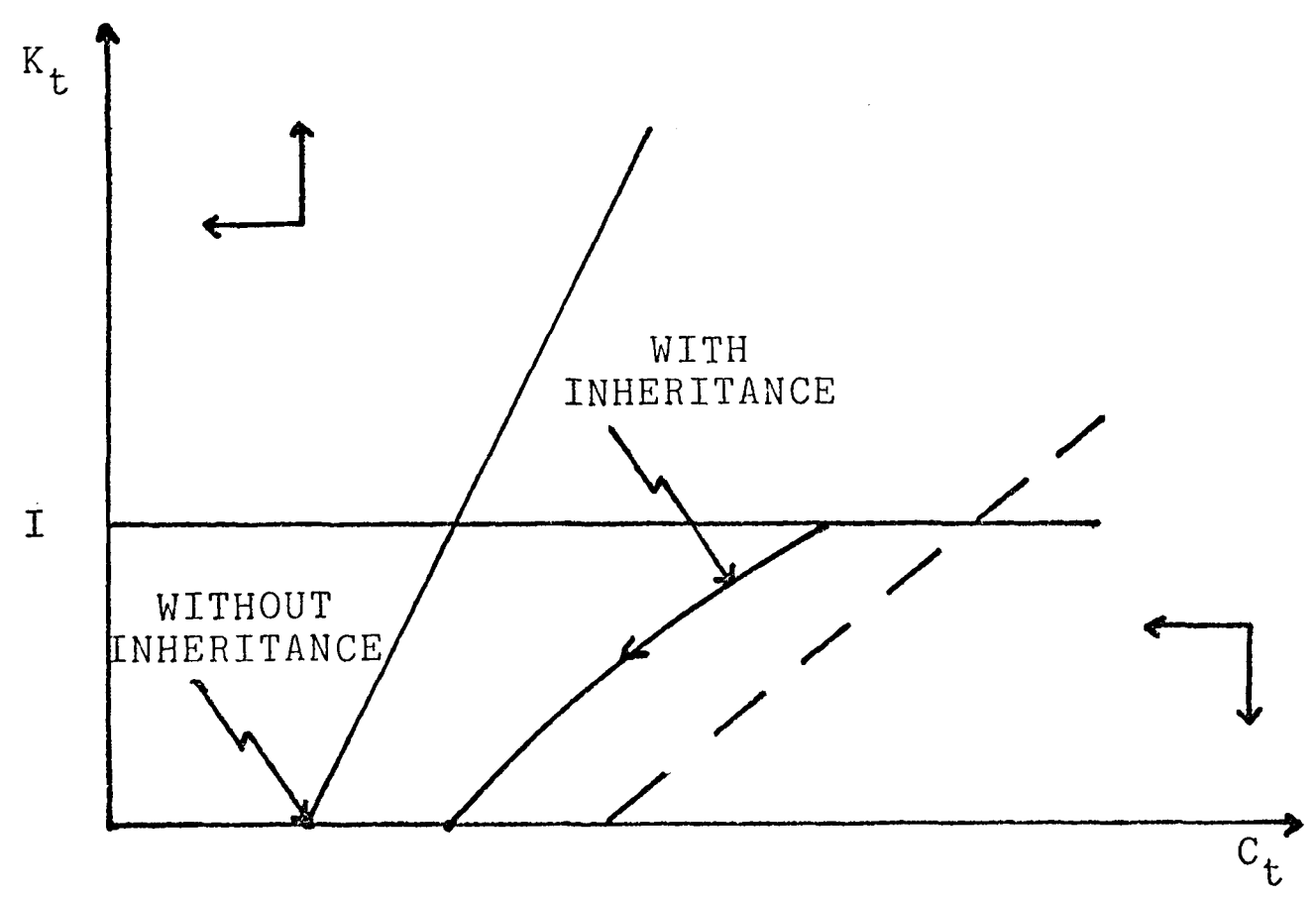
$$(28) \quad K = \int_0^T \frac{1}{T} K_t dt = K(I, r_i, w, \tau_B)$$

The bequest a person leaves can similarly be expressed as a function of these variables

$$(29) \quad B = B(I, r_i, w, \tau_B)$$



(a) A_1



(b) A_2

FIGURE 2

The utility of an individual can be expressed indirectly as a function of his inheritance, the interest rate he can earn, the wage and the tax on bequests

$$(30) \quad V = V(I, r_i, w, \tau_B)$$

(24) and (25) imply that the indirect utility of a property-less worker V^* can be expressed as a function of the wage alone.

$$(31) \quad V^* = V^*(w)$$

The relationship between the ability of parents and children is assumed to be Markovian. If α_1 is the proportion of A_1 's in the population, α_{11} is the probability of an A_1 parent having an A_1 child and α_{21} is the probability of A_2 parents having A_1 children, then in the steady state

$$(32) \quad \begin{bmatrix} \alpha_1 \\ 1 - \alpha_1 \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{21} \\ 1 - \alpha_{11} & 1 - \alpha_{21} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ 1 - \alpha_1 \end{bmatrix}$$

Hence

$$(33) \quad \alpha_1 = \frac{\alpha_{21}}{1 + \alpha_{21} - \alpha_{11}}$$

For simplicity the population is assumed to be constant through time and children are born as parents die. There is therefore a uniform distribution over the life cycle and the inheritance of each generation of a family is the bequest

of the one before. For the g^{th} generation

$$(34) \quad I_g = (1 - \tau_B) B_{g-1}$$

The population is normalised at 1.

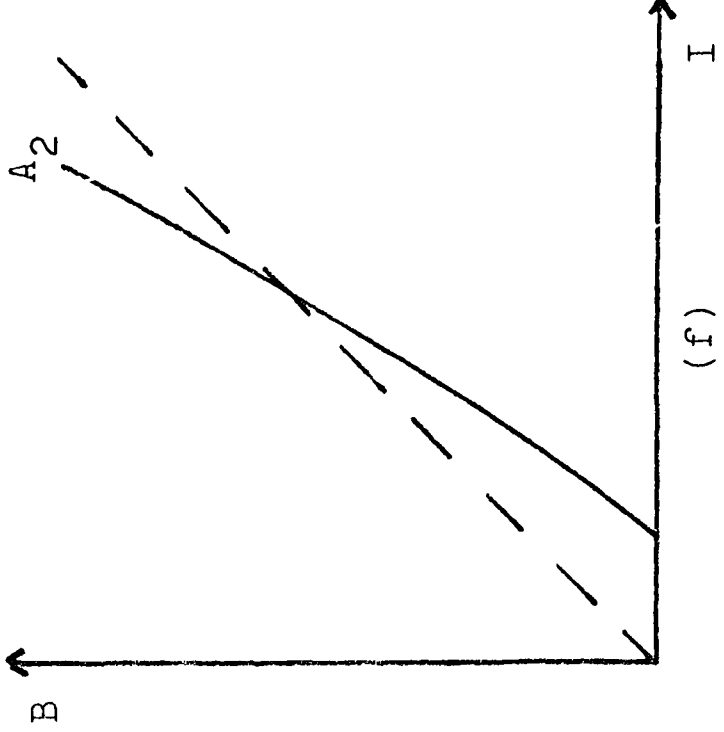
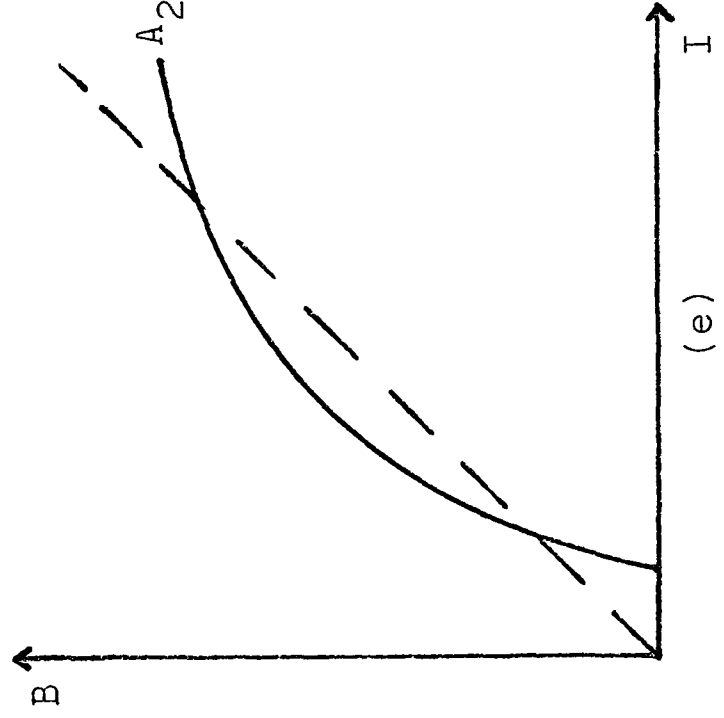
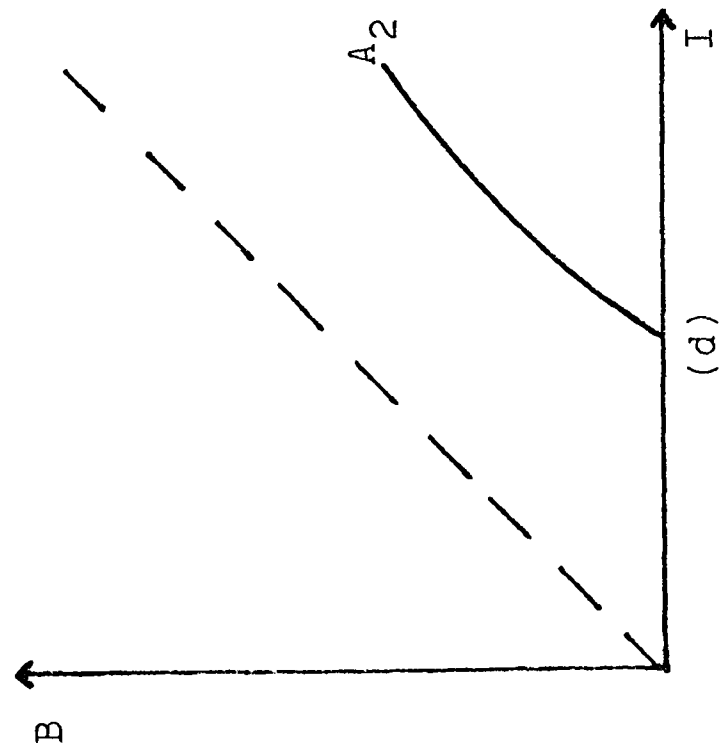
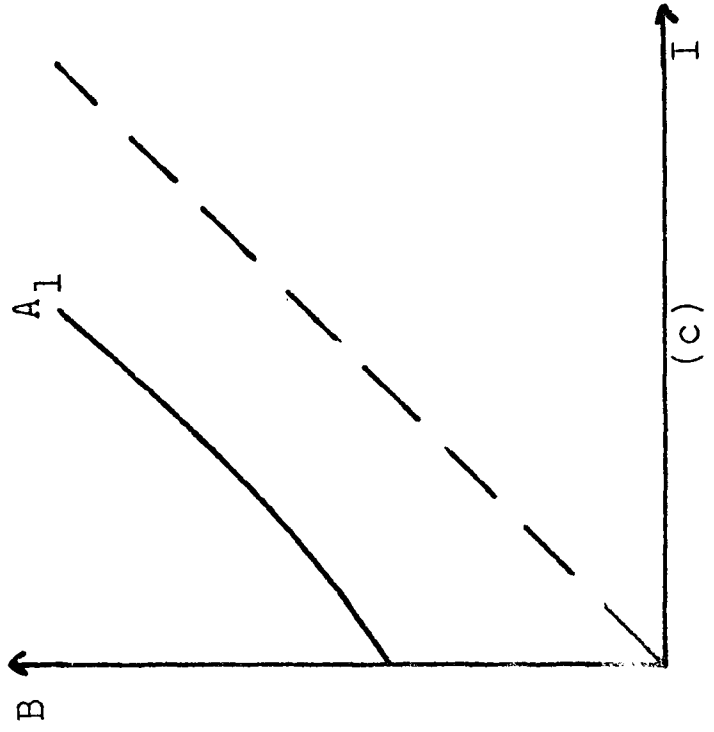
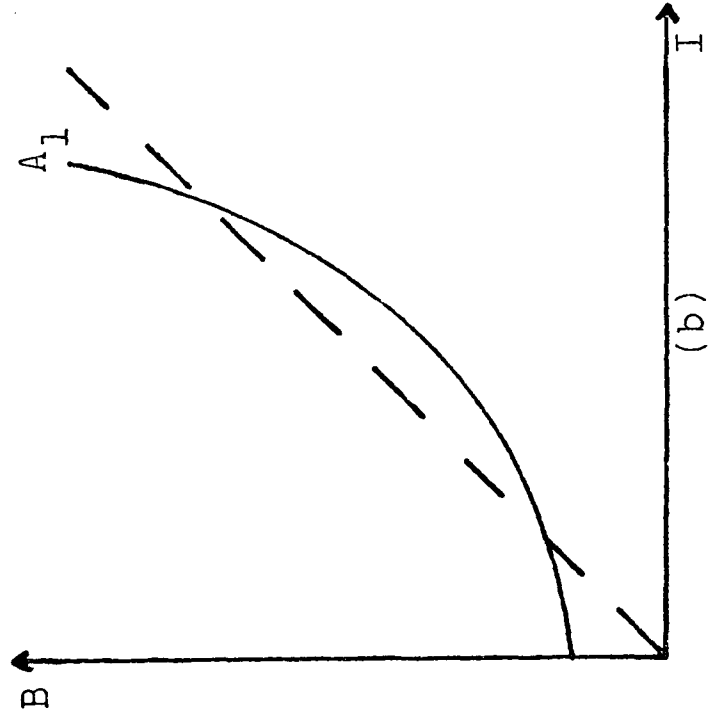
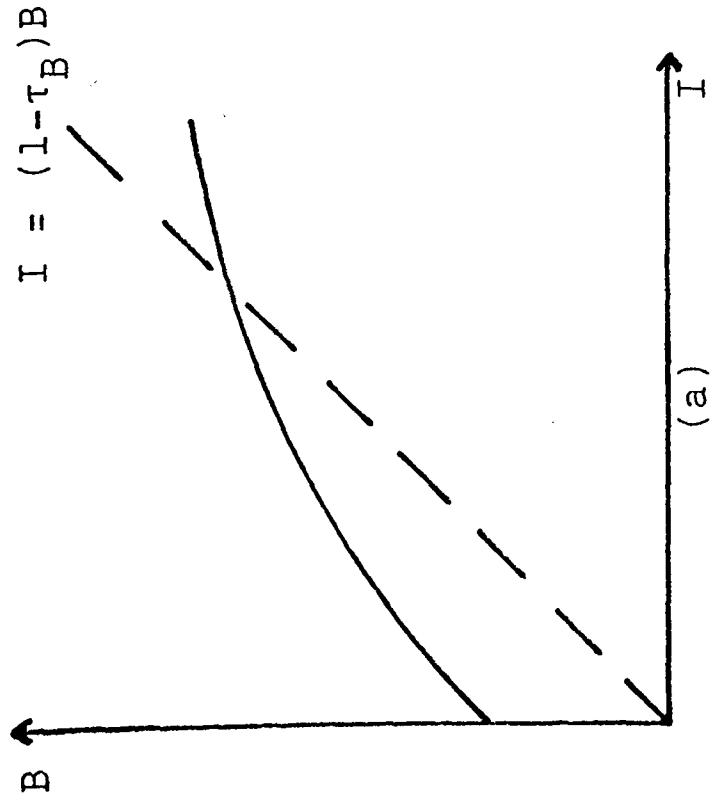
The development of wealthholding over generations depends on the outcome of the Markovian process and the form of the bequest function. It can be seen from the analysis of consumers' behaviour above that an increase in the inheritance of a person results in a strict increase in his bequest provided this is strictly positive, so that $\partial B / \partial I > 0$ for $B > 0$. If inflection points are disregarded, then for each ability group there are therefore three possibilities for the shape of $B(I)$ as shown in Figure 3.

It follows that there are nine types of situation that can arise in this case. Given the same inheritance A_1 's must leave a larger bequest than A_2 's since they have the same utility function and greater factor payments. The curves cannot then cross and it is possible to rule out the combination of (a) and (f) so that there are in fact only eight possibilities.

As an example of the type of steady state that arises consider the combination of (a) and (d) as shown in Figure 4.

Consider a group of A_1 's who are born without an inheritance. They build up a fortune and leave a bequest B_1 . α_{11} of their heirs are also A_1 and they build up the family fortunes and leave a bequest B_2 . The other $(1 - \alpha_{11})$ of the heirs are A_2 ; these people dissipate their inheritance

FIGURE 3



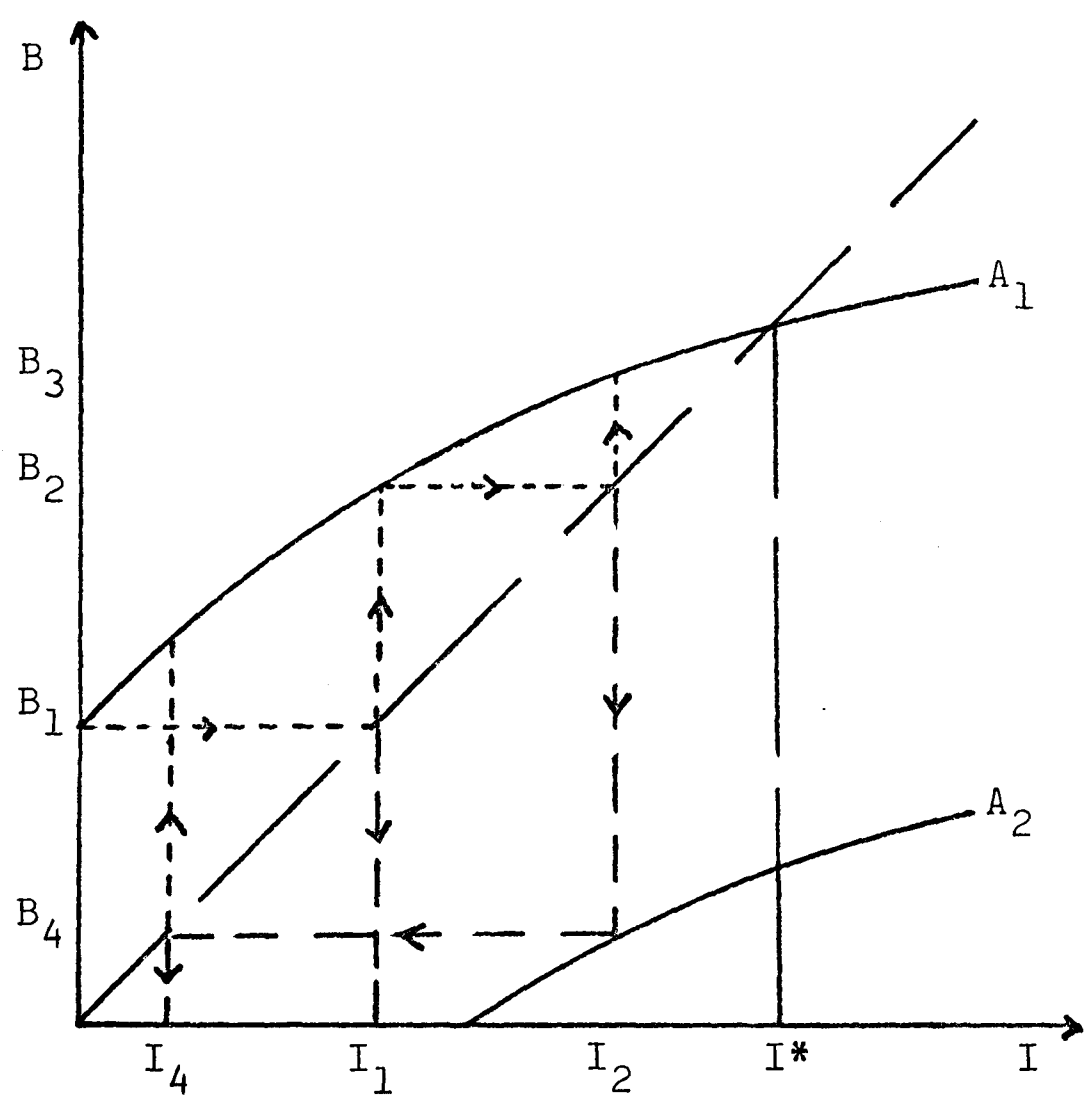


FIGURE 4

and their children are born propertyless. Of those who receive B_2 , α_{11} again expand the fortune and leave B_3 . The A_2 heirs dissipate their inheritance but still leave a bequest B_4 . Of those with B_4 , α_{21} are A_1 's and expand their inheritances, $(1 - \alpha_{21})$ are A_2 's and dissipate theirs. The process continues in a similar way.

Some families expand to I^* but none go above this. Eventually they all have an A_2 heir and the family fortune is diminished. With a long enough run of A_2 's it is entirely dissipated but large fortunes may be able to survive a limited number of A_2 heirs. This process leads to a steady state distribution of wealth.

The positions of the curves depend on factor prices. If these are fixed at their steady state values then any families which start out with wealth above I^* will in this case tend to I^* even if they consist of a string of A_1 's.

Other combinations of the curves can be analysed in the same way. (b) and (d) is similar to the case above, the difference being that any family that starts out with wealth above the point where the bequest-inheritance curve and the $I = (1 - \tau_B) B$ line cross for the second time will expand its fortune as long as it continues to have A_1 heirs. However, the occurrence of A_2 heirs will eventually cause the family to fall back.

With (c) and (d) there is no upper bound to the size of the fortune no matter where the family starts off. However, the number that attain more than a certain amount becomes very small because of the occurrence of A_2 heirs. Provided the A_1 curve is not too convex the large wealth-holders own only a small proportion of the total capital in

the economy and the process will again tend to a steady state.

There are no propertyless workers in the steady state with (a) and (e); everybody has capital. Fortunes oscillate between the two points where the A_1 and A_2 curves cut the $I = (1 - \tau_B) B$ line. Similarly for (b) and (e).

With (c) and (e) there are again no propertyless workers in the steady state but in this case there is no upper bound to larger fortunes.

The analysis of (b) and (f) and (c) and (f) is like that of (b) and (d). However, if families gain enough wealth they will continuously expand their fortunes whether heirs are A_1 or A_2 .

If the possibility of inflection points is included then the number of conceivable situations is greatly increased since each bequest curve may cross the $I = (1 - \tau_B) B$ line more than twice. However, the analysis of such cases is similar.

In this general form the rigorous analysis of the effect of taxes is rather complicated because of the problem of identifying the distribution of wealth without assuming more specific functional forms. For this reason a number of simplifying assumptions are made which greatly ease the analysis while at the same time retaining the essential features of the model.

The first of these is that

$$(35) \quad r_2 < d < \rho$$

so that A_2 's who inherit do not become entrepreneurs at all. They hold their wealth in the form of storable commodities or government assets rather than productive capital. This means that as far as the determination of factor prices is concerned only the capital and production function of A_1 's need be considered.

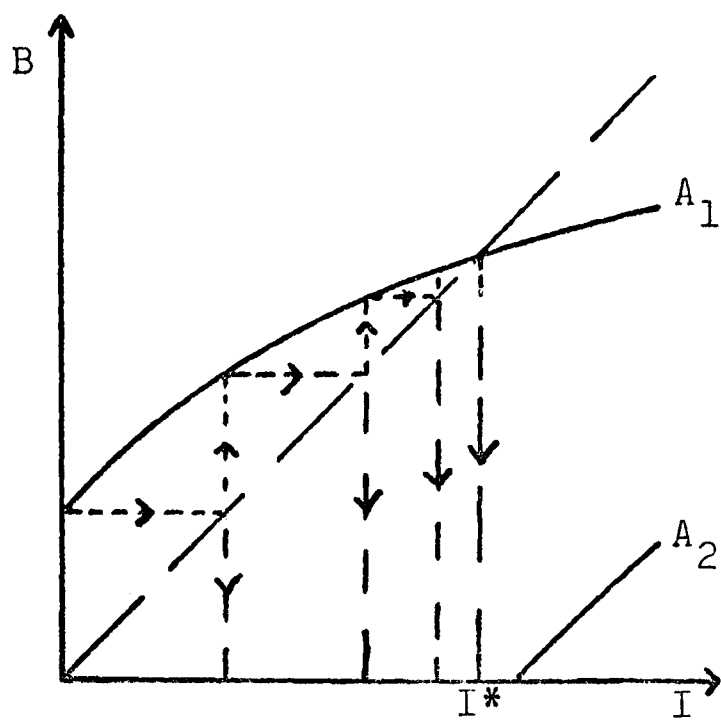
The second assumption is that d is sufficiently below ρ that the A_2 bequest-inheritance curve coincides with the axis for all the relevant levels of A_1 bequests in the steady state as shown in Figure 5. In (a) and (b) all fortunes received by A_2 heirs are squandered in a generation. In (c) it is assumed that the number of fortunes above I^* and their total as a proportion of the national wealth are negligible. Except very occasionally all fortunes are dissipated in one generation; those which are not are taken to be sufficiently rare to be ignored.

Most manual labour is taken to be supplied by property-less workers. Aggregate labour supply thus depends only on the wage.

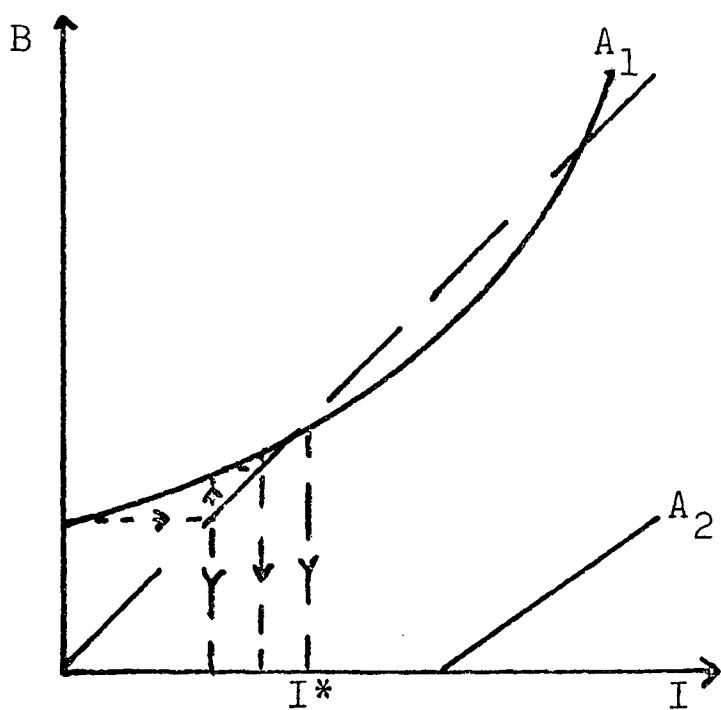
$$(36) \quad L = L(w)$$

Also the time entrepreneurs spend having to work for wages when they originally found their fortunes is taken to be negligible so that w can be excluded from the life-time capital supply, bequest and indirect utility functions of A_1 's.

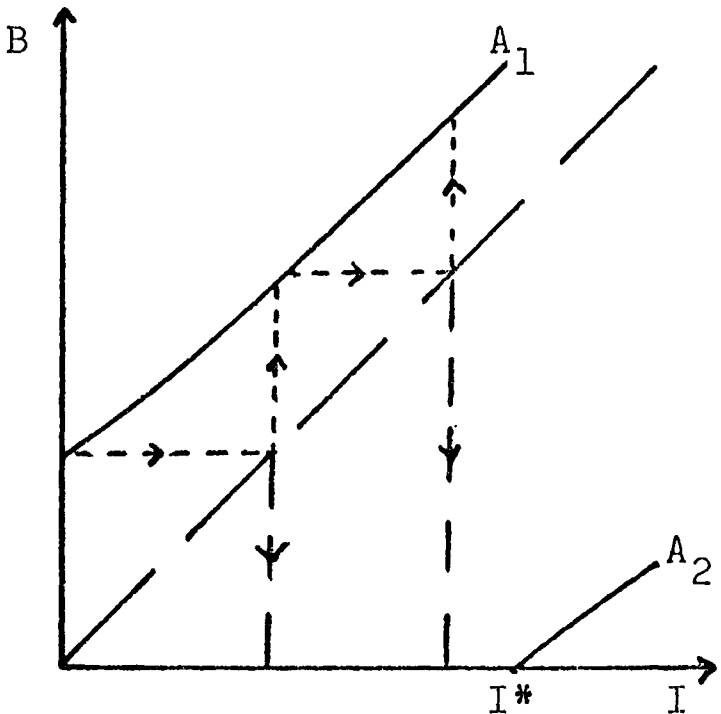
$$(37) \quad K^1 = K^1(I, r_1, \tau_B)$$



(a)



(b)



(c)

FIGURE 5

$$(38) \quad B^1 = B^1(I, r_1, \tau_B)$$

If A_1 's of the current generation whose parents were A_2 are said to be A_1 's of generation 0, those whose parents were A_1 but grandparents were A_2 to be of generation 1 and so on, then the proportion p_g of the population who are alive at any time who are A_1 's of generation g is given by

$$(39) \quad p_g = \alpha_{11}^g \alpha^*$$

where $\alpha^* = \alpha_{21}(1 - \alpha_1)$.

The number and proportion of wealth held by A_1 's above generation n is taken to be negligible. Hence in the steady state the aggregate capital supply is given by

$$(40) \quad K = \sum_{g=0}^n p_g K_g^1$$

where $K_g^1 = K^1((1 - \tau_B) B_{g-1}^1, r_1, \tau_B)$

with

$$(41) \quad B_g^1 = B^1((1 - \tau_B) B_{g-1}^1, r_1, \tau_B)$$

and $B_{-1}^1 = 0$

Since there are constant returns to scale the determination of factor prices can be represented as if there were a single firm with a production function given by

(3) ($i = 1$), a capital supply given by (40) and an aggregate labour supply given by (36).

This is the version of the model that is used to consider the effects of taxation. For ease of notation the subscripts and superscripts referring to A_1 's will be omitted in the next subsections.

3.2.3. Inheritance taxation

In this subsection the effect of the introduction of a small tax on inheritance, or equivalently in this model a capital transfer tax, is considered.

Differentiating (5) and (6) with respect to τ_B gives

$$(42) \quad \frac{d \log w}{d \tau_B} = \xi_w \frac{d \log k}{d \tau_B}$$

$$(43) \quad \frac{d \log r}{d \tau_B} = - \xi_r \frac{d \log k}{d \tau_B}$$

where

$$(44) \quad \xi_w = \frac{k}{w} \frac{\partial w}{\partial k} > 0$$

$$(45) \quad \xi_r = - \frac{k}{r} \frac{\partial r}{\partial k} > 0$$

the elasticities of the marginal products of labour and capital with respect to the capital labour ratio.

From the definition of k

$$(46) \quad \frac{d \log k}{d \tau_B} = \frac{d \log K}{d \tau_B} - \frac{d \log L}{d \tau_B}$$

From (36)

$$(47) \quad \frac{d \log L}{d \tau_B} = \epsilon \frac{d \log w}{d \tau_B}$$

where $\epsilon = \frac{w}{L} \frac{\partial L}{\partial w}$ the elasticity of labour supply.

Differentiating (40) and evaluating at $\tau_B = 0$

$$(48) \quad \frac{d \log K}{d \tau_B} = \frac{1}{K} \sum_{g=0}^n p_g K_g \left[\kappa_{Ig} \frac{d \log B_{g-1}}{d \tau_B} + \kappa_{rg} \frac{d \log r}{d \tau_B} + \kappa_{\tau g} - \kappa_{Ig} \right]$$

where

$$(49) \quad \kappa_{Ig} = \frac{I}{K_g} \frac{\partial K_g}{\partial I} ; \kappa_{rg} = \frac{r}{K_g} \frac{\partial K_g}{\partial r} ; \kappa_{\tau g} = \frac{1}{K_g} \frac{\partial K_g}{\partial \tau_B}$$

Differentiating (41)

$$(50) \quad \frac{d \log B_g}{d \tau_B} = \beta_{Ig} \frac{d \log B_{g-1}}{d \tau_B} + \beta_{rg} \frac{d \log r}{d \tau_B} + \beta_{\tau g} - \beta_{Ig}$$

where

$$(51) \quad \beta_{Ig} = \frac{I}{B_g} \frac{\partial B_g}{\partial I} ; \beta_{rg} = \frac{r}{B_g} \frac{\partial B_g}{\partial r} ; \beta_{\tau g} = \frac{1}{B_g} \frac{\partial B_g}{\partial \tau_B}$$

(50) is a difference equation of the form

$$(52) \quad x_{t+1} = a_{t+1} x_t + c_{t+1}$$

which has the solution

$$(53) \quad x_t = \prod_{j=1}^t a_j x_0 + \sum_{h=1}^{t-1} \prod_{j=h+1}^t a_j c_h + c_t$$

This gives a rather complex expression for $d \log B_{g-1} / d\tau_B$ and more insight is perhaps gained by considering the constant elasticity case where β_I , β_r and β_τ are independent of g . The form of (50) then simplifies to

$$(54) \quad x_{t+1} = a x_t + c$$

which taking $a \neq 1$ has the solution

$$(55) \quad x_t = (x_0 - \frac{c}{1-a}) a^t + \frac{c}{1-a}$$

Taking $\beta_I \neq 1$, substituting the terms from (50) into (55) with

$$(56) \quad x_0 = \frac{d \log B_0}{d\tau_B} = \beta_r \frac{d \log r}{d\tau_B} + \beta_\tau$$

and rearranging

$$(57) \quad \frac{d \log B_{g-1}}{d\tau_B} = Q_g \beta_r \frac{d \log r}{d\tau_B} + Q_g (\beta_\tau - 1) + 1$$

where

$$(58) \quad Q_g = \frac{(1 - \beta_I^g)}{(1 - \beta_I)} = \sum_{j=0}^g \beta_I^j$$

Substituting from (57) into (48) and using (42), (43), (46) and (47) it is possible to solve simultaneously to find the changes in w , r , k , K and L in terms of the elasticities. Solving these for $d \log k / d\tau_B$ gives

$$(59) \quad \frac{d \log k}{d\tau_B} = \frac{\sum p_g \frac{K_g}{K_g} \frac{D_g}{G_g}}{\sum p_g \frac{K_g}{K_g} \frac{D_g}{G_g}}$$

where

$$(60) \quad D_g = \kappa_{\tau g} + \kappa_{Ig} Q_g (\beta_{\tau} - 1)$$

$$(61) \quad G_g = 1 + \epsilon \xi_w + \kappa_{rg} \xi_r + \kappa_{Ig} Q_g \beta_r \xi_r$$

It can be seen from (42), (43), (46) and (47) that once the sign of $d \log k / d\tau_B$ has been found it is a fairly simple matter to determine the direction of the change in the other variables.

Atkinson (1971a) showed that for a utility function similar to (7) but with L held constant and with slightly different assumptions concerning ψ that certain restrictions can be placed on the sign and magnitude of the elasticities. Since it was assumed that the 0th generation entrepreneurs only supply a negligible amount of manual labour at the very

beginning of their careers these results are applicable here. They are summarised in Table 1.

It follows from the concavity of the production function that $\xi_w, \xi_r > 0$. Also with constant returns to scale it can be shown

$$(62) \quad \xi_r = \frac{w L / Y}{\sigma}$$

$$(63) \quad \xi_w = \frac{r K / Y}{\sigma}$$

where σ is the elasticity of substitution of the production function.

Since the labour is assumed to be predominantly supplied by propertyless workers who have identical preferences, the aggregate labour supply elasticity ϵ is the same as the individual one. It is well known that this can be of either sign. Also if the elasticity of substitution between consumption and leisure is positive then $\epsilon > -1$.

The terms in the expression for $d \log k / d\tau_B$ in (59) can be intuitively explained in the following way. The numerator represents the direct effect of the tax. It is the sum of n terms because the overall change is the combination of the effects on each generation of A_1 's. $p_g K_g$ is the amount of capital owned by the g^{th} generation and the term D_g is the change in this caused by the direct effect of the tax.

People are interested in the net bequest received by their heirs. A change in the tax rate on transfers alters this amount. β_τ represents the change made in their bequests

TABLE 1

WITH RESPECT TO ELASTICITY				
	I	r	τ	
	β	κ		
	$\begin{matrix} > 0 \\ < 1 \text{ as } \mu > \eta_C \end{matrix}$	$\begin{matrix} > 0 \\ < 0 \end{matrix}$	$\begin{matrix} < 1 \\ > 0 \text{ as } \mu > 1 \\ < 0 \text{ as } \mu < 1 \end{matrix}$	
			$\begin{matrix} > 0 \\ < 0 \end{matrix}$	

* $\mu = - \psi_{BB} \text{ B} / \psi_B$, the elasticity of the marginal utility of bequests.
 $\neq$ Constant elasticity case

to offset this; it can be positive or negative. Either they leave more to compensate for the effect of the tax or they leave less to avoid paying it: which depends on whether the elasticity of marginal utility of bequests $\mu > 1$. It can be shown that the net effect on bequests $\beta_\tau - 1$ is always negative. $Q_g (\beta_\tau - 1)$ is the cumulative effect of the tax on the bequests of previous generations; since $Q_g > 1$ for $g > 0$ it is negative. $\kappa_{Ig} Q_g (\beta_\tau - 1)$ is therefore the direct effect of the tax on average lifetime capital supply caused by the cumulative change in the net bequests of the previous generations and is always negative. $\kappa_{\tau g}$ is the alteration in capital supply associated with the change in bequests; it has the same sign as β_τ and can be either positive or negative.

The direct effect changes relative factor supplies which alter factor prices which again in turn change factor supplies. The term G_g in the denominator represents this indirect general equilibrium effect on each generation. $\epsilon \xi_w$ is the change in labour supply which results from the change in wages and can be of either sign. Similarly $\kappa_{rg} \xi_r$ is the change in capital supply resulting from the change in the interest rate and it can also be of either sign. Bequests depend on the interest rate; $\kappa_{Ig} Q_g \beta_r \xi_r$ represents the effect on the capital supply due to the cumulative change in bequests resulting from the variation in the interest rate. It is positive.

It follows that $d \log k / d\tau_B$ can be of either sign in plausible situations as the following example demonstrates.

Example 1

- (a) If for all g , $\kappa_{\tau g} < 0$, $\kappa_{rg} > 0$ and $\epsilon > 0$ then
 $d \log k / d\tau_B < 0$.
- (b) If for all g , $\kappa_{\tau g} < 0$, $\kappa_{rg} < 0$ and $\epsilon < 0$, and the magnitudes of the terms are such that $G_g < 0$ then
 $d \log k / d\tau_B > 0$.

The demonstration of this example follows directly from the conditions given, those in Table 1, (44), (45) and (59). All that (a) requires is that an inheritance tax reduces bequests, and the capital and labour supply elasticities be positive. (b) differs in that the factor supply elasticities are negative and there are requirements on the magnitudes of the terms in G_g . The elasticity of substitution of the production function σ should be low and the income share of the capital high so that ξ_w and ξ_r are large with the former being the greater of the two. The cumulative alteration in the capital supply due to the effect of the change in interest rates on bequests, $\kappa_{Ig} Q_g \beta_r \xi_r$, should also be small. If these requirements are satisfied then $G_g < 0$. Neither of these cases would appear to be unreasonable.

Using (39) it can be seen that (59) simplifies to

$$(64) \quad \frac{d \log k}{d\tau_B} = \frac{\sum \alpha_{ll}^g K_g D_g}{\sum \alpha_{ll}^g K_g G_g}$$

The only demographic factor that enters the expression is

α_{11} so that it is the likelihood that an A_1 parent has an A_1 child that is important. This is perhaps not surprising given the assumption that only A_1 's use their capital for production and A_2 's do not leave an inheritance. However, it can be seen that the effect of a change in α_{11} on the influence of an inheritance tax on the capital labour ratio and other terms is again ambiguous. A high correlation between the ability of parents and children does not therefore necessarily provide an argument against inheritance taxation. Finally it also follows from (64) that if the direct effect D_g and general equilibrium effect G_g are the same for each generation then demographic factors play no role at all in determining $d \log k / d\tau_B$.

As is perhaps to be expected in a general equilibrium model it is not possible to say whether the introduction of a small inheritance tax will cause a rise or fall in the capital labour ratio and hence how the other quantities, prices and utilities of interest will be affected. The importance of the formulae of the type given in (59) is rather that they give some indication of the situations in which there will be movements in a particular direction.

3.2.4. Wage taxation

The model can be used to analyse the effect of a number of other taxes. As an example, the case of a proportional wage tax τ_w is considered.

The expression for aggregate labour supply (36) is changed to

$$(36') \quad L = L((1 - \tau_w) w)$$

Differentiating with respect to τ_w

$$(47') \quad \frac{d \log L}{d \tau_w} = \epsilon \left[\frac{d \log w}{d \tau_w} - 1 \right]$$

Holding τ_B constant and evaluating at τ_B , $\tau_w = 0$ (48) becomes

$$(48') \quad \frac{d \log K}{d \tau_w} = \frac{1}{K} \sum_{g=0}^n p_g K_g \left[\kappa_{Ig} \frac{d \log B_{g-1}}{d \tau_w} + \kappa_{rg} \frac{d \log r}{d \tau_w} \right]$$

The effect on bequests is given by

$$(50') \quad \frac{d \log B_g}{d \tau_w} = \beta_{Ig} \frac{d \log B_{g-1}}{d \tau_w} + \beta_{rg} \frac{d \log r}{d \tau_w}$$

This can be solved in a similar way to (50). Thus in the constant elasticity case with $\beta_I \neq 1$, this has the solution

$$(57') \quad \frac{d \log B_{g-1}}{d \tau_w} = Q_g \beta_r \frac{d \log r}{d \tau_w}$$

The expressions for changes in w , r and k are as in (42), (43) and (46) with τ_w replacing τ_B . Solving these (47'), (48') and (57') simultaneously as before and using (39) gives

$$(65) \quad \frac{d \log k}{d \tau_w} = \frac{\epsilon \sum \alpha_{11}^g K_g}{\sum \alpha_{11}^g K_g G_g}$$

The expression is very similar to that for $d \log k / d \tau_B$. The general equilibrium indirect effect which is represented by the denominator is the same. The numerator is the direct

effect of the tax on labour supply. As before a wage tax can effect the capital labour ratio and hence the other variables of interest in either direction, as the following example demonstrates.

Example 2

- (a) If for all g , $\kappa_{rg} > 0$ and $\epsilon > 0$ then $d \log k / d\tau_w > 0$.
- (b) If for all g , $\kappa_{rg} > 0$ and $\epsilon < 0$ and magnitudes are such that $G_g > 0$ then $d \log k / d\tau_w < 0$.

The demonstration of this again follows directly from the conditions and (65). The only requirement for (a) is that both factor supply elasticities are positive. The difference in (b) is that the labour supply elasticity is negative and magnitudes must be such that $G_g > 0$. Since $\epsilon > -1$, a sufficient condition for the latter to hold is that $\sigma > 1$. Thus both (a) and (b) are plausible.

The effect of other taxes such as a tax on income from capital or on consumption can be found in a similar way.

3.2.5. A Rawlsian comparison

One of the purposes of considering incidence is to be able to compare the effects of various taxes on a given objective. As an example for this model this section compares the effect on Rawlsian social welfare of an inheritance tax with that of a wage tax which raises the same revenue.

Since propertyless workers are the worst-off people in the model Rawlsian social welfare S is given by

$$(66) \quad S = V^*((1 - \tau_w) w)$$

An equal yield comparison at $\tau_w, \tau_B = 0$ implies that

$$(67) \quad w L d\tau_w = \frac{1}{T} \sum_{g=0}^n p_g B_g d\tau_B$$

or

$$(68) \quad \frac{d\tau_w}{d\tau_B} = \frac{F}{w L}$$

where F is the flow of bequests.

Hence whether a wage tax or an inheritance tax is preferable as far as the workers are concerned depends on whether

$$(69) \quad Z = \frac{dS(\tau_w)}{d\tau_w} \frac{d\tau_w}{d\tau_B} - \frac{dS(\tau_B)}{d\tau_B} > 0$$

Evaluating this using (42), (59), (66) and (68) gives

$$(70) \quad Z = - \frac{1}{L} \frac{\partial V^*}{\partial w} \frac{\sum \alpha_{11}^g K_g N_g}{\sum \alpha_{11}^g K_g G_g}$$

where

$$(71) \quad N_g = F(1 + \kappa_{rg} \xi_r + \kappa_{Ig} Q_g \beta_r \xi_r) + w L \xi_w D_g$$

Once again both signs are possible for Z as the following example demonstrates.

Example 3

- (a) If for all g , $\kappa_{tg} < 0$, $\kappa_{rg} > 0$, $\epsilon > 0$ and magnitudes are such that $N_g < 0$, then $Z > 0$ and a wage tax is the preferred means of raising revenue.
- (b) If the conditions are as in (a) except $\kappa_{rg} < 0$, $\epsilon < 0$ and magnitudes are such that $G_g < 0$, then $Z < 0$ and an inheritance tax is preferable.

As before the demonstration of this follows directly from the conditions and (70). Empirical work suggests that plausible values for the average bequest are of the same order of magnitude as annual earnings (see, for example, Farrell (1970)). Since only a small proportion of the population dies each year this implies that wL is much larger than F . Hence if κ_{tg} is negative it is likely that N_g will also be. If factor supplies are positive as in (a) then $G_g > 0$ and $Z > 0$ so that a wage tax is a preferable means of raising revenue on the basis of Rawlsian social welfare. The difference in (b) is that factor supplies are negative and σ is small so that $G_g < 0$ and $Z < 0$; an inheritance tax is then preferred.

These are just two of many possibilities. However, they do illustrate the importance of careful analysis. Although workers' only source of income is wages and they leave no inheritance, a wage tax may well be preferable to an inheritance tax as far as they are concerned and hence on the grounds of Rawlsian social welfare; this can be so even if much of the inherited capital is squandered. Direct incidence is thus a very unreliable guide to the effects and

desirability of a tax. In case (a) for example the wage tax is more desirable because of the general equilibrium effects: with an inheritance tax these reduce the capital labour ratio and hence the wage whereas with a wage tax they increase the latter and this helps to offset the reduction in net income due to the direct effect.

3.2.6. Concluding remarks

The analysis above has developed a simple general equilibrium model of inheritance in the context of an imperfect capital market. This has illustrated the way in which inequality of wealth arises from the combination of differences in ability in organising factors of production and an imperfect capital market.

The effect of inheritance taxation was also considered and this was compared to a wage tax. It was shown that even in this simplified model arguments of the type given in the introduction are too simplistic; the determination of the effects and merits of taxes is a complex issue. The importance of the results is that they give some indication of the channels through which the taxes operate, their relative importance and the roles of the various elasticities.

CHAPTER 4

OPTIMAL LINEAR INCOME TAXATION WITH GENERAL
EQUILIBRIUM EFFECTS ON WAGES4.1. Introduction and a survey of theories of income
taxation

One of the most important consequences of differences in ability is inequality of income; this chapter is concerned with the use of taxation for the redistribution of income.

There are two main approaches to problems of taxation. The first concerns the incidence or effects of the tax on the variables of interest. These can be summarised by considering the change in an appropriate measure of social welfare. This was the method used in the previous chapter. The second takes this one stage further and considers the taxes which would be required in order to maximise the measure of social welfare.

This type of exercise can be justified on a number of grounds. Philosophies such as utilitarianism or Rawls' (1971) theory of justice may represent either the whole or elements of individuals' attitudes towards distribution; the method can be useful in assessing such ethical foundations and help point to inconsistencies in attitudes.

The method may also be valuable in designing policies since it makes the ethical foundations of tax schedules explicit and separates them from the operation of the model of the economy. Even in relatively simple examples it is very difficult to judge the relationship between ethics and the final schedule. Mirrlees (1971), for example, wrote 'I

must confess that I had expected the rigorous analysis of income-taxation in the utilitarian manner to provide an argument for high tax rates. It has not done so.' (p. 207). As models become more complex this problem is greatly exacerbated. It is argued in the next section that because of general equilibrium effects the optimal schedules implied by Rawlsian and utilitarian social welfare functions may involve redistribution from poor to rich through the tax system. The use of an approach where social welfare is explicitly specified highlights the important point that such a result arises from positive rather than normative assumptions.

Much of the work concerned with income taxation has adopted this approach of considering the tax schedules that would result from various ethical positions. From the early literature, Edgeworth (1897) identified three main principles of taxation. These are equal sacrifice, proportional sacrifice and minimum sacrifice. Equal sacrifice requires that tax schedules to raise the government revenue requirement should be such that everybody's utility is reduced by the same absolute amount. Similarly proportional sacrifice requires that taxation is such that everybody's utility is reduced by the same proportion. Minimum sacrifice is derived from the utilitarian rule and requires that the utility sum of a society be reduced by the least amount possible given the revenue requirement.

Edgeworth argued for the superiority of utilitarianism over the principles of equal and proportional sacrifice. This is a question of ethics and cannot be settled conclusively.

Nevertheless, the notions of equality to which equal and proportional sacrifice correspond, and the use of these as bases for action do not appear to be in vogue at the present time and the principles have not received much attention in recent analyses of optimal income taxation.

Edgeworth analysed the implications of the minimum sacrifice principle in a model where everybody had the same utility function, which displayed increasing marginal utility of income but at a diminishing rate. Ignoring incentive problems he showed that the optimal schedule required that revenue be raised by not taxing incomes below a certain level Y^* at all, and by taxing away completely those above it. However, the required amount of government revenue might be such that incomes are not equated, so that there will be scope for increasing the sum of utilities further by lowering Y^* , and redistributing the income so raised, until all incomes are equalised. He recognised the question of incentives and argued that this should be taken into account when designing an actual schedule but did not formulate the problem rigorously.

The basic approach of Edgeworth has been used in most subsequent contributions to the theory of optimal income taxation. The developments have been in two main directions. The first concerns the use of social welfare functions which weight lower utilities more heavily than in a simple sum, with the Rawls maximin approach as a limiting case.

The second involves the use of more realistic models which take into account incentives and consider the distortions caused by the tax. Income can arise from a number

of different sources: the supply of labour, interest, rents and so on; the distortions which result from taxing these various components are different and are usually considered separately. In most actual economies income from labour is proportionately the most important and many recent analyses have concentrated on this type of earnings. The taxation of revenue from other factors, particularly capital (see, for example, Diamond (1970) and Feldstein (1974, 1978)) and the relationship between the taxes on income from different sources (see, for example, Atkinson and Stiglitz (1976)) have also received attention. The focus of this chapter is on the taxation of earnings from labour although with suitable modification the model of the next section is applicable to the taxation of other factors.

Mirrlees (1971) formulated a model for the investigation of the distortion on the choice between labour and leisure caused by an income tax which has been the basis of much of the subsequent discussion. The model is often referred to as either the labour or effort incentive model.

There is assumed to be one type of labour which is the only input to production and there are either constant or diminishing returns to scale. Output consists of one consumption good which is taken as numeraire. There is a distribution of ability A which has a density function $f(A)$ over the range zero to infinity. The population is normalised at 1. When a person of ability A works for time L he provides AL efficiency units of labour. The wage w for an efficiency unit of labour is equal to its marginal product which is determined by the aggregate labour supply. Any profits that

arise are assumed to accrue to the government. The income of a person with ability A is

$$(1) \quad Y = w A L$$

The time to be allocated between working and leisure is taken to be constant and the same for everybody. Utility is assumed to depend on consumption C and leisure, and can be represented by a function

$$(2) \quad U = U(C, L)$$

$$(3) \quad U_C > 0 ; U_L, U_{CC}, U_{LL} < 0$$

where a subscript denotes a partial derivative.

The tax schedule is assumed to depend on income and is represented by a function $t(Y)$. The budget constraint is then given by

$$(4) \quad C = Y - t(Y)$$

Each person chooses C and L to maximise (2) subject to the constraints (1) and (4), taking the tax schedule and their ability as given. Indirect utility can be expressed as a function of the tax schedule and ability.

$$(5) \quad V = V(t(Y), A)$$

It is assumed that the government's aims can be

The amount of education received does not enter the utility function directly. However, it does entail a cost $e(Z)$ in terms of consumption foregone. The function e is assumed to be convex so marginal costs are positive and rising. If F denotes after tax income then consumption is given by

$$(8) \quad C = F - e(Z)$$

Since labour supply is fixed utility depends only on consumption. Substituting from (8) utility can be expressed as a function of F and Z . Using the convexity of e the signs of the partial derivatives can also be found.

$$(2') \quad U = U(F, Z)$$

$$(3') \quad U_F > 0 ; U_Z, U_{FF}, U_{ZZ} < 0$$

It can be seen that (2') and (3') are the same as (2) and (3) except that F replaces C and Z replaces L . Since F is after tax income the budget constraint of the individual can be expressed as

$$(4') \quad F = Y - t(Y)$$

where F again corresponds to C in (4). As before each individual chooses F and Z to maximise (2') subject to the constraints (1') and (4'). Indirect utility can be

expressed as in (5) and the problem facing the government is as before, the only difference being the change of variables in individuals' maximisation of utility.

It can therefore be seen that the education incentive model is the same as the labour incentive model only with a reinterpretation of the variables. Any results proved for the labour incentive model are thus also applicable to the education incentive model.

Mirrlees (1971) concentrated on the utilitarian case with $h(V) = V$. He showed that two of the general properties of the tax schedule $t(Y)$ were that the optimal marginal tax rate lies between zero and one and that in most cases it will be optimal for some people not to work. To obtain further results he used specific utility and density functions. The utility function used was log-linear in leisure and consumption. The numerical solutions involved a lognormal distribution of abilities although the case where the tail of the skill distribution is Paretian was also considered analytically. Using parameters for the distribution of skills derived from the work of Lydall (1968), Mirrlees found from his numerical solutions that for the cases he considered the optimal tax structure was approximately linear and that the marginal tax rates were rather low and tended to fall rather than rise. Typical maximum marginal rates were around 20 per cent to 30 per cent. Also it was found that most of the population was included in the workforce.

Phelps (1973) and Sadka (1974) showed that with a finite number of skill levels or individuals, the marginal tax rate on income from labour at the very top of the earnings

distribution is zero. The result which holds for all Pareto-inclusive social welfare functions has a fairly simple intuitive explanation. Assume that the maximum earned by anybody is Y^* and consider the marginal tax rates on incomes greater than Y^* . If these are above zero then a reduction may induce somebody who earns Y^* to work more in which case he must be better off. Similarly if others below Y^* are induced to work harder they are made better off and, also, since revenue increases if their incomes move through bands where marginal tax rates are positive the welfare of others may also be improved.

Most of the other results which have been obtained are concerned with the case where the tax schedule is linear so that

$$(9) \quad t(Y) = t Y - I$$

This is of intrinsic interest since the nonlinear case presents a number of additional informational problems which may make it infeasible or too administratively costly. In particular it is necessary for a person's total income to be revealed whereas with linear taxation it can be taxed at source without it being necessary to observe individual incomes. However, perhaps the main reason for considering linear taxation is that it provides a simplification which allows analytical consideration of the various influences to be considerably extended.

Sheshinski (1971, 1972) showed in the case with constant returns to scale and no government expenditure that if it

is assumed firstly that leisure is a normal good, so that an increase in the lump sum grant I reduces the labour supply, and secondly that the elasticity of labour supply is non-negative, t is positive and has an upper bound that depends on the elasticity of labour supply.

Helpman and Sadka (1978) considered the effect of increasing the degree of inequality aversion in the social welfare function where there are again constant returns to scale in the production function. They showed that if h in (6) is strictly increasing, strictly concave and differentiable, the marginal tax rate and lump sum grant are higher than when the social welfare function consists of the sum of utilities. This transformation is equivalent to increasing the degree of inequality aversion. It therefore follows from repeating the transformation that the greater the degree of inequality aversion the higher are the optimal marginal tax rate and lump sum grant. As the degree of inequality aversion approaches infinity the social welfare function approaches the Rawls criterion of maximising the well-being of the worst-off person. This suggests that the marginal tax rate and lump sum grant corresponding to the sum of utilities social welfare function are less than the Rawlsian values, and Helpman and Sadka are able to show this directly. Their analysis confirmed the results of earlier numerical studies by Atkinson (1973) and Stern (1976).

These results thus tend to confirm the intuitive notion that with utilitarianism and similar ethical bases, taxes should redistribute from rich to poor and the greater the

weights put on lower utilities the more redistribution there should be. However, Mirrlees' (1971) original conclusion that utilitarianism implied low rates has been shown to be fairly sensitive to the formulation of the model. Stern (1976) considered the effect on linear tax rates of relaxing the assumption of a log-linear utility function and using a CES form instead. His survey of empirical estimates suggested that values of the elasticity of substitution between leisure and consumption of around 0.5 are of interest. Using this and other plausible parameter values he found that even with mild concavity of h , optimal tax rates of around 50 per cent are reasonable.

One of the strongest assumptions of the basic Mirrlees model on which these results are all based is that there is only one type of labour and differences in wages are due to differences in the amount of this which can be supplied in a given time period. It can perhaps be argued that a more natural representation of the production process involves the assumption that there are different types of labour which are combined together to produce output and that differences in wages are due to a difference in type, rather than quantity, of labour supplied by a person. The inclusion of a production function with more than one type of labour introduces an interdependence of labour supplies and incomes which significantly changes the nature of the problem of redistribution. Now, not only can income be redistributed through the fiscal system but there is also the possibility of redistribution through the production function by the changing of relative labour supplies and hence wages.

$$(1) \quad Y = Y(L_E, n L_R)$$

It is assumed that there are constant returns to scale. The production function can then be written in terms of output per unit of raw labour.

$$(2) \quad y = y(\ell)$$

where $y = Y / n L_R$ and $\ell = L_E / n L_R$. Factors are paid their marginal products so that the wages of the two groups also depend on the ratio of labour supplies.

$$(3a) \quad w_E = y'$$

$$(3b) \quad w_R = y - y' \ell$$

where y' denotes $\partial y / \partial \ell$.

The utility of both groups is dependent on their consumption and the labour they supply.

$$(4) \quad U_i = U(C_i, L_i) \quad i = E, R$$

with the usual assumptions concerning the partial derivatives.

The government redistributes income using a linear tax system. Everybody receives a lump sum grant which is financed by a proportionate tax on income. The government's revenue constraint is given by

$$(5) \quad I = t k - q$$

Feldstein (1973) investigated the effects of assuming that production required two types of labour by numerically solving for the optimal linear income tax in a specific model with a Cobb-Douglas production function. For the parameters he chose, he found that optimal linear tax rates were little changed from those obtained when wages were held constant. In the next section it is argued that these results are a special case. The use of a Cobb-Douglas production function ensures that redistribution via the fiscal system and via the production function are reinforcing. In general this does not happen.

There is a large range of possible assumptions concerning the labour supply of the two groups and the degree of substitutability between the different types of labour. However, even in the case considered below where both types of labour are regarded as the same in terms of the disutility of work, and there is a low elasticity of substitution, it is shown that the fiscal and production effects can be reinforcing or opposing depending on the nature of the labour supply curve. If the elasticities of both groups are positive the effects operate in the same direction and the optimal linear tax has a positive marginal rate. If there is a backward bending labour supply curve, such that the better paid group is on the portion with a negative elasticity and the lower paid group has a positive elasticity, the redistribution through the production function acts in the opposite direction to that through the fiscal system and may more than offset it. Thus the optimal linear tax may have a negative marginal rate so that there is a wage subsidy which is financed by a lump sum tax. In this case direct

fiscal redistribution is from poor to rich but causes a change in relative labour supplies and an improvement in the wages of the less prosperous group which overall makes them better off.

These results do not depend on the form of the social welfare function; they arise in both the Rawlsian and utilitarian cases. The important assumption is the inclusion of a production function.

In Subsection 2.1. the model is described. In 2.2. the incidence of the tax is investigated. 2.3. contains the normative analysis where the tax is considered in the context of Rawlsian and utilitarian social welfare. 2.4. consists of the cases described above and 2.5. contains some concluding remarks.

4.2. A theory of optimal linear income taxation with general equilibrium effects on wages

4.2.1. The model

The model developed below looks at the general equilibrium effects that income taxation has on wages. Attention is confined to the simplest case where there are two types of labour and the tax function is linear. The two types of labour required to produce a single homogeneous output are called entrepreneurial skills and raw labour. However, these are just labels and can easily be reinterpreted as, for example, skilled and unskilled labour. The characteristics of the entrepreneurs are denoted by the subscript E and those of the workers by the subscript R. The ratio of entrepreneurs to workers is one to n and the groups are assumed to be homogeneous.

In the production function the amount of output is dependent on the amounts of labour supplied by the two groups.

$$(1) \quad Y = Y(L_E, n L_R)$$

It is assumed that there are constant returns to scale. The production function can then be written in terms of output per unit of raw labour.

$$(2) \quad y = y(\ell)$$

where $y = Y / n L_R$ and $\ell = L_E / n L_R$. Factors are paid their marginal products so that the wages of the two groups also depend on the ratio of labour supplies.

$$(3a) \quad w_E = y'$$

$$(3b) \quad w_R = y - y' \ell$$

where y' denotes $\partial y / \partial \ell$.

The utility of both groups is dependent on their consumption and the labour they supply.

$$(4) \quad U_i = U(C_i, L_i) \quad i = E, R$$

with the usual assumptions concerning the partial derivatives.

The government redistributes income using a linear tax system. Everybody receives a lump sum grant which is financed by a proportionate tax on income. The government's revenue constraint is given by

$$(5) \quad I = t k - q$$

where I is the lump sum grant, t is the proportionate tax, q is the government's per capita revenue requirement for projects other than income redistribution and k is per capita income.

$$(6) \quad k = \frac{w_E L_E + n w_R L_R}{n + 1}$$

It is assumed that consumption C is equal to net income so that

$$(7) \quad C_i = (1 - t) w_i L_i + I \quad i = E, R$$

Each person is assumed to maximise their utility subject to their budget constraint, taking their wage, the proportionate tax and the lump sum grant as given. Since the wage rates are dependent on the labour supplies of both groups they are also implicitly constrained by the labour supply function of the other type of worker.

$$(8) \quad \text{Max}_{C_i, L_i} U(C_i, L_i)$$

subject to (7) and

$$(9) \quad L_j = L((1 - t) w_j, I)$$

for $i, j = E, R$; $i \neq j$.

(8) gives rise to an indirect utility function which is denoted by

$$(10) \quad V_i = V((1 - t) w_i, I) \quad i = E, R$$

It follows from the theory of indirect utility functions that (9) and (10) are related by

$$(11) \quad L_i = \frac{\partial V_i / \partial w_i}{\partial V_i / \partial I} \quad i = E, R$$

The characteristic of the entrepreneurs that distinguishes them from the workers is that they have greater entrepreneurial ability. It is assumed that they earn more than the workers, otherwise if the disutility of the work is the same, the model will degenerate into one of self-employment where each person supplies both types of labour himself. This self-employment model is like the labour incentive model where each person supplies one type of labour and some people are more efficient than others, but earnings cannot be represented by a simple linear function except in special cases. However, the situation that is looked at here is when the entrepreneurial wage is higher than the raw labour wage and the workers prefer working for the entrepreneur because their wages are higher than when they are self-employed, due to the entrepreneur's greater ability.

$$(12a) \quad w_E > w_R$$

or using (3)

$$(12b) \quad y'(1 + \ell) > y$$

It is also assumed that the total earnings of each member of the entrepreneurial group are larger than those of each member of the workers' group. That is

$$(13a) \quad w_E L_E > w_R L_R$$

or using (3)

$$(13b) \quad y' \ell (n + 1) > y$$

If these conditions are satisfied by the production function for the appropriate labour supplies then in the case where the two types of labour are regarded as the same as far as the person supplying them is concerned, (13) will hold when (12) does, provided the elasticity of substitution between consumption and leisure in the utility function is positive. This is usually taken to be the case.

4.2.2. The incidence of a linear tax

In this subsection the incidence of a linear tax is considered. The crucial variable is the ratio of labour supplies ℓ and $d \log \ell / dt$ is therefore found first.

From the definition of ℓ it follows that

$$(14) \quad \frac{d \log \ell}{dt} = \frac{d \log L_E}{dt} - \frac{d \log L_R}{dt}$$

From (9)

$$(15) \quad \frac{d \log L_i}{dt} = -\epsilon_i + \epsilon_i \frac{d \log w_i}{dt} + \zeta_i \frac{dI}{dt} \quad i = E, R$$

where $\epsilon_i = \frac{w_i}{L_i} \frac{\partial L_i}{\partial w_i}$, the elasticity of labour supply of the members of group i and $\zeta_i = \frac{1}{L_i} \frac{\partial L_i}{\partial I}$, the percentage change in labour supply caused by an increase in the lump sum grant.

From (3)

$$(16a) \quad \frac{d \log w_E}{dt} = -\xi_E \frac{d \log \ell}{dt}$$

$$(16b) \quad \frac{d \log w_R}{dt} = \xi_R \frac{d \log \ell}{dt}$$

where ξ is the elasticity of the wage with respect to the labour ratio.

$$(17a) \quad \xi_E = -\frac{\ell}{w_E} \frac{\partial w_E}{\partial \ell} = -\frac{\ell}{y'} \frac{y''}{y'} > 0$$

$$(17b) \quad \xi_R = \frac{\ell}{w_R} \frac{\partial w_R}{\partial \ell} = \frac{-\ell^2}{y - y' \ell} \frac{y''}{y'} > 0$$

Equations (14), (15) and (16) can be solved simultaneously to find $d \log \ell / dt$ in terms of the elasticities and dI/dt . This gives

$$(18) \quad \frac{d \log \ell}{dt} = \frac{D(dI/dt)}{G}$$

where

$$(19) \quad D\left(\frac{dI}{dt}\right) = (\epsilon_R - \epsilon_E) + \frac{dI}{dt} (\zeta_E - \zeta_R)$$

$$(20) \quad G = 1 + \xi_E \epsilon_E + \xi_R \epsilon_R$$

The term $d \log \ell / dt$ can be interpreted in the following way. The tax alters the labour supplies directly because the net wage received is altered. The term $D(dI/dt)$ represents this direct effect. $(\epsilon_R - \epsilon_E)$ is the change due to the proportionate part of the tax and $(dI/dt)(\zeta_E - \zeta_R)$ is due to the lump sum grant. However, the change in relative labour supplies alters marginal products and hence gross wages which in turn alter labour supplies. The denominator G represents this indirect general equilibrium effect. The signs of these terms are considered at greater length in Subsection 2.4. where it is shown that they can be such that $d \log \ell / dt$ can be either positive or negative.

The change in wages and labour supplies can be found using (18) in (14) and (15). Differentiating (10) gives the change in utilities.

$$(21) \quad \frac{dV_i}{dt} = -w_i \frac{\partial V_i}{\partial w_i} + \frac{\partial V_i}{\partial w_i} \frac{dw_i}{dt} + \frac{\partial V_i}{\partial I} \frac{dI}{dt} \quad i = E, R$$

Using (11) this simplifies to

$$(22) \quad \frac{dV_i}{dt} = \frac{\partial V_i}{\partial I} \left[w_i L_i \frac{d \log w_i}{dt} + \frac{dI}{dt} - w_i L_i \right] \quad i = E, R$$

The term in brackets can be split up into two components. The first is the redistributive effect via the production function due to the change in gross wages, $w_i L_i d \log w_i / dt$. The second is the effect via the fiscal system which is the

change in the net tax payment, $dI/dt - w_i L_i$.

Assuming the government's per capita revenue requirement q is constant, it follows from (5) that

$$(23) \quad \frac{dI}{dt} = k + t \frac{dk}{dt}$$

where from (6)

$$(24) \quad \frac{dk}{dt} = \frac{1}{n+1} \left[w_E L_E \left[\frac{d \log w_E}{dt} + \frac{d \log L_E}{dt} \right] + n w_R L_R \left[\frac{d \log w_R}{dt} + \frac{d \log L_R}{dt} \right] \right]$$

A special case of particular interest is that where $t = 0$ so that (23) becomes

$$(25) \quad \frac{dI}{dt} = k$$

and it is the effect of introducing a small linear tax at the market solution which is considered. Using (6), (16), (18), (22) and (25) it follows that the changes in entrepreneurs' and workers' utilities are given by

$$(26) \quad \left(\frac{dV_E}{dt} \right)_{t=0} = - w_E L_E \frac{\partial V_E}{\partial I} \left[\frac{\xi_E D(k)}{G} + \frac{n(X-1)}{(n+1)X} \right]$$

$$(27) \quad \left(\frac{dV_R}{dt} \right)_{t=0} = w_R L_R \frac{\partial V_R}{\partial I} Z$$

where

$$(28) \quad Z = \frac{\xi_R D(k)}{G} + \frac{X-1}{n+1}$$

and $X = w_E L_E / w_R L_R$ the ratio of an entrepreneur's earnings to a worker's.

It follows from (3) and (17) that

$$(29) \quad \frac{n}{X} = \frac{\xi_E}{\xi_R}$$

Using this to simplify (26) it can be seen that

$$(30) \quad \left(\frac{dV_E}{dt} \right)_{t=0} = - n w_R L_R \frac{\partial V_E}{\partial I} Z$$

Hence since $\partial V_i / \partial I > 0$, the crucial determinant of the direction of movement of utilities is the condition

$$(31) \quad Z \begin{matrix} > \\ < \end{matrix} 0$$

This determines whether or not the workers are made better off by the tax and the entrepreneurs worse off, or vice versa. $\xi_R D(k) / G$ is the redistributive effect via the production function and $(X - 1) / (n + 1)$ is the redistribution through the fiscal system. It follows from (13) that the latter is always positive. However, it is shown in Subsection 2.4. that $\xi_r D(k) / G$ can plausibly have either sign and can dominate $(X - 1) / (n + 1)$ when negative. Thus because of the change in wages caused by the alteration in relative labour supplies it is not clear that a positive linear tax which redistributes income through the fiscal system from entrepreneurs to workers will make the latter better off.

It can be seen that the labour incentive model with constant returns to scale is a special case of the one considered here. Since wages then only depend on the ability of the person supplying the labour ξ_E and $\xi_R = 0$, and there is no redistribution through the production function. Thus using (13) it follows from (26) and (27) that the workers are always made better off and the entrepreneurs worse off by the introduction of a small linear tax. It is the inclusion of the production effect which is the important feature of the model considered here.

4.2.3. The normative analysis

In this subsection the results of 4.2.2. are used to investigate the effect of a linear tax on Rawlsian and utilitarian social welfare, which are denoted S and S_u respectively, and to derive the first order condition for the Rawlsian optimal schedule. S and S_u are defined by

$$(32) \quad S = V_R$$

$$(33) \quad S_u = n V_R + V_E$$

It can be seen directly from (27) that whether the introduction of a small linear tax at the market solution improves Rawlsian social welfare depends on the condition in (31).

Differentiating (33) with respect to t and evaluating at $t = 0$, it follows from (27) and (30) that

$$(34) \quad \left(\frac{dS_u}{dt} \right)_{t=0} = n w_R L_R \frac{\partial V_R}{\partial I} (1 - M) Z$$

where $M = \frac{\partial V_E / \partial I}{\partial V_R / \partial I}$, the ratio of marginal utilities of income. It is usual to assume

$$(35) \quad M < 1$$

so that the marginal utility of income of the entrepreneurs is less than that of the workers. This is certainly the case where the direct utility function (4) is additive in consumption and labour. However, there may be values of the cross partial derivative U_{CL} such that (35) is not satisfied and in this case the results stated below will not hold for the utilitarian social welfare function.

It is now possible to state Theorem 1.

Theorem 1

$(dS/dt)_{t=0}$ and $(dS_u/dt)_{t=0}$ have the same sign which is positive if and only if

$$(36) \quad - \frac{\xi_R D(k)}{G} < \frac{X - 1}{n + 1}$$

and negative if and only if the inequality is reversed.

The extent to which a small linear tax improves or reduces Rawlsian and utilitarian social welfare thus depends on whether the production effect reinforces or opposes the fiscal effect and their relative magnitudes; some possible cases are considered in the following subsection.

The next step is the determination of an expression which the optimal tax rate t must satisfy. These are similar for both Rawlsian and utilitarian social welfare, the difference

being that in the latter case there are extra terms to take into account the utility of the entrepreneurs. Since the expressions are algebraically complex only the Rawlsian case is considered here, the utilitarian case is given in Allen (1979). The important point is that the results arise from positive rather than normative assumptions and do not depend on the form of the social welfare function.

The government's problem in the Rawlsian case is

$$(37) \quad \underset{t}{\text{Max}} S = V_R((1 - t) w_R, I)$$

The first order condition requires

$$(38) \quad \frac{dS}{dt} = \frac{dV_R}{dt} = 0$$

Taking $\partial V_R / \partial I \neq 0$ and using (22) with $i = R$, and also (16) and (18), it follows that for t which are such that (38) is satisfied

$$(39) \quad \frac{dI}{dt} = \frac{w_R L_R \{G + \xi_R (\epsilon_E - \epsilon_R)\}}{G + w_R L_R \xi_R (\zeta_E - \zeta_R)}$$

Solving (15), (16), (18), (23) and (24) simultaneously for dI/dt , using (29) and rearranging it can be shown that

$$(40) \quad \frac{dI}{dt} = \frac{(n + 1)k/w_R L_R + t\{n\xi_R(\epsilon_E - \epsilon_R)^2/G - (X\epsilon_E + n\epsilon_R)\}}{(n + 1)/w_R L_R + t\{n\xi_R(\epsilon_E - \epsilon_R)(\zeta_E - \zeta_R)/G - (X\zeta_E + n\zeta_R)\}}$$

It is then possible to eliminate dI/dt by equating (39) and (40) and simplify to give Theorem 2.

Theorem 2

For all t such that $dS/dt = 0$

$$(41) \quad (n + 1)\xi_R D(k) + G(X - 1) = t \{ n\xi_R(\epsilon_E - \epsilon_R)D(w_R L_R) \\ + (G + w_R L_R \xi_R(\zeta_E - \zeta_R))(X\epsilon_E + n\epsilon_R) \\ - w_R L_R(G + \xi_R(\epsilon_E - \epsilon_R))(X\zeta_E + n\zeta_R) \}$$

Hence if there is an interior solution the Rawlsian optimal tax must satisfy this equation.

Equation (41) defines t in terms of $w_E L_E$, $w_R L_R$, ϵ_E , ϵ_R , ζ_E , ζ_R , ξ_E and ξ_R . In general these terms will also depend on t so that there may be many solutions to (41) apart from the optimal one. Also the t which satisfy (41) may be local minima as well as local maxima.

4.2.4. Some possible cases

In considering possible cases, a large number of assumptions concerning the labour supply of the two groups are conceivable. The enjoyment of the two types of work may be different, leisure opportunities may not be the same because of different abilities and so on. The simplest case where both types of labour are regarded as the same, as far as the person supplying them is concerned, is considered below.

It can be seen from Theorem 1 that the production effect will reinforce the fiscal effect if $D(k)$ and G have the same signs, and oppose it if they have different signs.

With a conventional type of labour supply curve, a

higher wage corresponds to a lower elasticity, so that $(\epsilon_R - \epsilon_E)$ will be positive. Also, since a lump sum grant is a smaller component of an entrepreneur's income than of a worker's, it seems likely it will have a smaller proportionate effect on his labour supply than on the worker's. Since the effect of a lump sum grant is usually to reduce labour supply, the ζ terms are negative, and $(\zeta_E - \zeta_R)$ will therefore also be positive. On the basis of this reasoning, $D(k)$ will be positive. It is shown in Allen (1979) that when labour supply is derived from a general utility function which has consumption, leisure and the two types of labour as arguments, both signs of $D(k)$ are in fact possible, although a positive value is more plausible. With a CES utility function in consumption and leisure of the type Feldstein (1973) used, it is also shown that $D(k)$ is positive except in the unusual case where government expenditure and the elasticity of substitution between consumption and leisure are high. $D(k)$ is thus taken to be positive in the examples below.

Even with this restriction it is possible to show that the production effect can operate in both directions because G can plausibly have both signs. A positive elasticity of substitution between consumption and leisure implies that ϵ_E and $\epsilon_R > -1$. Since ξ_E and ξ_R are positive, G can be negative only if $\xi_E + \xi_R > 1$. It follows from (17) that

$$(42) \quad \xi_E + \xi_R = \frac{-y y'' l}{y' (y - y' l)} = \frac{1}{\sigma}$$

where σ is the elasticity of substitution. With a Cobb-Douglas production function of the type Feldstein (1973) used it follows that $\xi_E + \xi_R = 1$ so that G must be positive. It is for this reason that his solutions were such that the fiscal and production effects were always reinforcing. However, for production functions where σ is less than one, the possibility arises that G will be negative. The values of ξ_E and ξ_R can be more readily seen by solving (29) and (42) simultaneously. This gives

$$(43a) \quad \xi_E = \frac{n w_R L_R / Y}{\sigma}$$

$$(43b) \quad \xi_R = \frac{w_E L_E / Y}{\sigma}$$

Whether σ is greater or less than one would seem to depend to some extent on the interpretation of the two types of labour. If entrepreneurial and raw labour are interpreted as skilled and unskilled labour respectively, there may be scope for substitution and σ could be large. In this case G will be positive. If entrepreneurial and raw labour are taken to refer to these actual types it seems likely that there will not be much scope for substitution so that σ will be less than one. If either ϵ_E or ϵ_R , or both, are negative G may also be.

In the case where the production effect opposes the fiscal effect, however, this is not yet sufficient to demonstrate the possibility that a small linear tax will actually reduce social welfare. The following theorem shows

that in theoretically plausible cases, both signs are possible for $(dS/dt)_{t=0}$ and $(dS_u/dt)_{t=0}$, and illustrates how sensitive the condition is to the particular assumptions made.

Theorem 3

- (a) If at $t = 0$, $\epsilon_E, \epsilon_R > 0$; $\zeta_E, \zeta_R < 0$; and the magnitudes of these, $n, w_E L_E$ and $w_R L_R$ are such that $D(k) > 0$, the condition in Theorem 1 is satisfied with the inequality ' $<$ ' so that a small positive linear tax at the market solution improves Rawlsian and utilitarian social welfare.
- (b) If the requirements are as in part (a) except that $\epsilon_E < 0$, $(\zeta_E - \zeta_R) > 0$, and $G < 0$, the condition in Theorem 1 is satisfied with the inequality ' $>$ ' so that a small positive tax at the market solution reduces Rawlsian and utilitarian social welfare.

Proof

- (a) Since ξ_E, ξ_R, ϵ_E , and ϵ_R are positive, G must be positive. Combining this with the positive sign of $D(k)$ implies the left hand side of the condition in Theorem 1 is negative. Since (13) implies $(X - 1)$ is positive the condition must be satisfied with the inequality ' $<$ '.

- (b) Using (29) it can be shown that

$$\begin{aligned}
 (44) \quad A &= G (X - 1) + (n + 1) \xi_R D(k) \\
 &= (1 + \xi_R \epsilon_R) (X - 1) - \xi_E \epsilon_E - \xi_R \epsilon_E \\
 &\quad + (n + 1) \xi_R \epsilon_R + (n + 1) \xi_R k (\zeta_E - \zeta_R)
 \end{aligned}$$

Combining the assumptions in the theorem with (13) and (17) it follows that all the terms in this last expression are positive. Hence, since $G < 0$ this implies that the condition in Theorem 1 is satisfied with the inequality '>'.

The situations described in the theorem can both arise if, for example, the types of labour are taken to have a low elasticity of substitution in the production function and the labour supply curve is backward bending. The requirement in (b) concerning magnitudes that $G < 0$ while $\epsilon_R > 0$, means that either $-\epsilon_E > \epsilon_R$ or $\xi_E > \xi_R$, or both. The former depends on the relative positions on the labour supply curve. Looking at (29) it can be seen that the latter will be fulfilled if n is larger than X . Either or both of these could plausibly be satisfied.

The difference between the parts of the theorem is that in (a) wages are such that both are on the positive portion of the curve and in (b) the entrepreneurs have moved to the negative portion. The sensitivity of the sign of the change in social welfare when a small linear tax is introduced is due to the term G which represents the indirect general equilibrium component of the change in relative labour supplies. It arises because a change in taxes alters relative

labour supplies which changes wages, which in turn again alter labour supplies. When the elasticity of substitution σ is small, ξ_E and ξ_R are large, and this indirect effect can become important.

Theorem 3 applies both to Rawlsian and utilitarian social welfare. The next result is concerned with the absence of turning points in the function expressing Rawlsian social welfare in terms of t . It can be used together with Theorem 3 to demonstrate the Rawlsian optimal tax can be either positive or negative. Similar results are given for the utilitarian case in Allen (1979); the perverse examples where the optimal tax is negative occur because of the inclusion of a production function not the specification of social welfare.

Theorem 4

- (a) For all negative t such that the conditions in Theorem 3(a) are satisfied together with the additional requirement that $(\zeta_E - \zeta_R) > 0$, there are no values of S such that $dS/dt = 0$.
- (b) For all positive t such that the conditions in Theorem 3(b) are satisfied, there are no values of S such that $dS/dt = 0$.

Proof

- (a) It can be seen directly that the assumptions imply that the left hand side of (41) is positive. Rearranging the coefficient of t in (41) which will be denoted B ,

$$\begin{aligned}
(45) \quad B = & 2 n \xi_R \epsilon_E \epsilon_R + n w_R L_R \xi_R \epsilon_E (\zeta_E - \zeta_R) \\
& + (1 + \xi_R \epsilon_R) X \epsilon_E + (1 + \xi_E \epsilon_E) n \epsilon_R \\
& + w_R L_R \xi_R (\zeta_E - \zeta_R) X \epsilon_E \\
& - w_R L_R \{(1 + \xi_E \epsilon_E) + \xi_R \epsilon_E\} (X \zeta_E + n \zeta_R)
\end{aligned}$$

It follows directly from the assumptions in the theorem that this is positive. Hence there can be no turning points for negative t .

(b) A in (44) is the numerator of t and it thus follows that it is positive. It can also be seen from (45) that since the assumptions imply that $1 + \xi_E \epsilon_E < 0$, B is negative. Hence there can be no turning points for positive t .

If the conditions in Theorem 4(a) are satisfied for $t = 0$ and all feasible negative t it is possible to prove that the optimal tax is positive since Theorem 4(a) then rules out negative interior maxima and Theorem 3(a) rules out negative corner maxima. Similarly if the conditions for Theorem 4(b) are satisfied for $t = 0$ and all feasible positive t the optimal tax is negative. However, whereas it is possible for the conditions in 4(a) to be satisfied for all negative t it is not for those in 4(b) to be satisfied for all positive t . Although the production effect causes entrepreneurs' gross wages to rise as t becomes higher, with a backward bending labour supply curve, as t approaches one, G and ϵ_E will eventually switch to being positive as net wages become

small. There may then be a number of turning points for positive t .

It is possible to argue that the failure of the conditions to hold at extreme values of t is unimportant since the optimum is unlikely to be in this region. If it were to be, it would be necessary for there first to be a minimum, a fairly steep rise and then a maximum, all at values of t near one. This would also have to be greater than any maximum for negative t . A situation of this type is shown in Figure 1 and it might be thought reasonable to regard possibilities of this sort as being unlikely.

Theorem 5 then follows.

Theorem 5

- (a) If the conditions in Theorem 4(a) are satisfied for $t = 0$ and all feasible negative t , the Rawlsian optimal value of t is positive.
- (b) If the conditions in Theorem 4(b) are satisfied for $t = 0$ and all, except extreme, feasible positive t and situations such as those in Figure 1 are disregarded, the Rawlsian optimal value of t is negative.

Carruth (1980) has undertaken numerical simulations which confirm these results. He found no examples of multiple turning points.

4.2.5. Concluding remarks

It has been argued that a serious deficiency of many of

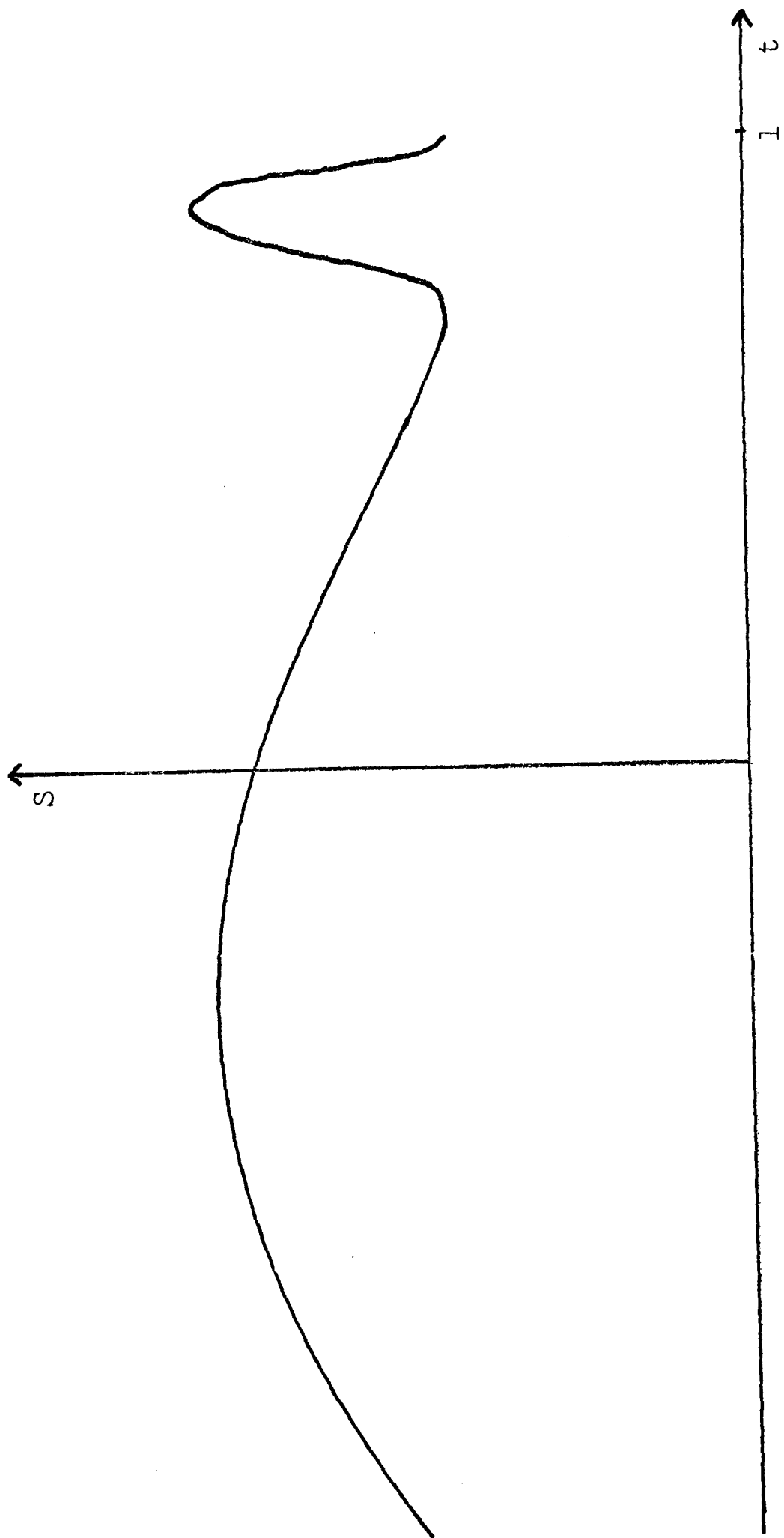


FIGURE 1

the recent models that have been used to investigate the taxation of income from labour has been their production assumptions. An exception was Feldstein (1973) but his use of a CES utility function and a Cobb-Douglas production function was a special case. The former ensures $D(k)$ will be positive except in very unusual circumstances and the latter combined with a positive elasticity of substitution between consumption and leisure implies G must be positive. The production effect at the market solution will then always reinforce the fiscal effect. It is therefore not surprising that, although one cannot rule out negative optimal taxes in his model, they do not occur in his calculations.

In Subsection 2.4. it was demonstrated not only that the production effect at the market solution may have both signs, but that this could outweigh the fiscal effect. It was also shown that it was possible for the optimal tax to have either sign. A case was described in which the production effect is predominant, characterised by a low elasticity of substitution between the types of labour and a negative labour supply elasticity for the entrepreneurs. It seems by no means implausible. In this case all that is required for the sign to change is for the entrepreneurs' labour supply elasticity to switch from a positive to a negative value.

Stern (1979) has derived a similar result for the optimal nonlinear income tax. He showed analytically that, with a production function in two types of labour, this involves a positive marginal tax rate for the lowest paid workers if leisure is a normal good and a negative marginal rate for the highest paid if the marginal social utility of

their consumption is lower than that of the other group.

However, there are a number of problems if the optimal linear tax is negative and involves a wage subsidy which is financed by a lump sum tax. One of the features of a positive linear income tax consisting of a lump sum grant and a proportionate tax is that everybody is guaranteed a minimum income. But with a negative linear income tax which increases social welfare by exploiting the production effect, it is important that everybody is an employed wage earner since otherwise the worst-off cannot benefit from the higher wage the tax induces. Whereas a positive tax could possibly be implemented on its own, a negative tax would need to be supplemented by a system for those who did not work, as is usually done in actual tax systems.

There are also problems in implementing a wage subsidy. One of the difficulties of enforcing a positive tax is that people tend to conceal their true incomes and declare their earnings to be lower than they actually are. However, the problems of doing this on a large scale are usually such that it is worthwhile only for those with high incomes to do it; and even then there is an upper bound to the amount of tax that can be avoided. With a negative tax, where the proportionate component takes the form of a subsidy, everybody has an unlimited incentive to conceal their true incomes and inflate their declared earnings. It would be very difficult to enforce such a system since artificial jobs paying large amounts can usually be created without problems and there is no upper bound to the amount that can be claimed, because the purpose of the system is to subsidise higher incomes the most.

A negative linear tax is equivalent to a proportionate subsidy on consumption and it might be more feasible to implement the tax in this way. However, it would still be fairly easy to create false transactions and exploit the system. Only if all income came from the government and there was no private sector might it be possible for a negative linear income tax to be implemented. In cases where the optimal linear tax was negative and there were no positive interior or corner maxima, a market system, with uniform lump sum taxation to raise the government revenue requirement, would be preferable to a positive linear tax as far as those on low incomes were concerned.

Whereas in the application of the labour incentive model it is possible to draw on observed distributions of earnings for the specification of $f(A)$, it is more difficult to use available statistics as the basis for the parameters of the model presented above - notably for those of the production function. In reality there are many different types of labour with many different magnitudes for the elasticity of substitution between them, and with a whole range of incomes. There is also the problem that only income from labour is considered. Nevertheless, it may be of interest to briefly consider parameter values that have been suggested as appropriate.

Stern's (1976) survey of empirical work suggests that if a CES utility function in consumption and leisure is used an elasticity of substitution of around 0.5 is of interest. This implies negative labour supply elasticities at all wages, and also that $D(k)$ is positive. A division of earners into

entrepreneurs or senior managers and workers would seem appropriate for this exercise. In this case the scope for substitution would seem small so that ξ_E and ξ_R will be large. Combining this with the negative values of ϵ_E and ϵ_R implies G will be negative so that the sign of the production effect at the market solution will be negative. It will act in the opposite direction to the fiscal effect and tend to offset it. It is not possible to say which effect will predominate without numerical values but it seems likely that overall the introduction of a positive linear tax will have little redistributive effect. Although it would be wrong to place much, if any, emphasis on this application of the model, it does illustrate an important conclusion. Because of changes in wages caused by changes in relative labour supplies attempts to alleviate poverty by progressive income taxation may not be very successful, and may even exacerbate it, especially if labour supply elasticities are negative and different types of labour are complements rather than substitutes.

CHAPTER 5

CONCLUSIONS

In organising its economic activity a society faces two main problems: the allocation of resources for production and the distribution of income for consumption. If abilities can be identified, then provided the usual conditions concerning externalities and so on are satisfied, competitive markets ensure the efficient allocation of resources for production: prices of factors are bid up until only the best entrepreneurs can produce enough to pay and marginal products are equated. There remains the problem of achieving the desired distribution of income while at the same time providing the correct incentives for the supply of factors but this can be solved by the appropriate use of lump sum taxes by the government.

The thesis has been concerned with the case where abilities cannot be identified and this ideal solution is not achieved. Chapters 2 and 3 considered the allocation of control of productive resources where transactors were unable to distinguish between people with different abilities and Chapter 4 the distribution of income and incentives where the government was unable to identify ability.

In Chapter 2, the focus was on the allocation of control arising in the capital services market. If there are differences in the amount people produce because of varying abilities, lenders will be unsure of whether borrowers who have insufficient collateral will be able to make the

payment for the use of their capital. It seems unlikely that anybody would lend to a person who had already defaulted since this is an indicator that he would probably do so again. It is therefore assumed that the penalty for people who default is that they will in future be unable to borrow. This implies that in order for a contract to be enforceable the current payment must be less than the value of future access to the capital market.

The main result of the chapter was to show that provided borrowers have similar time horizons and rates of discount, the contract which ensures that lenders receive the maximum amount possible from people of different abilities, without precipitating default, is one which makes the payment for capital a share of the output produced. Furthermore, it was argued that this is the type of contract which would be used since the differing interests of borrowers and lenders concerning default ensures that the latter specify the contract.

The lenders' share depends on the time horizon and rate of discount of borrowers; the amount of capital lent is determined by the requirement that the expected return from share contracts be equal to that from renting to people with collateral. If the output of a borrower is very low then lenders are better off allowing them to default and then in subsequent periods renting out the capital to people with collateral; this means that lenders impose a minimum requirement on output.

The structure of the share contract is such that it may lead to borrowers having more or less capital than they

would choose if renting. They might have more because of the share in pure profits such a contract allows the lender compared to renting. They may receive less because of the low level of the share or because of the necessity of compensating for the defaults of those who produce below the minimum requirement.

Owners of capital may be better off borrowing more than their collateral allows them to, by using a share contract. In order to indicate that they expect a return at least as great as that of renting to people with collateral, they must use their own capital and rent all that this allows them to; anything more than this is borrowed on the basis of a share of the difference between output and rental payments. In this case the debt equity ratio is determined by the amount the borrower owns; both this ratio and the total amount that can be borrowed are higher the greater this is.

The analysis of Chapter 2 points to the difficulties that arise in borrowing capital when abilities differ and suggests that the distribution of ownership of wealth among the different groups may be important in determining the productive capacity of an economy. Chapter 3 was concerned with the relationship between ability and wealth and the role of inheritance, in the extreme case where people can only use their own capital for production, and with the effect of inheritance taxation. In the model used, one of the ability groups is assumed to be much better than the other at organising factors of production and the intergenerational transmission of ability is Markovian in form. The theory of capital supply and bequest behaviour is based on that of

Atkinson (1971a). People of high ability either found, expand or maintain the family fortune. If born with an inheritance low ability people squander it; if born without one their income is earned by supplying the manual labour needed for production.

The model thus describes an economy of capitalists and workers in which firms are associated with families and their history depends on the outcome of the transmission of ability between generations. One of the main tasks of the chapter was to describe the various types of steady states that could arise with this structure.

A number of further simplifying assumptions allowed the effect of taxation to be considered. The focus was on inheritance taxation since this might be expected to have a significant impact on the relationship between ability and wealth-holding. The effect of such a tax was found to be complex and ambiguous. The various direct and indirect general equilibrium effects were identified and it was shown that these might plausibly be such that the variables of interest could move in either direction. The importance of these results lies in their giving some idea of when a variable will move in a particular direction.

A wage tax was also considered as an illustration of the effect of another means of raising revenue. This was compared with the inheritance tax on the basis of the change in Rawlsian social welfare for an equal yield in revenue. It was shown that even though workers leave no inheritance and their only source of income is wages, the changes caused by the two taxes in Rawlsian social welfare, which just takes

into account their utility, can be such that a wage tax is preferable. An example of when this perversity arises is the plausible case where labour and capital supplies are positive and an inheritance tax reduces bequests.

If abilities are observable then lump sum taxation can be used by the government to achieve the desired distribution of income while at the same time providing the correct incentives for the supply of factors of production. Without this information other characteristics must be used as the basis for taxation; the most commonly suggested being the amount of income received. However, the use of such proxies can lead to distortions which may offset any improvements in distribution. Chapter 4 was concerned with the use of linear income taxation for redistribution.

The model considered involves one ability group supplying entrepreneurial labour and the other raw labour: these are combined together to produce output. Both types are paid their marginal products which depend on relative labour supplies. This endogeneity of wages implies that if the tax alters labour supplies then not only is there direct redistribution through the fiscal system but there is also an indirect general equilibrium effect caused by the change in wages.

It was shown that not only is it plausible for these two effects to be reinforcing or opposing, but in the latter case the general equilibrium effect can dominate. Thus the introduction of a small proportionate wage tax to finance a lump sum grant so that there is redistribution through the fiscal system from rich to poor, may actually make the

latter worse off because of the change in wages. It was also argued that if the government is interested in maximising Rawlsian or utilitarian social welfare then for similar reasons a negative linear tax, consisting of a wage subsidy financed by a lump sum tax so that there is redistribution through the fiscal system from poor to rich, may well be optimal. The plausible example given where this occurs involves a low elasticity of substitution for the production function, a negative labour supply elasticity for the highly paid people and a positive one for the other group.

In conclusion, the two-ability model used in the thesis has thus allowed three main insights into the operation of economies. The first concerns the way in which share contracts can simultaneously prevent default and maximise the return to lenders. The second is the illustration of how the combination of an imperfect capital market and differences in entrepreneurial ability lead to inequality of wealth and that attempts to reduce this may be undesirable. The third is the importance of the general equilibrium effects of taxes since these can offset and dominate the direct effects.

REFERENCES

- Allen, H. F., (1979), 'Optimal income taxation with general equilibrium effects on wages', M.Phil. thesis, Oxford.
- Atkinson, A. B., (1970), 'On the measurement of inequality', Journal of Economic Theory, 2, pp. 244-263.
- Atkinson, A. B., (1971), 'The distribution of wealth and the individual life cycle', Oxford Economic Papers, 23, pp. 239-254.
- Atkinson, A. B., (1971a), 'Capital taxes, the redistribution of wealth and individual savings', Review of Economic Studies, 38, pp. 209-227.
- Atkinson, A. B., (1972), Unequal Shares, Allen Lane, The Penguin Press.
- Atkinson, A. B., (1973), 'How progressive should the income tax be?', in M. Parkin (ed.), Essays on Modern Economics, Longman.
- Atkinson, A. B. and A. J. Harrison, (1978), Distribution of Personal Wealth in Britain, Cambridge University Press.
- Atkinson, A. B. and J. E. Stiglitz, (1976), 'The design of tax structure: direct versus indirect taxation', Journal of Public Economics, 6, pp. 55-75.
- Bardhan, P. and T. N. Srinivasan, (1971), 'Crop sharing tenancy in agriculture: a theoretical and empirical analysis', American Economic Review, 61, 1, March, pp. 48-64.
- Bell, C., (1977), 'Alternative theories of sharecropping: some tests using evidence from Northeast India', Journal of Development Studies, 13, 4, July, pp. 317-346.
- Bell, C. and P. Zusman, (1976), 'A bargaining theoretic approach to cropsharing contracts', American Economic Review, 66, 4, September, pp. 578-587.
- Bevan, D. L., (1979), 'Inheritance and the distribution of wealth', Economica, 46, pp. 381-402.
- Bevan, D. L. and J. E. Stiglitz, (1980), 'Intergenerational transfer and inequality', Greek Economic Review, 1, pp. 9-26.
- Blinder, A. S., (1973), 'A model of inherited wealth', Quarterly Journal of Economics, 87, pp. 608-626.
- Blinder, A. S., (1976), 'Inequality and mobility in the distribution of wealth', Kyklos, 29, pp. 607-638.

- Bliss, C. J. and N. H. Stern, (1980), Palanpur, (forthcoming), Oxford University Press.
- Carruth, A., (1980), 'Optimum negative marginal tax rates: a linear tax model appraisal', mimeo, University of Kent.
- Cheung, S. N. S., (1969), The Theory of Share Tenancy, The University of Chicago Press.
- Debreu, G., (1959), Theory of Value, Yale University Press.
- De Tocqueville, A., (1850), Démocratie en Amérique, I, 13th edtn.
- Diamond, P. A., (1970), 'Incidence of an interest income tax', Journal of Economic Theory, 2, pp. 211-224.
- Edgeworth, F. Y., (1897), 'The pure theory of taxation', Economic Journal, 7, pp. 46-70.
- Farrell, M. J., (1970), 'The magnitude of 'rate of growth' effects on aggregate savings', Economic Journal, 80, pp. 873-894.
- Feldstein, M. S., (1973), 'On the optimal progressivity of the income tax', Journal of Public Economics, 2, pp. 357-376.
- Feldstein, M. S., (1974), 'Incidence of a capital income tax in a growing economy with variable savings rates', Review of Economic Studies, 41, pp. 505-513.
- Feldstein, M. S., (1978), 'The welfare cost of capital income taxation', Journal of Political Economy, 86, pp. S29-S51.
- Flemming, J. S., (1976), 'The assessment of the inequality of wealth', in Royal Commission on the Distribution of Income and Wealth: selected evidence submitted for Report Number 1, H.M.S.O.
- Flemming, J. S., (1979), 'The effects of earnings inequality, imperfect capital markets and dynastic altruism on the distribution of wealth in life cycle models', Economica, 46, pp. 363-380.
- Goldberger, A. S., (1979), 'Heritability', Economica, 46, pp. 327-347.
- Hallagan, W., (1978), 'Self-selection by contractual choice and the theory of sharecropping', The Bell Journal of Economics, 9, 2, Autumn, pp. 344-354.
- Harbury, C. D., (1962), 'Inheritance and the distribution of personal wealth in Britain', Economic Journal, 72, pp. 845-868.
- Harbury, C. D. and D.M.W.M. Hitchens, (1976), 'The inheritance of top wealth leavers: some further evidence', Economic Journal, 86, pp. 321-326.

- Harbury, C. D. and P. C. McMahon, (1973), 'Inheritance and the characteristics of top wealth leavers in Britain', Economic Journal, 83, pp. 810-833.
- Heady, E., (1947), 'Economics of farm leasing systems', Journal of Farm Economics, XXIX, August, pp. 659-678.
- Helpman, E. and E. Sadka, (1978), 'The optimal income tax: some comparative statics results', Journal of Public Economics, 9, pp. 383-394.
- Johnson, D. G., (1950), 'Resource allocation under share contracts', Journal of Political Economy, 58, pp. 118-123.
- Laitner, J. P., (1979), 'Bequests, golden-age capital accumulation and government debt', Economica, 46, pp. 403-414.
- Lydall, H. F., (1968), The Structure of Earnings, Oxford University Press.
- Marshall, A., (1920), Principles of Economics, 8th edtn., (1956), Macmillan.
- Meade, J. E., (1964), Efficiency, Equality and the Ownership of Property, George Allen and Unwin.
- Menchik, P. L., (1979), 'Inter-generational transmission of inequality: an empirical study of wealth mobility', Economica, 46, pp. 349-362.
- Mill, J. S., (1848), Principles of Political Economy, J. W. Parker.
- Mirrlees, J. A., (1971), 'An exploration in the theory of optimal income taxation', Review of Economic Studies, 38, pp. 179-208.
- Newbery, D. M. G., (1977), 'Risk sharing, sharecropping and uncertain labour markets', Review of Economic Studies, 44, pp. 585-594.
- Newbery, D. M. G. and J. E. Stiglitz, (1978), 'Sharecropping, risk sharing and the importance of imperfect information', University of Cambridge, Economic theory discussion paper 8.
- Phelps, E. S., (1973), 'Taxation of wage income for economic justice', Quarterly Journal of Economics, 87, pp. 331-354.
- Pryor, F. L., (1973), 'Simulation of the impact of social and economic institutions on the size distribution of income and wealth', American Economic Review, 63, 1, pp. 50-72.
- Rao, C. H. H., (1971), 'Uncertainty, entrepreneurship and sharecropping in India', Journal of Political Economy, 79, May/June, pp. 578-595.

- Rawls, J., (1971), A Theory of Justice, Harvard University Press.
- Rothschild, M. and J. E. Stiglitz, (1976), 'Equilibrium in competitive insurance markets: an essay on the economics of imperfect information', Quarterly Journal of Economics, 90, Symposium on the economics of information, pp. 629-649.
- Sadka, E., (1974), 'Income distribution, incentive effects and optimal income taxation', Ph.D. dissertation, M.I.T., Cambridge, Mass.
- Sandford, C. T., (1971), Taxing Personal Wealth, George Allen and Unwin.
- Sheshinski, E., (1971), 'On the theory of optimal income taxation', discussion paper 172, Harvard Institute for Economic Research.
- Sheshinski, E., (1972), 'The optimal linear income tax', Review of Economic Studies, 38, pp. 179-208.
- Sismondi, J. C. L., Simonde de, (1814), Political Economy, (1966), Augustus M. Kelley.
- Smith, A., (1776), An Inquiry into the Nature and Causes of the Wealth of Nations, R. H. Campbell and A. S. Skinner (eds.), (1976), Clarendon.
- Stern, N. H., (1976), 'On the specification of models of optimum income taxation', Journal of Public Economics, 6, pp. 123-162.
- Stern, N. H., (1979), 'Optimum taxation with errors in administration', University of Warwick discussion paper 157.
- Stiglitz, J. E., (1969), 'Distribution of income and wealth among individuals', Econometrica, 37, pp. 382-397.
- Stiglitz, J. E., (1974), 'Incentives and risk sharing in sharecropping', Review of Economic Studies, 41, April, pp. 219-255.
- Stiglitz, J. E., (1978), 'Notes on estate taxes, redistribution and the concept of balanced growth path incidence', Journal of Political Economy, 86, pp. S137-S150.
- Wedgwood, J., (1929), The Economics of Inheritance, Routledge.
- Yaari, M., (1965), 'Uncertain lifetime, life insurance and the theory of the consumer', Review of Economic Studies, 32, pp. 137-150.