Fiscal Tiers: The Economics of Multi-Level Government

DAVID N. KING
University of Stirling

London
GEORGE ALLEN & UNWIN
Boston Sydney

in association with
The Centre for Research on Federal Financial Relations,
The Australian National University

© David King, 1984
This book is copyright under the Berne Convention. No reproduction without permission. All rights reserved.

George Allen & Unwin (Publishers) Ltd,
40 Museum Street, London WC1A 1LU, UK

George Allen & Unwin (Publishers) Ltd,
Park Lane, Hemel Hempstead, Herts HP2 4TE, UK

Allen & Unwin Inc.,
9 Winchester Terrace, Winchester, Mass. 01890, USA

George Allen & Unwin Australia Pty Ltd,
8 Napier Street, North Sydney, NSW 2060, Australia

First published in 1984

British Library Cataloguing in Publication Data

King, David
Fiscal tiers.
1. Local finance 2. State governments
I. Title
336'.01'091821 HJ9105
ISBN 0–04–336081–5
ISBN 0–04–336084–X Pbk

Library of Congress Cataloging in Publication Data

King, David N. (David Neden).
Fiscal tiers.
Bibliography: p.
Includes indexes.
1. Intergovernmental fiscal relations. I. Title.
HJ197.K56 1984 336.1'85 84–9157
ISBN 0–04–336081–5
ISBN 0–04–336084–X (pbk.)

Set in 10 on 11 point Times by Grove Graphics, Tring, Hertfordshire
and printed in Great Britain
by Butler & Tanner Ltd, Frome and London

THE AUSTRALIAN NATIONAL UNIVERSITY LIBRARY

Contents

Preface *page* vii

Introduction 1

1 The Economic Role of Subcentral Authorities 6

Introduction – The Allocation Function – Public Goods – Output Levels for Public Goods – Local Public Goods – Arguments for Subcentral Provision of Local Public Goods – Arguments against Subcentral Provision of Local Public Goods – Voting-with-the-feet – The Distribution Function – The Stabilisation Function – Regional Development – Subcentral Authority Functions in Practice – Summary – Notes

2 The Optimal Size of Subcentral Authorities 50

Introduction – The Effects of the Major Gains from Collective Provision – The Effect of the Major Loss from Collective Provision – Organisational Aspects – Consumer Mobility – Purchasing Public Services – Structures in Cases of Many Subcentral Services – Public Goods whose Benefits are Confined to Fixed Areas – Impure Local Public Goods – Functions Other than Local Public Goods – Size and Subcentral Revenue – Other Factors Affecting Size – Sizes in Practice – Reforms in the United Kingdom – Summary – Notes

3 Grants: Their Types and Effects 86

Introduction – Types of Grants – The Framework of the Traditional Models of Grant Effects – Lump-sum Grants – Matching Grants – The Effects of Grantor Taxes – General Grants – The Background to the New Models – The Niskanen Model – The Romer and Rosenthal Model – The Oates Model – The Break Model – A Problem with the New Models – Summary – Notes

4 The Purposes of Grants 121

Introduction – Grants to Internalise Externalities – The Effects of Externality Correction on Subcentral Service Provision – Grants to Enforce Grantor Preferences for Merit Goods – Grants to Enforce Grantor Preferences for Minimum Standards – Grants to Enforce Grantor Preferences for Macroeconomic Reasons – Specific Grants in Practice – Grants to Correct for Fiscal Imbalance – The Equity Case for Equalisation Grants – The Efficiency Case for Equalisation Grants – Capitalisation as an Alternative to Equalisation – Benefit Taxes (Type I) as an Alternative to Equalisation – Benefit Taxes (Type II) as an Alternative to Equalisation – Differential Central Tax Rates as an Alternative to Equalisation – Equalisation Grants and Locational Efficiency – Equalisation Grants and Development – Summary – Notes

5 Equalisation and Revenue-Sharing Grant Schemes 164

Introduction – A Scheme to Equalise Fiscal Potential – A Scheme to Equalise Fiscal Capacity – Equalisation of Fiscal Potential and Equalisation of Fiscal Capacity Compared – An Alternative Equalisation Scheme – Measuring Expenditure Requirements – Measuring Resources – Measuring Revenue Effort – The Determination of Aggregate Grants in Effort-related Schemes – Equalisation and Revenue-sharing Grants in Practice – The West German Arrangements – The Canadian Arrangements – The Australian Arrangements – The United States' Arrangements – The United Kingdom's Arrangements – Summary – Notes

6 Taxes 199

Introduction – The Case for a Harmonised Structure of Subcentral Taxes – Taxes and the Sizes of their Yields – The Allocation of Tax Yields to Different Authorities – The Geographical Distribution of Tax Yields – The Elasticity of Tax Yields – Administration and Compliance Costs – Excess Burdens of Subcentral Taxes levied at Uniform Rates – Excess Burdens of Subcentral Taxes levied at Varying Rates – Tax Exporting – Tax Perceptibility – Tax Competition – The Spread of Tax Incidence – Taxes and their Macroeconomic Implications – Taxes and Equity – Taxes in Practice – Tax-sharing in West Germany – Summary – Notes

7 Charges 244

Introduction – The Finance of Subcentral Authority Enterprises – The Finance of Externality Correction and Public Good Provision – The Finance of Stimulating Merit Good Provision – The Finance of Subcentral Stabilisation Policies and Law-making – The Main Arguments in Favour of Increased Charges by Subcentral Authorities – Charging for Education – Charging for Roads – Charging for Housing – Charging for Police Services – Charging for Fire Services – Charging for Refuse Services – Charging for Other Services – The Possibility of Experiments with Charging – Benefit Taxation and Earmarked Revenues – Summary – Notes

8 Loans 270

Introduction – Loan Finance and Inter-temporal Equity – The Effects of Migration on Inter-temporal Equity – Loan Finance and Intra-generational Equity – The Effects of Migration on Intra-generational Equity – Summary – Notes

9 The Impact of Subcentral Government on Central Government Economic Policy 286

Introduction – Central Government Resource Allocation Policy – Central Government Redistribution Policy – Central Government Stabilisation Policy – The Impact of Subcentral Authorities on Demand, Output and Unemployment – The Impact of Subcentral Authorities on Inflation, the Balance of Payments and Growth – Regulating Borrowing by Subcentral Authorities – Regulating Subcentral Taxation and Expenditure – Summary – Notes

Bibliography 305

Name Index 317

Subject Index 320

Preface

In 1982, I spent some months as a visiting fellow at the Australian National University in Canberra at the invitation of the Centre for Research on Federal Financial Relations. The original intention was that I should write a monograph for them that would outline the main aspects of economic theory pertaining to the design and operation of systems of multi-level government, and that would also include some new theoretical developments in this field. It soon became clear, however, that it would need more than a monograph to do justice to all this theory, and that it would be desirable to include some empirical material to give an idea of how far the present practices of a number of countries could be related to it. The present book is the result, and the manuscript was finally completed late in 1983.

I must thank first Professor R. L. Mathews and the Centre for Research on Federal Financial Relations for asking me to Canberra, for enabling me to work in such attractive surroundings with access to excellent libraries, and for their unfailing support of my work up to and including its publication. I would like to thank Professor Mathews also for his numerous helpful comments on my ideas for this book and on many parts of the text. In addition, I must express my appreciation to the University of Stirling for granting me a semester's sabbatical leave to concentrate on this project.

I am indebted to many people for sparing their time to discuss various points with me, particularly Dr B. S. Grewal and the other visiting fellows I met in Canberra, and Mr. R. Shone and many of my other colleagues in the Economics Department at Stirling. I am grateful, too, to Professor C. V. Brown and Mr R. A. Jackman for reading parts of the text and for making a number of helpful suggestions. Naturally, I am fully responsible for any errors that remain.

I owe both thanks and sympathy to all those in Canberra and Stirling who have been involved in typing the first draft of the text and in incorporating innumerable subsequent changes, and I wish particularly to acknowledge my gratitude to Mrs A. Cowie who was responsible for the greater part of this work; her ability to decipher my writing, and to do so at speed, is a remarkable achievement. I would like also to express my appreciation of all the help given by Mrs S. Hewitt in the preparation of the indices and in the production of the diagrams ready for publication.

Lastly, I would like to record my debt to Lady Hicks, who first interested me in this branch of economics at a seminar in Oxford in 1967, and to Professor A. T. Peacock and Professor J. Wiseman, who, in diverse ways, have encouraged me to work in this field in the succeeding years.

To Susan

Introduction

My main intention in this book is to discuss the principal developments in economic theory relating to the design and operation of systems of multi-level government. In part, this discussion has involved bringing together and synthesising the existing material in this field. The list of references at the end of the book gives some idea of how many people have done relevant work, and of how widely their contributions are scattered in the literature. However, the discussion has also involved identifying gaps in the existing theory and attempting to fill some of them myself. Accordingly, a number of new theoretical developments are put forward here for the first time.

It is my hope that this book will be of interest both to economists engaged in research on this topic and to students engaged in a study of it. I hope, too, that it will be of value to politicians and public officials who are concerned with the design, running and oversight of devolved authorities, for I have attempted to concentrate on those parts of the theory that throw some light on the major policy issues that they will confront. Each of the following chapters is concerned with a particular group of questions that will arise from time to time.

Chapter 1 concerns the most basic issue, namely what functions should be performed by lower tiers of government, while Chapter 2 considers the issues of how many tiers of authorities should be established to perform those functions and how large the authorities in any one tier should be. The next three chapters examine grants, which are often a major source of finance for subcentral authorities. Chapter 3 describes the chief types of grant and their likely effects on recipients' spending levels; Chapter 4 looks at the various purposes of grants and the appropriate type of grant for each of those purposes; and Chapter 5 considers in detail the construction of grant schemes concerned with equalisation and revenue-sharing.

There follow three chapters that focus on the main sources of income for subcentral authorities other than grants. Chapter 6 examines the case for a harmonised tax structure for those authorities, and it considers the criteria by which taxes should be judged to see how suitable they are for use by them. Chapter 7 considers the circumstances in which charges would be a suitable form of finance for subcentral government activities, and discusses which services or parts of services that are often provided at subsidised or zero prices by such authorities might be financed to a greater degree by charges. Chapter 8 examines the extent to which these authorities should use loan finance. Finally, Chapter 9 considers the effects that subcentral authorities' combined revenues and expenditures might have on the central government's own economic policies, in particular its

macroeconomic policies, and examines how far subcentral authorities need to be controlled as a result, and also what sorts of controls might be appropriate.

It might be hoped that each of these chapters would end with some detailed conclusions that, taken together, would form a useful outline of an ideal system of devolved government. However, even if it were accepted that economics as a discipline could, in principle, produce some normative results, there are, I believe, major difficulties in trying to produce many here. One problem is that, although the theory brings to light the questions that need to be discussed when designing or operating a system of devolved government, and it indicates the possible advantages and disadvantages of doing things in different ways, the balance of the arguments may well tilt one way in one country and another way elsewhere, and it may move from time to time even within a particular country. Thus an attempt to find a unique definitive system would be absurd. A further problem is that, in this book, it is not possible to do much more than outline the arguments that need to be weighed up and give some indication of how their strengths might be assessed. While I hope this is of some use to policy-makers in all countries, it has to be conceded that much of the book is chiefly concerned with indicating the empirical work that it would be valuable to have done. Actually doing that work would require a series of separate exercises in each country, exercises that should perhaps be repeated periodically to see if the passage of time has altered the results.

It is inevitable that varying circumstances will cause different countries to adopt different practices. I am nevertheless convinced that all countries could learn a great deal by studying the practices of others, and that a detailed comparative study of the functions, structure, revenue arrangements and oversight of devolved governments in each of the world's major democracies would be of great value, but such a study is beyond the scope of the present work. I have, however, felt it worthwhile to include brief outlines of the practices of a number of countries for which there is a wide literature in the English language – namely, Australia, Canada, the United Kingdom, the United States and, as it happens, West Germany.

It is worth considering how far the relevance of the material in this book is affected by the various constitutional arrangements that might apply in different countries. The principal division ordinarily made is between unitary countries, in which the whole structure, powers and operation of subcentral authorities are subject to overall central government supervision, and federal countries, as defined by Wheare (1946, pp. 32–3), in which the central government and at least one tier of subcentral authorities are co-ordinate and, in principle, independent. All the material in this book might be seen as relevant to central governments in unitary states that wished to consider creating or modifying systems of multi-level government; but some of it, at least, might appear irrelevant in federal countries. For instance, it might be asked why anyone in a federal country would be interested in an economic analysis of what powers should be delegated and what sizes of authorities should handle these powers if the constitution prevents the federal government from changing the status quo. Why should they have much interest in the relative merits of different

subcentral taxes, or in charging policies, or in the possibility that subcentral authorities might issue excess debt, if there were no powers over their choice of taxes, their use of charges or their issue of debt? Why should they consider the case for controls to aid central government macroeconomic policies if the independence of the lower-tier authorities was constitutionally guaranteed? Further, given that true independence must entail independence in revenue-raising, why should they consider the whole question of grants?

The principal answer to these questions is that no so-called federal country has a tier of authorities that are as fully independent as Wheare's theoretical view of federalism suggests. In 1946, Wheare (pp. 32–3) himself was able to find only four countries where one tier of authority had a status generally co-ordinate and independent of the central government – namely Australia, Canada, Switzerland and the United States – and he conceded that in none of these were the lower-tier authorities wholly independent in terms of revenue. This, in turn, means that even in these countries there has necessarily long been an interest in grants; and, since grants can be made conditional on recipients' behaviour (or, in other words, since they can be used as a means of control), there is likely to be interest in all the other issues as well, except perhaps whether the states, provinces or cantons are of optimal size. In later editions of his book, Wheare did not feel entitled to extend his list to include any of the more recent so-called federations. This is perhaps a little surprising in the case of West Germany, but less so in the case of the new federations attempted in a number of ex-colonies in Africa, the Far East and the West Indies; of these, Hicks (1978, p. 4) has claimed that some were failures and the others only partial successes.

Moreover, times have changed since 1946. For example, Reagan and Sanzone (1981, p. 4) have argued that, in the United States at least, it no longer makes sense to conceptualise a federal system in which there is a wall separating the state and national levels of government; they argue that the two levels are now interdependent and share functions, a new situation seen most clearly in the explosive growth of federal grants-in-aid. In a similar vein, Hicks (1978, p. viii) has claimed that, in the modern world, federal systems have generally abandoned co-ordinate federalism and moved towards co-operative federalism, in which the relations between the centre and the states are much closer. Foster, Jackman and Perlman (1980, p. 37) go beyond this: they claim that the federal notions that central and subcentral governments are co-ordinate and independent could never be made operational; thus they imply that the wall could never have existed. Oates (1972, p. 18) rescues the term federalism by arguing that all countries are federal, though not in Wheare's sense, and differ only in the degree to which central governments interfere with lower tiers; in his view, constitutional constraints serve only to make central intervention more difficult or more costly.

I believe that there are two further reasons why the issues discussed in this book should concern citizens in so-called federal countries. First, the states in many federal countries have great powers over subordinate local authorities. In their regulatory role, they may well wish to consider issues

such as the independent sources of revenue of these authorities, and the purpose and structure of any grants they may require. They may also wish to reform the size and functions of these subordinate local authorities. Indeed, local government reforms have been made on a substantial scale in recent years in a number of federal countries. In the United States, for example, the number of local authorities fell from over 155,000 in 1942 to under 80,000 in 1977 (United States Department of Commerce, 1981, p. 275), and in West Germany the number fell from over 22,000 in 1970 to under 11,000 in 1976 (ACIR, 1981b, p. 34); in both these countries, local authority boundaries within any state are determined by that state and not by the federal government.

Secondly, federal constitutions do not make change impossible. Thus it may be possible to alter the powers, and even perhaps the boundaries, of the states in a federation, and it may be possible to alter the revenue sources available to them. As far as boundaries are concerned, for example, Article III (Section 3.1) of the Constitution of the United States allows states to merge or to be split up provided their legislatures and Congress approve; in Switzerland, on the other hand, the member states are listed in Article 1 of the Constitution, and mergers or splits require constitutional amendments that need the approval of a majority of the country's citizens, although as recently as 1978 a new Canton of Jura was created out of part of the Canton of Berne. The constitutions of the United States and Switzerland give the states and cantons almost unlimited powers of taxation, although in Switzerland Article 3 of the Constitution allows some tax powers to be explicitly reserved for federal use, and the list of such powers is occasionally modified. In Australia, the tax powers of the states have been changed significantly from time to time.[1] The Constitution gave the states wide tax powers, but it excluded them from customs and excise duties, and in 1942 the federal government was able to pass legislation that effectively prohibited states from raising income taxes. In 1971, however, each state was given access to a payroll tax, and from 1976 each was able to impose a surcharge or give a rebate (at its benefit or cost) on federal income taxes levied in its area, although none has yet done so.

This analysis suggests that all the issues discussed in this book are of importance in any country with one or more levels of subcentral government, irrespective of the degree to which the higher tiers of government interfere with the lower tiers, provided only that the degree exceeds zero, which it always does.

A zero degree of control is but one end of the spectrum of possibilities, however. At the other end there is a situation of total control where the subcentral authorities are, in effect, merely agents of the central government.[2] In practice, subcentral authorities do sometimes act purely as agents of the central government in assisting it to pursue one or other of its policies on a uniform national basis. For example, local authorities in the United Kingdom act effectively as agents in paying maintenance grants to students engaged in higher education, for the eligibility rules and the levels of these grants are determined by the central government, and local authorities are themselves given grants to cover the whole cost of

their grant payments. Another example occurs in Switzerland, where the cantons act as federal agents in the matter of national road construction and are reimbursed for their relevant expenditure (Mathews, 1980a, p. 27).

In general, though, agency activities tend to form a very small part of any subcentral authority's functions; certainly few functions appear to be so clearly those of agents that they would never be considered when subcentral elections take place. Nevertheless, I do not wish to suggest that those who discuss the agency theory of subcentral government are discussing what is virtually a non-issue. These writers are essentially discussing the possibility that, in some countries, increases in central control have meant that subcentral authorities have moved very much closer to the agency end of the spectrum than they used to be. How far such a move is desirable for any particular function of any particular tier of authorities in any particular country is clearly a matter of importance. Unfortunately, it is not possible to discuss individual cases in this book, but I hope that Chapter 1 will help those who might be concerned with such moves by outlining the issues they should bear in mind. Of course, provided some autonomy remains, all the issues examined in the remaining chapters will continue to be relevant.

It is appropriate to indicate two limitations of my own discussions, quite apart from any errors and omissions there may be. First, I have frequently ignored the particular qualifications that may be needed before some parts of the theory can be usefully applied in any particular country; I have no doubt that economists interested in this field will be well aware of the major qualifications that apply to their own countries. Secondly, I should point out that I have assumed that the reader has studied economics for at least one or two years, and that s/he has made some study of public finance; notwithstanding this, I have included a discussion of public goods that starts from first principles because it is my impression that the analyses given in most books rest on assumptions that are more useful in developing elegant theories than in explaining the nature of most of the goods and services that subcentral authorities typically provide.

Notes

1 For further details, see Mathews (in ACIR, 1981c, pp. 6–14).
2 For a discussion of the agency theory of subcentral authorities, see Foster, Jackman and Perlman (1980, pp. 34–7).

1

The Economic Role of Subcentral Authorities

Introduction

This chapter sets out to consider what economic functions should be performed by subcentral authorities. In principle, all the responsibilities of these authorities could be handled by the central government. In a unitary country, therefore, the material in this chapter can be seen as exploring the case for decentralising some responsibilities. Conversely, in a federal country, the chapter can be seen as presenting the case for retaining some responsibilities at a subcentral level.

Most discussion of this topic[1] focuses on three groups of government functions, which are often labelled the allocation, distribution and stabilisation functions. The general consensus seems to be that subcentral authorities may have a considerable part to play in the allocation function (most importantly by providing what have been termed local public goods), that they will have little or no part to play in the distribution function, and that they will have no part at all to play in the stabilisation function. For this reason, the role they play in the allocation function is considered here at greatest length. The discussion of it falls into several parts. First, there is a broad discussion of the allocation function and the various ways in which subcentral authorities might be empowered to participate in it. Then there is an examination of public goods in general, and an outline of the problems of determining the optimal amounts to produce. This is followed by a discussion of local public goods and a consideration of the arguments for and against subcentral provision. Finally, there is a closer look at one of the more frequently debated advantages of subcentral provision – namely, that citizens can vote with their feet to live in the authorities whose expenditure levels and patterns most closely match their preferences. The chapter then continues with discussions of the distribution function and the stabilisation function, particularly in connection with regional development, and looks both at the reasons why subcentral authorities are generally held to have little part to play in these functions and also at the reasons more recently advanced in support of some subcentral participation. The last section of the chapter attempts to give some idea of the range of activities that are undertaken by subcentral authorities in practice.

It should be observed that subcentral authorities may undertake some activities that do not clearly reflect any discretionary powers they may have in respect of resource allocation, distribution or stabilisation. For example, they may perform some functions as agents of the central

government and so have no degree of autonomy, and they may have law-making powers in fields with little or no relevance to economic policy. The case for entrusting these activities to them is not explored in this chapter. As far as agency functions are concerned, it is to be supposed that there is little to say except that they arise principally as a matter of administrative convenience. As for law-making functions, it is likely that discussions of the advantages and disadvantages of decentralisation would cover much of the ground discussed in a similar connection with local public goods, although this would be an interesting topic for future analysis. It is perhaps worth pointing out that there is little of relevance to agency functions or law-making powers in the following chapters. The former are likely to be financed by specific lump-sum grants from the central government and the latter are likely to cost little, so there is not really anything in the discussion of grants, taxes, charges or loans in Chapters 3–8 to add. Moreover, the lack of autonomy in agency functions should mean that they cannot be performed in ways that disturb the central government's macroeconomic policies, the issue of Chapter 9, and these policies are seldom likely to be upset by decentralised law-making. However, both types of function need to be considered when deciding the optimal sizes for subcentral authorities, so they are borne in mind in Chapter 2.

It should be observed, too, that in attempting to delineate the role of subcentral authorities it is necessarily assumed that this role is a variable. In so-called federal countries, this assumption may in practice be questionably valid. Whatever the merits of allocating some currently subcentral functions to the central government, the states (provinces or cantons) may resist attempts to reduce their powers. In such instances, there is a particularly strong case for developing co-operation between the centre and the states in formulating policies.[2]

The allocation function

It is a major proposition of welfare economics that a perfectly competitive economy will, under certain assumptions, secure a Pareto-optimal or Pareto-efficient allocation of resources; that is to say, it will produce a situation in which it is not possible to make one person better off without making at least one other person worse off. A key feature of this situation is that each product's price will equal its marginal cost. It is probably fair to say that the governments in most Western economies interfere with the allocation of resources chiefly in situations where the market is far from perfectly competitive, or where the necessary assumptions are not met. Interference of the first type will occur most particularly in cases of national or local monopolies (or oligopolies); in these cases, central or subcentral governments may seek to regulate the activities of private producers, or they may seek to nationalise or municipalise production and operate it themselves. Interference of the second type may occur, for example, when the production or consumption of an item produces external costs or benefits to other producers or consumers, in which case the appropriate

action may be to place a tax or subsidy on the item or to control the level of externality production in some other way. There are closely related occasions when an item possesses one or both of the properties of public goods, in which case the appropriate action may be to undertake public provision. This chapter looks closely at the role of subcentral authorities in providing local public goods, and this issue is considered in detail in the following six sections.

By way of contrast, little will be said about subcentral authorities interfering with local monopolies. There are really two separate issues here. First, there is the question of whether intervention should take the form of regulation or public ownership. Secondly, there is the question of whether this intervention should be handled by the central or subcentral authorities. The first issue is not considered here because it is probable that decisions will be made in each case on the basis of criteria that make little explicit reference to the existence or operation of subcentral government; such criteria are outside the scope of this work. As for the second issue, it is probable that the arguments for and against subcentral responsibility for intervention in local monopolies are similar to those for and against the provision of local public goods, but there is perhaps scope for more research in this area.

This chapter also gives little attention to the possibility of a role for subcentral authorities in handling problems created by externalities. In this respect, it echoes the literature on subcentral government, which also says little about this issue, except perhaps to cite alleged external benefits of services such as education and refuse collection as one reason for public provision, or to cite the fact that motorists in certain areas do not face the full costs of their journeys (since they do not have to pay anything towards the congestion costs they impose on others) as a reason for subcentral (or central) authorities subsidising urban public transport to make the relative prices of public and private transport reflect their relative costs. This feature of the literature arises, perhaps, because the correction of externalities is generally seen as a matter of devising taxes, subsidies or controls that will result in production occurring where each item's price equals its marginal social cost, and it may be felt that there is little point in entrusting to subcentral authorities a function that allows little scope for discretion.

It could be argued, however, that some external effects are of primary concern to local citizens, and that in such cases subcentral authorities are the appropriate ones to deal with them because they may have more information on local opinions and hence a clearer idea of the value of the external effects. This line of reasoning might be the chief reason why subcentral authorities are so often given powers over land use and controls over new buildings, and why they sometimes have the authority to control the discharge of effluents. Moreover, subcentral authorities may not only have varying views on the need for controlling externalities, they could also have views on the appropriate policy instruments to use in their areas, and, if subsidies or taxes were used, then they could have views on how these should be financed or spent. It appears, therefore, that there is scope for further theoretical analysis of the case for some decentralisation of

this function. It is likely, though, that such an analysis would incorporate many of the arguments raised in discussions of the case for decentralising the provision of local public goods. For the present text, it will generally be assumed that, when subcentral authorities do handle externalities, they do so either directly by regulations, which can be seen as a form of law-making, or indirectly by providing services, such as refuse collection.

It has been observed by Musgrave and Musgrave (1976, pp. 65–6) that not all interferences in resource allocation by governments can be explained in terms of regulating imperfect markets, of arranging for externalities to be taken into account or of supplying public goods. In addition, the state has a tendency to encourage – sometimes perhaps by supplying at subsidised or zero prices or by stimulating private provision – the consumption of items that it would like to see consumed in greater quantities, items that are termed merit goods; likewise, it has a tendency to discourage – perhaps by taxes or laws – the consumption of items that it would like to see consumed in smaller quantities, items that may be termed demerit goods.

There seem to be three possible reasons for the provision of merit goods (such as low-cost housing or subsidised school lunches). First, the purchasers may have low incomes. In this case, though, the provision is essentially part of an income redistribution policy rather than a resource allocation one. Secondly, the purchasers may not be the consumers, as, for example, would be the case with school education bought by parents on behalf of their children. This situation might well lead to some parents buying very little education. This problem is no doubt overcome more effectively with the normal practice of having laws to make education compulsory up to a certain age than it would be by subsidising education, for even zero price schooling would not necessarily tempt all parents, as some might prefer their children to be gainfully employed. Thirdly, as Allan (1971, p. 15) points out, the consumers may be ignorant of the benefits of certain goods and services, unaware, perhaps, that some foods are much better for them than others. Governments could generally try to improve knowledge in such cases by publicity campaigns.

Conversely, demerit goods (such as alcohol, certain drugs and tobacco) are generally considered to be consumed in excess by consumers solely because they do not fully appreciate their effects or do not take the effects fully into account. Consumption may be discouraged by taxes or laws.

This analysis of merit goods and demerit goods suggests that, as far as pure resource allocation policy (as opposed to income redistribution policy) is concerned, the most likely government activities are regulations, publicity campaigns and selective taxes. If several tiers of government exist, then there may be a case for giving the responsibility for some of these activities to lower tiers; the arguments for and against devolving such functions will be similar to some of the arguments applying to local public goods, which are discussed shortly. However, Musgrave and Musgrave (1976, p. 65) note that the provision of merit goods is frequently directed at the poor, and it is probable that their distribution aspects are really much the more important. The role of subcentral authorities in the distribution function is also considered later in this chapter.

Public goods

It was noted earlier that most discussions of the allocation role of subcentral authorities concern their part in providing public goods. Before discussing this role, it is necessary to consider the nature of public goods themselves in some detail. The need for this detailed discussion stems from the fact that it implies that there is a case for the public provision of a greater range of items than those that fully meet the strict criteria generally used for defining public goods. It is therefore useful to examine those criteria, and then to see how far they need to be met by any item for there to be a public good type of argument in favour of public provision of that item.

A public good is defined strictly as one exhibiting two properties. First, its consumption is non-rival. Consider an item such as an apple. The apple (or at least any particular part of it) can be consumed by only one person. If it is consumed by Mr X, then it cannot be consumed by Mrs Y; accordingly, their consumption of apples is held to be rival. In contrast, consider an item such as defence. If a nation provides a defence service, then the benefits of that service may be available to all citizens, including Mr X and Mrs Y, and their consumption will be non-rival. Secondly, consumption of a public good is held to be non-excludable; that is to say, producers cannot prevent benefits accruing to consumers who refuse to pay. In the case of an apple, for instance, a greengrocer can refuse to supply would-be consumers who refuse to pay, and hence non-payers can be excluded from consumption. In the case of defence, though, an entrepreneur who attempted to finance an army from voluntary contributions would find that the services he provided (in the form, perhaps, of keeping a foreign invasion at bay) would be consumed by all citizens, whether or not they had paid contributions, so that non-payers could not be excluded from consumption.

A good that exhibits both rivalry and excludability is termed a private good, and, in general, the free market is well suited for providing such goods. Problems arise, however, with goods (or, as is more usually the case, services) that exhibit either non-rivalry or non-excludability or both. Consider, for example, a radio broadcast, which is inherently non-rival but which could be made excludable by transmitting 'scrambled' signals and renting 'unscramblers' to listeners. The problem here is that such exclusion is inefficient, for it means that the price of listening is positive in a situation where the marginal cost of providing the service to an extra listener is zero. Private priced production may occur in these cases, but only free (here defined as zero-priced) provision, presumably financed by taxes, would be efficient. Some people might argue that free public provision is also fairer or more equitable in the sense that payments may be related to ability to pay, whilst others might argue that it is less equitable in the sense that it requires most or all citizens to pay taxes to finance services, whereas priced private provision confines payments to beneficiaries alone and would relate the individual payments to the individual levels of consumption. A final decision on the type of provision

should take both efficiency and equity considerations into account.

Next, consider a case such as farmers who could grow plants to provide nectar for the honey production of local bee-keepers. The nectar is certainly rival, for once taken by one bee it is not available to others, but exclusion is impossible since the farmer cannot in practice drive off the bees of any particular keepers who refuse to pay him for nectar. The farmer may ask the keepers for a donation, but many may decide to give him nothing and become 'free-riders'. In principle, the consequence could be that, with private production, farmers would provide few, if any, nectar-bearing crops; zero-price provision financed through taxes could be much more efficient. It is not inconceivable, though, that an equity case could be made for finance by voluntary contributions; for example, it might be found that contributions were made chiefly by beneficiaries, whereas tax finance might exact payments from everyone.

Finally, consider a service such as defence, which exhibits both non-rivalry and non-excludability. The principal difficulty with private provision will stem from the non-excludable property because an entrepreneur seeking to finance his provision through voluntary contributions may secure little in the way of funds if consumers attempt to be free-riders. Of course, reliance on voluntary contributions does not mean that underprovision is inevitable. For example, this source of funds seems to have been largely sufficient to meet the costs of cleaning the exterior of some English cathedrals, an activity whose benefits would seem to be essentially non-rival and non-excludable. Nevertheless, in cases where underprovision seems a real threat, free provision financed by taxes is likely to be called for, though equity considerations might result in the calls being ignored.

Many discussions of public goods suggest that the case for budget-financed provision is confined to items that possess both non-rival and non-excludable

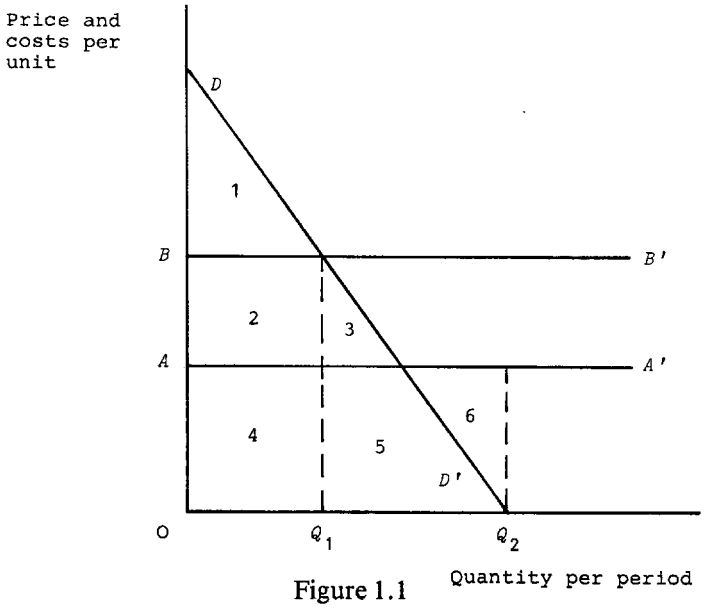

Figure 1.1

qualities. It can be seen, however, that the case exists for any product that displays just one of these properties. Moreover, if exclusion were costly, then a case could be made for budgetary provision of a good exhibiting neither property. It might, for instance, be possible to charge householders for water supply by installing meters in all houses and employing a number of meter readers, but the costs of such an exercise might well exceed the benefits, in which case free provision and budgetary finance might again be called for. Figure 1.1 shows more precisely the factors involved in deciding whether private priced provision or public free provision should be used in the case of a particular item (X).[3] The market demand curve for X is shown by DD' and the unit production costs (assumed for convenience to be constant) by AA'.

Now, if X were to be supplied privately and priced, then suppliers would want to cover exclusion costs as well as production costs. Unit exclusion costs (also assumed for convenience to be constant) are shown by the gap between AA' and BB', so that BB' represents total unit costs. It will be supposed for the present discussion that private suppliers would supply the quantity OQ_1 at a price OB. This situation would occur if the suppliers were in perfect competition with each firm having constant (total) marginal costs of OB, for in that case BB' would represent the industry's supply curve and the equilibrium position would be where that curve intersected the demand curve DD'. It would occur, also, if there was a single private monopolist subject to a regulation requiring it to adopt marginal cost pricing, for such a firm would set its output and price where its (total) marginal cost curve (BB') intersected its demand curve (DD'). With a quantity OQ_1 and a price OB, the total consumer benefits would be represented by area $(1+2+4)$ and the total costs of production and exclusion by area $(2+4)$, so that the net social benefit (assuming there were no external costs) would be represented by area 1.

If, instead, X were to be supplied publicly free of charge, then it will be supposed that the quantity OQ_2 would be supplied. This assumption means that some units, those to the right of the intersection of AA' and DD', are valued by consumers at less than their marginal cost, If, however, the authorities supplied less than this quantity, then there would be a permanent situation of excess demand, and it is probable that the authorities would wish to avoid the problems associated with such situations. With a quantity of OQ_2 supplied free, the total consumer benefits would be represented by area $(1+2+3+4+5)$ and the total costs of production by area $(4+5+6)$. There would be no exclusion costs, so (again assuming no external costs) the production costs would equal social costs and the net social benefit would be represented by area $(1+2+3-6)$. It follows that there would be a case for priced private provision if area $(2+3)$ were smaller than area 6, and for free public provision if it were larger. Not surprisingly, the case for public provision is stronger where exclusion costs are relatively high. More surprisingly, perhaps, the case is also stronger if, as in Figure 1.1, demand is relatively inelastic than if, as in the analogous Figure 1.2, it is relatively elastic. These results arise because high exclusion costs increase areas 2 and 3 whilst inelastic demand reduces area 6.

It may be noted that, if private production were in the hands of an unregulated private monopolist, then output would be below OQ_1 and

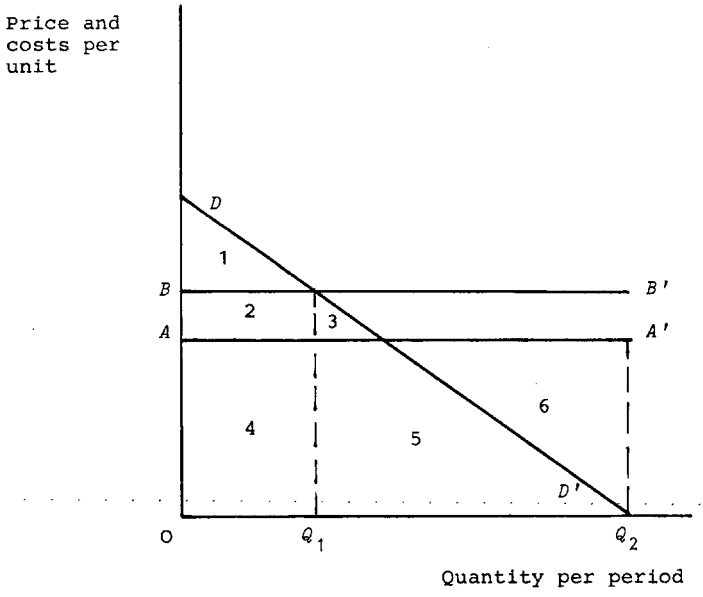

Figure 1.2

price above O*B*, so that the net social benefit (which in this case would include both producer's and consumers' surpluses) would be smaller than area 1. It is possible that the net benefits from free public provision could exceed, or fall short of, those attainable from both an unregulated monopoly and a situation of regulated monopoly (or competitive supply). However, it is also possible that, although the net benefits attainable from free public provision might exceed those arising from an existing unregulated monopoly, still higher net benefits might emerge if that monopoly were instead subjected to regulation. It should be noted, of course, that, while these considerations of net benefits suggest instances where there is an efficiency case for free public provision as opposed to any form of priced private provision, the different methods of provision have different distributional implications. Free public provision must be financed by taxes, which may be related to ability to pay but will probably not be related to usage of the service concerned, while priced private provision will result in citizens paying sums that are related to usage but not necessarily to ability to pay. Once again, final decisions on the choice of provision must take both equity and efficiency considerations into account.

It can be seen that governments are likely to consider providing goods that exhibit both non-rivalry and non-excludability, or just one property (though non-excludability is rare for rival goods) or even, perhaps, neither property. For the purposes of this text, a public good will generally be taken as one where a decision in favour of budget-financed provision has been made, even if it is not a public good in the strict sense of exhibiting both non-rivalry and non-excludability. For the moment, it will be assumed that such goods are provided on a nationwide basis by the central government, and that they are intended to benefit the nation's citizens and no one else.

Output levels for public goods

A major problem for governments providing public goods is determining the Pareto-optimal quantity to produce. The optimality conditions differ from those for private goods. In a world of private goods, the optimal quantity of one good, say X, is the one at which the marginal rate of transformation between X and any other good, say Y (that is MRT_{XY}) equals the marginal rate of substitution between X and Y (MRS_{XY}) for any consumer. With public goods, Samuelson (1954) has shown that the optimal quantity of a public good, P, occurs when $\Sigma\, MRS_{PY}$ for all citizens equals MRT_{PY};[4] in essence, extra units of P are worth producing up to the point where their marginal cost in terms of Y forgone (MRT_{PY}) precisely equals the sum of the marginal benefits that all users simultaneously gain in terms of Y ($\Sigma\, MRS_{PY}$). A somewhat simpler optimality condition is often encountered in the literature, and is used for convenience from time to time in this text. This is that the optimal level of output of P occurs when the sum of the citizens' marginal valuations for P, $\Sigma\, MV_P$ (or the sum of the marginal benefits they derive from P, $\Sigma\, MB_P$) equals the marginal cost of P, MC_P; strictly speaking, this alternative formulation applies only when the price of each private good or service equals its marginal cost, as would be the case if such items were supplied by perfectly competitive industries.

The problem for the policy-makers is that ascertaining the optimal quantity requires a knowledge of each consumer's preferences for P. If the prices for private goods and services equalled their respective marginal costs, then it would help governments greatly if consumers would agree to pay a unit price for P equal to the marginal benefit they derive from its production; if the total price so offered (reflecting the total marginal valuation) exceeded the marginal cost, then it would be clear that a suboptimal amount was being provided and vice versa. However, such an arrangement of marginal benefit charging or taxing is most unlikely to work in practice because, in Samuelson's words (1954, pp. 388–9), 'it is in the selfish interest of each person to give *false* signals, to pretend to have less interest in a given collection consumption activity than he really has'; a reduction in his budgetary contribution will significantly increase his spending power on other items while having no perceptible effect on the level of P provided. The problem for governments is compounded by the fact that, even if a marginal benefit pricing scheme were to work and help to determine the quantity of P to provide, it might not appeal on equity grounds. For example, there is no *a priori* reason for wishing to charge in relation to marginal benefits rather than total benefits; after all, a citizen with a relatively inelastic demand for P might receive a higher total benefit and yet have a lower marginal benefit than one with a relatively elastic demand. In addition, a government might prefer to have charges that bear no explicit relation to benefits but that instead relate to the ability to pay of each citizen. Another difficulty is that different financing schemes would leave individuals with different disposable incomes and hence alter their demands for all items, including P, so that the valuations of the marginal benefits from P would vary and hence the optimal quantity of P would vary. There would, in fact, be a different optimal quantity of P for each possible financing scheme, and the government would need to apply

a social welfare function to decide which of these quantities and schemes it preferred.[5]

The analysis of public good provision can be developed further by distinguishing between pure public goods and impure public goods. A pure public good may be defined as one from which all consumers derive equal benefits and with which the benefit each consumer derives does not alter when the number of consumers changes. Unequal benefits arise when consumers make different amounts of use of a public good, either as a result of choice, as might occur with the use of roads, or as a result of their distance from the public good, as might occur with a park. In such cases, the public goods should still be produced up to the point where the total marginal benefits from providing an extra unit of road or park equal its marginal cost; this time, though, many consumers might use the extra unit little, if at all, so that they might derive little or no marginal benefit from it. In cases such as the park, the authorities should note also that the total marginal benefit may vary according to the location of the extra unit of good.

In cases where congestion occurs, the benefits derived by individual consumers will diminish when the number of consumers rises. A motorway network forms an obvious example; in this case, the item concerned is not wholly non-rival in consumption. Once again, however, the optimality condition requires further units to be provided up to the point where the total marginal benefits equal marginal cost.

It follows from this discussion that the preferences of citizens for any public good are unlikely to be revealed by any process that attempts to link them to individual contributions. Samuelson therefore concluded (1954, p. 388) that they would have to be ascertained through a mechanism such as the voting system, though citizens might also use such activities as lobbying and joining pressure groups. Traditionally, the analysis of such non-market decision-making has been the province of political science rather than economics, although recent years have witnessed the growth of public choice theory,[6] which uses economic methodology to study the same field. This theory is generally beyond the scope of the present discussion, but as Musgrave and Musgrave (1976, p. 56) have noted, the political mechanism is imperfect, and its results are likely to lead only approximately to the optimal budget choice. Their conclusion may need to be modified in time, for work has been done recently on new processes for eliciting preferences on public goods;[7] the processes so far developed have weaknesses, but it is possible that, in time, they will pave the way for one that is satisfactory in both theory and practice.

At present, it seems that the best that can be hoped for is that governments will be able to interpret citizens' voting and other activities sufficiently well to provide approximately the optimal quantities of public goods. Most of the analysis discussed in this volume assumes that this will be the case. However, some theoretical arguments considered later are based on the more explicit assumption that governments will seek to satisfy the preferences of their median voters. It is perhaps useful to make four observations about this assumption.

First, it can be shown (see Mueller, 1979, pp. 98–105) that in a

representative democracy, under a number of strict assumptions, the candidate (or party) whose policies most closely coincide with those of the median voter is likely to win, and therefore each candidate has an incentive to put forward policies favoured by the median voter. It should be noted, however, that the assumptions needed to ensure this result include the assumption that voting takes place on one issue about which preferences vary in a single dimension, and where preferences are unimodal and symmetrically distributed.

Secondly, there is some empirical support for the median voter hypothesis as an explanation of subcentral authority expenditures. Until the mid-1960s, empirical analysis of such expenditures sought to explain them in relation to factors such as urbanisation indices, population densities and mean community incomes, but, following Barr and Davis (1966) and Davis and Haines (1966), a number of studies have included the median voter's income and his tax-price, that is the tax cost to him of each unit of local public goods. The ratio of the median voter's income to the mean income is often similar in many authorities of a given country, which makes it hard to see which is more important in determining expenditure levels. Some interesting light has been thrown on this question by Pommerehne (1978, pp. 262 and 275) who studied 110 Swiss municipalities and found that the median voter's income performed significantly better than the mean income in explaining variations in expenditure levels. However, this superiority was more marked in municipalities employing the direct democracy method of town hall meetings to make decisions than in those using elections to choose representatives.

Thirdly, it should be observed that, whatever its theoretical or empirical support, the median voter assumption is often used in analyses out of convenience rather than necessity. Many of the results based on it would be little altered, if at all, if rather different assumptions were used. Fourthly, it is possible that the median voter's preferences over the levels of provision of public goods may frequently be close to the optimal levels; accordingly, those parts of the following analysis that assume that levels are close to the optimal ones give very reasonable predictions for any authorities seeking to provide the levels desired by their median voters.

However, very different levels of public goods provision may emerge if it is supposed that governments do not try to maximise the welfare of their citizens by seeking to produce optimal levels of public goods, or if they are not constrained by an efficient electoral process into meeting the preferences of the median voter. Instead, the politicians or bureaucrats may be more interested in their own welfare and this may be enhanced by high, super-optimal budget levels. There has been an extensive discussion of this possibility in the public choice literature. A number of the relevant major models of government behaviour are discussed in detail in Chapter 3 to see how they suggest that authorities would respond to grant receipts; it is shown there that, in fact, they do not square fully with the empirical evidence on the effects of grants, although this does not, of course, mean that they are wholly invalid.

Local public goods

It has been assumed so far that all public goods and services are provided by the central government for the benefit of the nation's citizens (though some citizens may derive no benefit from some impure public goods) and that no benefits accrue to citizens outside the nation. However, the issues of what levels and sizes of jurisdiction might provide public goods must now be raised. The presence or absence of external benefits to citizens in other jurisdictions is an important aspect of this discussion.

The simplest cases, in principle, are public goods or services where, for technological reasons, the benefits must accrue wholly to a precisely defined geographical area. A flood prevention scheme in a coastal area might be an example of this type of situation. Such cases are sometimes referred to as local public goods, though it is quite possible that such 'local' public goods could benefit citizens in more than one country. For example, a river may flow through several countries and frequently burst its banks in each; if the country where the river rises were to build a dam, perhaps to control floods in its own area, then technological factors might mean that benefits in fact accrued to a precisely defined geographical area that stretched over several countries. Public goods where technological factors precisely determine the area of benefits are rare; when one occurs, then there is much to be said for establishing a subcentral (or supranational) authority coinciding with that area to provide it, though this issue is discussed in more detail in Chapter 2. It may be observed that cases where the benefits from a service are precisely confined to a single nation are rarer still, because national boundaries are seldom drawn with technological considerations in mind; a possible example would be the attempts of a nation that was wholly confined to one island and free from rabies to prevent the introduction of rabid animals.

In most cases, the area over which a public good or service confers benefits is variable. It is often suggested that some items, such as defence, always confer benefits over a country as a whole, but this is not necessarily the case; for example, a nation might, in practice, be unable to defend the citizens on one of its distant offshore islands. Moreover, there is nothing technologically impossible about a situation in which each district of a country provided its own defence forces.

This means that determining which jurisdiction should provide any public good or service is an issue that needs detailed consideration in virtually every case. The issue can be divided into two sub-issues. First, it must be decided whether the service should be provided by a central or subcentral authority. Secondly, if there is to be subcentral provision, then it must be decided what size of subcentral authority should provide it.

One factor in analysing these issues will be how far the provision of the item by one area confers external benefits on the citizens in other areas. In some cases, such as fire protection or street lighting, the benefits may be largely confined to local citizens; in other cases, such as the defence of a major port or the provision of a stretch of motorway, substantial parts of the benefits may accrue to non-local citizens. The greater the

proportion of benefits that accrues to non-local citizens, the stronger the case for central provision, since subcentral authorities would be likely to ignore external benefits and underprovide. In practice, locally provided services almost invariably confer some external benefits, so underprovision is always a risk. However, as the later sections of this chapter show, there may be advantages in subcentral provision that offset this hazard, and, as Chapter 4 shows, grants from the central government can be used in attempts to correct for underprovision,[8] although the shortcomings of those attempts might be serious in cases where external benefits were substantial. Of course, the provision of grants gives the central government some reason to influence the operation of the service. When grants are large, this influence might be dominating and mean that, in effect, the service was centrally controlled.

Another factor relates to technological aspects of the service. In the case of items such as fire protection and street lighting, the provision of an effective service to all the nation's citizens means there must be fire engines and street lamps in many locations. In these cases, it is clearly operationally feasible to hand over control to one or more tiers of authorities, each tier's authorities covering the nation between them. In other cases, such as defence, the most cost-effective service might be provided by concentrating factors of production in a few strategically chosen locations, possibly in a few offshore nuclear submarines and a modest number of barracks and airfields, and it might well be best to ensure that the activities of all these locations were centrally co-ordinated. In these cases, subcentral control is not really feasible.

For the purpose of further analysis, a division will be made between local public goods – taken here to be those where subcentral provision of an efficient service to all the nation's citizens is technologically feasible and where most of the benefits from provision in any one area would be confined to local citizens – and national public goods, where one or other of these conditions does not hold. The case for central provision of national public goods seems powerful. What needs to be considered is the case for subcentral provision of local public goods. This will be examined in the following three sections of this chapter. If, for some services, a decision were made in favour of subcentral provision, then the question of the optimal size of jurisdictions for the services would arise. That issue is discussed fully in Chapter 2.

It should be noted that if subcentral authorities were to provide some public goods then they would face the problem of determining the optimal quantities to produce. The optimality conditions will be the same as those discussed above. However, the marginal benefits to be included comprise those of local residents and, where applicable, citizens elsewhere enjoying spillover benefits. As noted earlier, grants from higher-tier authorities may be used to encourage lower-tier authorities to take these external benefits into account.

There is one further aspect of local public goods that needs to be mentioned. It is common in the literature to discuss examples of local public goods such as streets, fire protection, police services or education, and some of these have been used in the present chapter. These examples

reflect the major services provided in practice by subcentral authorities. However, they give a vastly oversimplified view of reality. Public fire protection services, for example, typically comprise several subservices, and yet between them they form only a part of any area's fire protection arrangements. The public authorities may, for instance, provide fire engines and crews, fire hydrants, building regulations and education in fire protection, while the private sector may provide night watchmen, insurance company regulations and fire alarms.[9] Thus, rather than deciding on the optimal size of authority to provide fire services, the correct procedure is to identify those parts of fire protection that need public provision, and then to estimate the optimal size of authority for each of those parts. At first, it may seem convenient to give all the public aspects to a single tier, but, as the next section of this chapter will note, there may be advantages in not doing so, apart from the fact that each aspect alone might optimally require a different sized authority. There is a further value in recognising that local services are, in reality, conglomerations of subservices, for it may be that some parts of services that are now locally provided are not really public goods at all, and that they could either be privatised or, at least, efficiently financed by charges rather than taxes; this issue is discussed in Chapter 7.

While it would not be surprising to find that parts of some actual subcentral government services are not really public goods, it is perhaps surprising that some research has led a number of people to question whether entire services such as police services, fire protection or parks are really public goods. The issue started, perhaps, with some work by Borcherding and Deacon (1972), who estimated the degree of divisibility of consumption, α, for a number of subcentral services provided in states in the United States in 1962. The coefficient α was estimated for any service from the equation $Q = X/N^{\alpha}$, where X was the physical output of the service, Q was the amount of the service 'captured by the median voter' and N was the state's population (p. 897). In essence, if $\alpha = 0$, then $Q = X$, so that the median voter is able to capture every unit of the service supplied, a situation indicating that the service is a public good. Conversely, if $\alpha = 1$, then $Q = X/N$, suggesting that the median voter captures no more units than he might expect with a private good.

The Borcherding and Deacon estimates for α were produced as a by-product of a regression study concerned with estimating the demand for public services in the United States. For these purposes, all local data were aggregated at the state level, and the resulting estimates of α were between 1.05 and 1.08 for police services, fire services and parks and recreation (p. 897).[10] Bergstrom and Goodman (1973) conducted very similar analyses for ten separate states in 1962 using data for a total of 826 municipalities with populations between 10,000 and 150,000. They produced α values for police services ranging from 0.69 to 1.48, for parks and recreation ranging from 1.00 to 3.63, and for annual general expenditures of municipalities (excluding education and welfare) ranging from 0.65 to 1.50 (pp. 288–9). Individual state values of α were frequently near to one, and suggested in turn that there might be no increasing returns to scale in the provision of the municipal services under study.

These results give some support to the proposition that the services concerned are not really public goods or, at least, are quasi-public goods. However, Bergstrom and Goodman suggested that economies of scale, and hence evidence of publicness, might be found for units of smaller size than the 10,000 lower limit used by them. McMillan, Wilson and Arthur (1981) examined this hypothesis by using a similar model to estimate α values for different groups of municipalities in Ontario in 1976. They found values of 0.68 and 1.15 respectively for towns and cities with more than one fire station, and a value of 0.30 for municipalities with less than 10,000 people and just one fire station (p. 606). Their results suggest that if publicness and economies of scale do exist, then they are more pronounced at the level of the individual facilities. They gave this contention further support with evidence on recreation and cultural services. These results also suggest a possible explanation for the observations of Foster *et al.* (1980, p. 569), who noted that there had been many studies on economies of scale in the provision of local government services in the United Kingdom and who felt that, broadly speaking, these studies produced little satisfactory evidence that such economies exist. The work of McMillan *et al.* suggests that most, if not all, of the authorities examined in these studies may have been more than large enough fully to exploit any available scale economies so that costs per head would not be higher in the smaller authorities.

If the results of McMillan *et al.* are typical for many subcentral services, then there could be important implications for the optimal size of local authorities. It is concluded in the next chapter that the optimal size for any service may well be below the level at which economies of scale are exhausted; if that is, in turn, below the population that typically supports a single fire station (or other facility) then it could be much smaller than many existing authorities.

Arguments for subcentral provision of local public goods

In the last section, it was shown that it would be feasible for local public goods to be provided by one or more tiers of subcentral authorities. This does not necessarily mean that subcentral provision is preferable to central provision; the cases for and against decentralisation of these functions must now be considered. One of the principal arguments in favour of subcentral provision is that citizens who are dissatisfied with the pattern of services in their own areas can, if they so wish, move to areas where the pattern suits them better; however, this possibility of migration also leads to some problems with subcentral provision. These benefits and costs of migration are discussed in a later section. The present section examines in turn the other main advantages of subcentral provision, while the following section examines the disadvantages.

Perhaps the most commonly claimed advantage of subcentral provision is that subcentral authorities are likely to produce services whose quality and quantity vary from area to area in accordance with the varying tastes of citizens in different areas, whereas, if those services were provided by

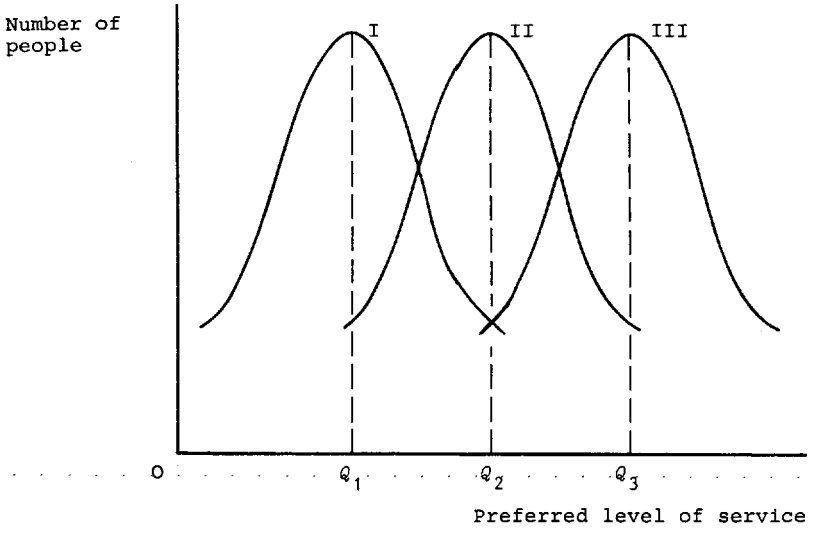

Figure 1.3

the central government, then the most probable result would be uniform provision in all areas. The central government is liable to create uniform provision partly for administrative convenience, and partly, also, because it is likely to levy taxes at uniform rates in all areas and therefore feel obliged to maintain visibly equal levels of provision in each area; further, it may have no explicit methods of ascertaining the extent to which tastes vary from area to area. By contrast, there is no particular administrative convenience for subcentral authorities in providing the same service levels as their neighbours, and they will (if they have some taxes of their own) be able to adjust their taxes upwards or downwards if local citizens demand more costly or cheaper services; also, their own local elections will be one way of ascertaining local preferences.

Figure 1.3 provides an illustration of this argument for subcentral provision. In a country with, say, five areas, it is possible that the distribution of preferences for the provision of a particular public good is shown by curve I for one area, curve II for three areas, and curve III for the remaining area. Central provision is likely to result in a level around OQ_2 being provided in each area, whilst local provision could result in a level around OQ_1 being produced in the area whose preferences are indicated by curve I, a level around OQ_3 in the area represented by curve III, and a level around OQ_2 elsewhere. As Oates (1972, pp. 11–12) has noted, economic efficiency is attained by providing the mix of output that best reflects the preferences of the individuals who make up society; if all individuals are compelled to consume the same level of output of a local public good or service when variations in consumption among different subsets of the population are possible, then the result is an inefficient allocation of resources, unless, of course, subcentral provision carries with it disadvantages outweighing this gain.

It is sometimes argued that a central government need not insist on uniform provision, and it has been claimed (for instance by Foster *et al.*, 1980, p. 577) that no government secures a majority most efficiently by providing exactly the same services for everyone where differentiation is possible. Certainly, a central government might, in principle, provide levels of OQ_1, OQ_2 and OQ_3 in different areas as appropriate, provided it could levy taxes at geographically discriminatory rates and had explicit means of ascertaining preferences in each area. This possibility raises two conceptual problems, however. First, it is not clear that it is very different from a system of local government; it is certainly much closer to that than it is to any existing system of central government. Secondly, it is hard to see how the system would work in the normal context of party politics, for the parties would have to claim at election time that their sole aim for local issues was to meet local wishes; they could not, for example, express desires to expand school education and contract housing subsidies since those policies might not be welcome in all areas. Thus, before abandoning subcentral authorities in favour of a central government operating non-uniform policies, it is necessary to show that a system of parliamentary democracy could be devised in which elections were not fought on policies. It is not clear that this demonstration has yet been made.

In addition to catering for geographical variations in taste, there are four further reasons why it can be argued that subcentral provision will generally be more in accordance with electors' wishes. First, it is possible that, in a system of multi-level government, politicians, particularly those in the lower tiers, are likely to have a greater understanding of electors' wishes than would be the case with a fully centralised system. There are a number of reasons for this. For example, a multi-level system means that individual citizens will have more opportunities to signal their preferences in terms of voting frequency, and more elected representives to lobby; indeed, it has been suggested that demands for some system of devolved government in Great Britain have been fuelled by desires to change the signalling system there (see Peacock, 1977, p. 51). Also, because any given election will be fought on a narrower range of issues than would be the case with fully centralised provision, it might be easier for the representatives to interpret voting patterns and ascertain preferences on individual issues. Moreover, as Jones and Stewart (1982, p. 74) have observed, local councillors ordinarily live and work in the areas where they are elected, and are more likely to be acquainted with local views than distant politicians in the nation's capital; they argue further that, in the United Kingdom at least, local councillors comprise proportionately more women and lower-class people than national politicians, and so may be closer and more responsive to the public. However, it may be observed that central governments might be able to deepen their understanding of electors' wishes, and thereby weaken this case for subcentral government, if they were willing to undertake referenda on particular issues.

Secondly, decentralisation may secure better democratic control of public officials. In the devolution debate in Great Britain in the 1970s, it was pointed out (HMSO, 1973b, p. 4) that the number of civil servants employed by the central government had grown from around 50,000 in

1900 to over 500,000 in 1973. This figure gives some idea of the increasing scale of central government power. Throughout this period, though, there was still one cabinet (whose size rose by around two-thirds) and one parliament (with a roughly constant number of members) to discuss the principles on which these people should base their work and to deal with complaints from constituents. It would be possible to increase the number of members of parliament, but there must be a limit to the number of issues on which they can acquire expertise. As for ministers, it was argued that the expanded scope of government meant that they might be unable to control adequately the full range of departmental business. In general, decentralisation with a number of tiers of government can simultaneously increase the ratio of representatives to officials and reduce the number of fields in which each representative has to operate.

The general theme behind the above advantages of subcentral provision has been that it would help representatives and officials in their efforts to secure services in accordance with their electorates' wishes. Much public choice analysis questions the competence of representatives and officials. This leads to a third advantage of decentralised provision, noted by Sproule-Jones (1975, pp. 23–30) – an advantage that is much greater when there is some overlap between the powers of the different tiers – that the authorities of different tiers will have to co-operate with one another, and, if one tier seems to pay little attention to consumer wishes, then consumers can apply pressure, not only directly, but also through the other tiers.

A fourth reason why subcentral provision may result in more efficient levels of local public goods and services is that electors may become more aware of the costs of the services and hence be in a better position to weigh up the benefits against the costs. Oates (1972, p. 13) has pointed out that in the United States, for example, proposals to improve local school systems are frequently submitted to the local electorate along with a proposed increase in tax rates to finance the improvements; the electorate is therefore well placed to weigh up the costs and benefits before voting for the improvement. By contrast, if school systems were centrally financed, then local communities would bear only a negligible part of the costs of improvements and would therefore be likely to lobby for more improvements whose marginal benefits did not match their marginal costs.

The advantages of decentralised provision so far discussed suggest that it may lead to public services being provided in closer conformity to electors' wishes; the next suggests that it could lead to greater productive efficiency. Specifically, as Oates has explained (1972, pp. 12–13), it is possible that decentralisation will result in more experimentation and innovation in the production of public goods. If, for instance, a new method of collecting refuse were invented and seemed worth trying, then it might be more likely that a local government would try it out than that the central government would. This could be partly because a central government would probably wish to try it out in one particular area, and has no ready source of information about which area or areas would be most willing to give it a try; it could be partly because the new method might cost less (or more) and the central government has no ready method of altering tax rates accordingly in one particular area; it could be partly because the central government might

feel unwilling to embark on the new method unless it thought it could eventually gain acceptance in all areas, since otherwise it could generate numerous complaints about inequitable activity in different areas; and it could be partly because a system in which subcentral authorities were allowed to choose their own methods might be a system in which more officials were actively looking out for alternatives. Of course, if just one area adopted a new method and found it popular with local residents, then the governments of other jurisdictions might well feel compelled to follow suit in order to avoid serious criticism from their local residents.

Arguments against subcentral provision of local public goods

Against these advantages of subcentral provision, it can be argued that decentralisation may have some disadvantages. Many of these suggest that decentralisation may not lead to more efficient levels of service provision. First, there is the possibility of tax exporting, whereby some of the burden of local taxes may be shifted on to non-residents, in which case the local citizens do not bear the full cost of local services. If, in this situation, they demand local services to be provided up to the point where the marginal benefit to them equals the marginal cost to them, then the full marginal cost of these services may well exceed their marginal benefit. Tax exporting can occur with virtually any tax for reasons that are explained in some detail in Chapter 6. Nevertheless, it is likely that the degree of exporting could be modest with many taxes. This was certainly the conclusion reached by McLure (1967, p. 73), who estimated the amount of exporting of various taxes by states in the United States in 1962. He considered that there would be very little exporting of taxes on personal income, taxes on expenditure (except those on motor fuel) or general property taxes if payments of these taxes could not be offset against federal tax liability. As it was, these offsets were found (p. 66) to be sufficient to help put the amount of most taxes exported between 15 and 25 per cent, though the figure was nearer 13 per cent for taxes on alcohol and tobacco; the amount of exporting with corporation tax was put at 75 per cent in the long run and 48 per cent in the short run. McLure considered that the degree of exporting was acceptable (or at least would be in the absence of federal offsets) for all major taxes except a corporation tax, a point that suggests that this particular tax should be confined to central governments.

Secondly, a related problem will arise if grants to subcentral authorities include a tax effort factor (so that increases in an area's taxes raise its grant receipts) because expenditure can again rise by more than electors' own tax payments. This problem can be avoided simply by not having a tax effort factor in the grants, although, as Chapter 5 explains, there is a case for this type of factor.

Thirdly, it is possible that subcentral authorities will provide services at less than efficient levels as a result of attempts either to attract industry into their areas or retain the industry that is already there. The most frequently discussed way in which this might happen is associated with tax competition. It is argued that industry is more interested in tax rates than service levels,

so that areas will be tempted to cut service levels in order to reduce their taxes. The problems of tax competition can be reduced, if not wholly avoided, by equipping subcentral authorities with taxes that do not fall greatly on the owners of businesses; this possibility is discussed further in Chapter 6. Another way in which attempts to attract industry could lead to service levels frequently being less than efficient arises if subcentral authorities cut their spending on most services in order to release funds for spending on what might be termed business lures, such as premises for industrial use leased at below market rents. This possibility is analysed in more detail towards the end of this chapter. For the moment, it may be observed that the problem is perhaps most likely to arise if the authorities regard the attraction of industry as an important function and are not given specific-purpose grants with which to finance their incentives. Even so, the problem may not be very serious if electoral pressure resists significant drops in service levels.

Fourthly, it is possible that subcentral authorities would have excessive spending levels, financed by excessive debt, if they were not required to finance current spending from taxes (or charges) and capital projects from loans that had to be paid off steadily during the lifetimes of the projects concerned. The relevant issues are discussed at length in Chapter 8. Here, it is sufficient to note that there are good reasons for supposing that the burden of any subcentral authority's debt (which is typically chiefly external debt) falls on the citizens required to pay taxes to finance its repayment. Thus an authority that responded to the wishes of its median voter might be encouraged to finance spending on items whose benefits were quickly exhausted from long-term loans, because the median voter might hope to enjoy most or all of the benefits while meeting little of the costs, as he might expect to leave the authority, or to die, before the loans were fully repaid. Excess spending could well arise if present voters were able to ignore part of the costs of the subcentral services they enjoy.

Fifthly, there is the possibility that overprovision would occur if the taxes levied by subcentral authorities were not perceptible to citizens, because they might then underestimate the costs of service provision. Sixthly, there is the further possibility that a subcentral authority that aimed to satisfy its median voter would overprovide if it was to levy a tax whose incidence was not widely spread and was largely escaped by the median voter. Chapter 6 explains how these problems can be largely, if not wholly, mitigated by a suitable choice of taxes. Meanwhile, it may be observed that the same problems can also arise with centrally provided services, since not all central government taxes are easily perceived, and not all have a widely spread incidence.

Seventhly, as noted earlier, the politicians or bureaucrats in any government may seek to provide excessive spending levels as a result of attempting to maximise their own welfare rather than that of their electorates. Interest in this aspect of government has been much stimulated by the recent tax and expenditure limitation proposals put forward, and frequently endorsed by referenda, in various parts of the United States, all of which have helped to focus attention on subcentral government there. Perhaps the most notable recent limitation proposals have been Proposition 13 in California in 1978[11] and Proposition 2½ in Massachusetts in 1980. The chief result of the former

was to cut property tax revenues in the state by, on average, 57 per cent, as a result of imposing a maximum aggregate local property tax rate of 1 per cent of each property's full cash value; the latter effectively imposed a maximum property tax rate of 2½ per cent for cities and towns, though areas exceeding that rate in 1980 were given some time to fall into line. Brennan and Buchanan (1979) have interpreted these limitations as attempts to control Leviathan governments seeking to maximise their own surpluses, defined as the difference between their tax revenues and public goods expenditures, a concept clearly related to some of the models of government behaviour outlined in Chapter 3. If this interpretation of the limitations is correct, however, then the problem might well apply to all levels of government. Accordingly, it would not seem to weaken the case for having some services provided by subcentral authorities rather than by the central government, although it might be felt to clinch the case against public provision of those items where the other arguments for and against are finely balanced.

It may be noted that there are other explanations for the recent crop of limitations in the United States. Brazer (1981, pp. 21–4) has pointed out that these include the feeling by middle-class voters that they have to pay too much in the way of taxes to support expenditures on services such as education or health that benefit minority groups, a widespread desire to limit spending on welfare and related programmes, particularly health and housing, opposition to the growth of public expenditure since 1973 while real income increases have been held close to zero, and a desire to reduce the relative burden of local property taxes. This last point was important in California where local property tax payments by home owners (though not by businesses) soared in the years prior to 1978 in the wake of property revaluations, and where the state government was running a surplus from which it was able, shortly after Proposition 13 was carried, to give local authorities generous aid and so replace much of their lost property tax revenue (see Gold, 1979, pp. 271–5). It was also important in Massachusetts where property tax payments as a proportion of income in 1980 were higher than in any other state except Alaska. Suggestions such as these indicate that subcentral governments may, from time to time, pursue policies that are out of kilter with their electors' preferences, but they do not really constitute a case against subcentral provision itself. Indeed, it is arguable that the most interesting feature of the limitation proposals in the United States is that citizens have felt obliged to secure their wishes through referenda rather than by voting the offending politicians out of office at elections.[12] Possibly citizens feel that state-imposed limitations will encourage generally acceptable politicians to cut waste and inefficiency, and strengthen their bargaining position vis-à-vis the unions; alternatively, the referenda may be sought by pressure groups with below-average demands, or by majorities who feel that existing budget levels reflect the preferences of pressure groups with above-average demands, or by individuals seeking vast amounts of publicity. Important and interesting as this feature of the limitations debate is, a more detailed analysis is beyond the scope of economics.

An eighth possible disadvantage of subcentral government could be lower efficiency of public service provision. It has been suggested (Peacock,

1977, pp. 56–9) that subcentral authorities might suffer from inefficiency, particularly if new authorities were established that lacked experience in expenditure control or project appraisal, or if authorities discriminated in favour of suppliers in their own regions. There might be a case for central government supervision, and possibly some controls, to reduce these risks.

A ninth problem of decentralised provision is that the authorities might be too small to exploit many of the economies of scale that may occur in service provision. In fact, as the next chapter shows, the optimal size for authorities might be one where they were not large enough to exploit all such economies. It could be that the gains from making the authorities larger, or, indeed, from having central government provision, would be more than offset by losses such as a reduced ability to cater for the disparate tastes of different small groups of people. However, the ability to exploit economies of scale must be borne in mind when determining the size of authorities (unless it is decided that they should purchase services from outside producers); in some cases this consideration may lead to central provision.

A tenth potential difficulty with subcentral provision is that subcentral authorities may ignore any external benefits to non-residents and so underprovide their services. It is possible that the authorities could negotiate between themselves and pay one another to provide such benefits, although this process would be unlikely to work at all well unless there were relatively few authorities concerned with each externality. Alternatively, a higher level of government could attempt to use grants to secure optimal service levels. Calculating the correct level of grants is a complex, costly and imperfect business (discussed in Chapter 4), and, for services where externalities are important, it may be preferable to have large authorities so that a relatively smaller proportion of benefits is external, or even to have central provision so that perhaps all externalities are internalised.

In addition to these ten main points, subcentral governments create some further problems which, though in principle soluble, incur administrative costs as a result. For instance, there is likely to be a need (discussed in Chapter 4) for some system of equalisation grants if citizens in different areas are not to risk finding that similar tax rates could result in very different service levels; a higher tier of government is likely to be involved in administering these grants. Another problem (discussed in Chapter 6) is that of finding suitable taxes for subcentral authorities; if not enough can be found, then the central government may be involved in administering some system of revenue-sharing grants (discussed in Chapters 4 and 5) to bolster local revenues. Further, the existence of subcentral authorities may pose a threat to the central government's own policies, and thus imply the need (discussed in Chapter 9) to administer some controls. The costs of administering all these programmes have to be seen as a disadvantage of subcentral provision.

Voting-with-the-feet

One feature of subcentral provision that has received a great deal of attention is the phenomenon of voting-with-the-feet. If local public goods were provided by subcentral authorities, and if preferences for quantities and qualities of services were to vary from area to area, then different authorities would be likely to have different service patterns and levels, and different tax levels, and this in itself would offer advantages in efficiency terms over uniform central service provision and tax levels. There is the further attraction that if some citizens were dissatisfied with the service and tax bundle in their own areas, then they could migrate to any area whose bundle appealed to them more. It was suggested in relation to Figure 1.3 that there might be a range of preferences in each area. Thus, a citizen who wanted a high level of service with high taxes, but who was living in an area where the median voter's preferences were low, might well be faced with a low level of provision and hence might move to an area where the level of provision was closer to his own preferred level. It is this process that is known as voting-with-the-feet.

This process has principally been discussed in reaction to a celebrated paper by Tiebout (1956),[13] and this debate must now be considered. In his paper, Tiebout pointed out that people who vote with their feet are revealing their preferences for public services, and he argued that, under certain assumptions,[14] consumer mobility would lead to a Pareto-optimal provision of local public goods without recourse to the imperfections of the voting process. In the extreme version of his model needed to produce this result, Tiebout assumed an infinite number of authorities providing each possible package of public services (pp. 420–1). Moreover, local subcentral authority services were not subject to any economies or diseconomies of scale and were financed by benefit taxes; more specifically, total service costs were held to be proportional to population, and poll taxes were used so that each citizen's tax payment equalled the cost incurred in providing him with services. Finally, consumer-voters had full knowledge of each area's budget package and were fully and costlessly mobile; even their incomes did not depend on their areas of residence. In this situation, each consumer was held to move to the area whose local service package precisely matched his tastes. No one's welfare from local services could be increased without making someone else worse off (perhaps by raising taxes in one area to increase service levels in another) so that there was Pareto-optimal provision of those services.

Tiebout himself acknowledged that there were severe limitations in this model. In particular, he argued that there would be economies of scale in the provision of local public services, and that instead of an infinite number of areas of infinitely small size there should be the number of areas consistent with each having the population that enabled it to provide its services at the lowest possible average cost per person.[15] He imagined that the areas would try to keep their populations constant, reacting to net emigration by, for example, attempting to attract new residents through chambers of commerce or other agencies. In this case, there would be only

an approximation to the ideal solution, for people might not find an area with a pattern precisely in accordance with their wishes, and might find zoning laws keeping them out of the best existing one. Tiebout also realised that people might not always move in practice, because moving would not be costless; however, he argued that the solution would be the best that could be obtained with given preferences and resource endowments (p. 421).

There are other limitations in the model. If, for instance, the pattern of services within each area was fixed, then, although citizens might end up in the areas closest to their preferences, the result might not be optimal, since, for example, the range of packages offered could be much narrower than people's desires. However, if areas sought to reflect the tastes of their median voters, then the result could conceivably be an unstable process; as Mueller (1979, p. 128) has pointed out, individuals might migrate into one area expecting one bundle of public goods, only to find that their presence altered the bundle offered, thereby inducing further changes in the bundle, further migration and so on. It may be added, too, that citizens might not find a package very close to their preferences. With just one local service, 101 areas would be needed to cover each percentile of preferences; with two services, the number would rise to 101^2; and with n services it would leap to 101^n. Yet consideration of factors such as economies of scale might suggest that there should be relatively few areas.

Tiebout's paper was presented as an answer to Samuelson's conclusion (noted earlier) that preferences for public goods could be ascertained only through a system such as the voting system with all its imperfections. In his reply, Samuelson (1958, p. 337) argued that people might not migrate to areas offering more tempting packages, first because they would rather improve their own communities than abdicate from them, and secondly because people often like heterogeneity even though it involves conflict. Formally, these points can be seen as restraints on mobility that reduce the effectiveness of the Tiebout process. On the first point, though, it may be asked whether non-median people should try to improve their communities if, for example, that means trying to make them conform to their own personal wishes. Samuelson also queried whether citizens should be free to migrate and abdicate their social responsibilities; however, few democratic countries, if any, have explicitly forbidden internal migration.

Most of the discussion of the Tiebout model has focused on two particular issues. On the one hand, from the individual's own point of view, there may be various factors discouraging a move to an area whose budget pattern appeals more than the one in his own area; these factors might mean that such migration is rare and its benefits to individuals modest. On the other hand, it is possible that, when individuals do so move in order to secure net benefits to themselves, they could be imposing external costs and benefits on others. In these cases, it is possible that the net external effect is a cost, and it is conceivable that this cost outweighs the net private gain.

There are many factors that may discourage an individual from moving. Job commitments, family ties and the benefits of living in an area known

well are some of the more obvious ones. These restrictions will be less forceful when local authorities are small, as might be the case with districts in a metropolitan area, and much of the literature on migration suggests that migration is most likely to occur in this type of situation. Oates (1972, p. 163), for example, has noted that individuals working in a city centre frequently have a wide choice of suburban communities in which to reside, and Tiebout (1956, p. 418) considered the case of a city resident who was about to move to the suburbs and who studied the availability in each of services such as beaches, parks, police protection, roads and parking facilities.

While migration may well be most likely between small areas, there are a number of barriers that can discourage people even then, apart from any desire to improve their own communities or any love of heterogeneity and conflict. First, there is the cost of moving, which includes removal costs, the costs of searching for new accommodation and, in the case of owner-occupiers, the costs of selling and purchasing properties. Secondly, there may be institutional rigidities. In the United States, for instance, proper zoning laws and implicit agreements between realtors were cited by Tiebout as ways in which immigration into an area might be discouraged (p. 420). In the United Kingdom, on the other hand, there is evidence that the council house system drastically reduces the amount of long-distance migration undertaken by the one-third of households who live in such houses (see Hughes and McCormick, 1981), and there is no reason to suppose that it has a different effect on short-distance inter-area migration.

Even if these barriers were sufficient to discourage anyone from moving solely to secure a preferred local budget, it could still be the case that people moving for other reasons would consider the local budgets of each area to which they might move and take them into account when making a final choice. Since many people move during a lifetime, often more than once, the private advantages of being able to vote with the feet might still be substantial. Of course, when a move is made, factors other than local budgets have to be considered, but there is evidence to suggest that local budgets are taken into account. There are problems in testing this proposition, but, in a review of many studies that have made such tests, Cebula (1979, p. 79) concluded that they supported the hypothesis that state and local policies do affect human migration patterns.[16] It is perhaps worth mentioning the conclusions of four studies to indicate the sorts of results obtained. Aronson (1974) conducted a survey of local government in the 1960s in urban areas in Harrisburg in the United States and in Leeds and Manchester in the United Kingdom, and found (p. 338) that fiscal factors affected the distribution of population in each country. Ellson (1980) reported a study of fifty metropolitan areas in the United States (1960–1970) and concluded (p. 208) that fiscal differentials were an important determinant of location, particularly for middle-income households. Pack and Pack (1978), in a study of suburban communities in Pennsylvania in 1970, found (p. 359), not surprisingly perhaps, that public goods questions were certainly not the sole or necessarily the major determinant of household location, but their results were consistent with the hypothesis that public goods questions were important at the margin.

In another test, Oates (1972) studied fifty-three residential communities in North-eastern New Jersey in the 1950s, and concluded that his results were consistent with the view that rational consumers weight, to some extent at least, the benefits from local public services against the costs of their tax liability (p. 176).[17]

Perhaps more serious questioning of the merits of the voting-with-the-feet process is made by those writers who have concentrated on the specific external effects of migration.[18] When an individual moves, he may be presumed to do so in the belief that he personally will be better off. His move would necessarily lead to a Pareto-improvement if it had no effect on anyone else, but it might have such effects. First, it might affect the tax rates in the area he leaves and in the one he enters. If, for example, local authorities provide roads, then the total expenditures in each area might not be altered by his move, so that he would confer an external benefit on the citizens of the new area equal to the tax he pays there, and an external cost on the citizens of the old area equal to the tax he used to pay there. Secondly, he may impose congestion costs on the users of the local public goods in the new area and reduce congestion costs on the users of them in the old area. In practice, the areas may respectively ease and restore congestion by raising and lowering their total expenditures, thereby simultaneously reducing both sets of externalities. Nevertheless, the fact that there are external effects that will not necessarily balance out means that the actual amount of migration, determined solely by private considerations, may be too great or too small. However, it might be possible to correct the situation by arranging for each area to tax or subsidise migrants in a way that internalised their externalities, or, alternatively, by arranging for the central government to make grants to areas with too few people and to levy taxes on areas with too many in order to stimulate the desired migration.

Other problems with migration can arise if the financial arrangements for subcentral authorities are not carefully devised. One of the principal attractions of the voting-with-the-feet process is that citizens are able to choose between areas with varying service levels and tax burdens. For this process to work well, citizens should be faced with a clear relationship between the level of taxes paid and the level of services offered. There are three situations in which this might not occur. First, if different areas were to levy different taxes, then, for example, a rich man might find he could reduce his tax bill by moving from an area with an income tax to one with a poll tax and not face a fall in service levels. Secondly, if each area were to levy the same tax, say an income tax, then tax rates might be lower in a rich area than in a poor one even when service levels were similar, thereby encouraging all citizens to move to the rich area. If the rich were better able to afford to move than the poor, then the rich area would get richer and the poor area poorer. Thirdly, if each area levied a tax where an individual's payment reflected his ability to pay rather than the costs of providing him with benefits, then the rich in high tax rate areas might pay far more than those costs and so envy their equals in low tax rate areas, while the poor in high tax rate areas might pay far less in taxes than the costs of their benefits and be envied by the poor in low

tax rate areas. If such envy exists, then it might be argued in general terms that there is horizontal inequity and a systematic encouragement for migration not directly related to preferences over service levels; however, the rich are likely to have higher valuations of public services than the poor, and as a result these feelings of envy might in practice be modest. A more specific problem with non-benefit taxes is that they encourage non-optimal location decisions, since citizens decide where to live by comparing the benefits expected in each area with the tax costs to them of living in each area rather than with the true costs each area would face in accommodating them. In general, the rich locate in areas where the benefits are lower than optimal for them, while the poor locate where the benefits are more than optimal for them.

These three problems are all discussed in later chapters, but it may be noted here that the first can be solved by harmonising the tax structure in each authority of any tier (as explained in Chapter 6) and the second can be handled by an effective equalisation scheme (as explained in Chapter 5[19]). The third can be eased by selecting taxes where payments are most closely related to the costs of benefits received (as explained in Chapter 6); these taxes might be regressive and so appear inequitable, but regressive subcentral government taxes could be offset by progressive central government ones.

One final issue with migration deserves mention here. It has been observed that people do not generally move specifically to get a preferred local budget. Indeed, moves are usually the result of one or more of a number of other factors. The fact that people do move from time to time, though, means that people may be reasonably concerned with tax rates and service levels in areas other than their own. They can influence these levels only through the central government (or some other higher tier) and may vote for minimum standards requirements to be enforced.

The distribution function

Whereas there is widespread support for the view that subcentral authorities have an important role to play in the allocation function, there is little support for their playing any part in the distribution function. No doubt some of the arguments considered earlier as objections to a subcentral allocative role could be raised also against their having a redistribution role. For instance, they might indulge in excessive handout programmes if their taxes could be exported or were imperceptible, or if their representatives or officials overruled electors' preferences in attempts to maximise their budgets. Also, there might be acute problems for areas with large proportions of poor or needy citizens and few rich citizens, although a suitable equalisation grant scheme could offset their difficulties. However, the principal reason for seeing no point in giving subcentral authorities redistributive powers is the feeling that there would be major problems for any area that attempted to achieve a more equal distribution of incomes than the others.

Suppose, for example, that subcentral authorities engaged in redistribution activities by means of positive and negative income taxation,

and that one particular area attempted a more ambitious redistribution programme than the others by taxing the rich more heavily in order to give more generous handouts to the poor. In this situation, there would be an incentive for its rich citizens to leave, and an incentive for poor people elsewhere to move in. In order to sustain its chosen level of handouts, the authority would have to increase the tax rates paid by the rich, and hence increase the incentive for them to leave and so exacerbate its problems. It is clear that, in the end, a more generous handout programme could be unsustainable.

The situation is not quite analogous for an area that opted for a less generous handout programme than the others, accompanied by a lower rate of tax on the rich. Doubtless the poor would be encouraged to leave, and rich people elsewhere would be encouraged to move in, but these moves would enable the chosen level of handouts to be maintained even more easily with lower rates of tax for the rich. The problem here is that the other areas would face greater difficulties in sustaining their original levels of handouts, and in the end they might have to follow the first area in setting less ambitious redistribution programmes.

The most general conclusions drawn from considering these problems (see Oates, 1972, p. 8 and Musgrave and Musgrave, 1976, p. 623) are that there should be a single nationwide redistribution policy and that this policy should be formulated and implemented solely by the central government. Actually, the second conclusion does not necessarily follow from the first: as Grewal (1981a, pp. 30–1) has pointed out, it is possible that the subcentral authorities could get together and formulate a single policy that each would agree to operate, though he notes that this arrangement does not have a great deal to commend it. One objection is that if there is going to be a uniform policy in the end, then it is likely that it could be formulated at a lower organisational cost by a single central authority acting on its own than by a group of subcentral authorities negotiating with one another. Another, and perhaps more important, objection is that decentralising the formulation of policy in this way could mean that the policy would not become more (or less) progressive, even if a clear majority of the population wanted this to happen. Suppose, for example, that there are three areas (A, B and C) each with 1,000 voters, and that the number of voters wanting a more progressive system is 750 in A and 450 in both B and C. In this case, a more progressive system is desired by 1,650 voters, or 55 per cent of the total electorate of 3,000. Nevertheless, areas B and C would oppose a more progressive system, and it is unlikely that this change would be made if the policy had to be formulated by the subcentral authorities. Conversely, the change probably would be introduced if the policy had to be formulated by the central government.

Another way of drawing up a uniform policy has been suggested by Dafflon (1977, pp. 29–30) in which the subcentral authorities would negotiate a common policy that they would like to see implemented, and then confer with the central government to resolve any differences there might be between the common subcentral policy and the central policy. While this has the intent of giving each citizen two channels through which his views could be taken into account, there is the danger that some of the

aspects of the subcentral policy that appeared in the final package might (for the reasons outlined in the previous paragraph) not be desired by a majority of the population.

It can be seen that the case for a uniform and (probably) central policy on redistribution relies chiefly on the fear that migration would cause problems with any non-uniform decentralised policy. However, it is arguable that the migration problem can be overstated. At the very least, as Musgrave and Musgrave (1976, p. 623) have argued, migration, which is a costly activity, might not take place if there were only very minor differences in distribution policies, and it could be argued further that migration would occur only when the policy differences in two areas were thought likely to persist. Moreover, it is reasonable to suppose that the larger the subcentral authorities were, the smaller the problem of migration would be. Also, it should be stressed that unless subcentral authorities rely wholly on benefit-related taxes (and none do) they will automatically produce some income redistribution as a side-product of producing goods and services;[20] there is no evidence that this redistribution causes serious problems. Further, it should be noted that tax rate differences could be given less effect on net incomes if subcentral taxes could be credited against central government tax liability, though this policy might be opposed on the grounds that it results in tax exporting (as explained in Chapter 6).[21] In addition, it is possible that there would be an equalisation scheme in operation that would give extra grants to any area that undertook relatively substantial redistribution and encouraged emigration of the rich and immigration of the poor and thereby eroded its per capita tax base; however, such an equalisation scheme would continue only if it commanded widespread support, and support for it might well fall if citizens in other areas found themselves required to give ever increasing aid to an authority whose actions systematically reduced its tax base (although the emigration of the rich from the area concerned would in fact increase the ability of citizens elsewhere to help it).

More importantly, it may be questioned whether the rich would regard paying taxes for a redistribution policy as a purely painful exercise; if not, then paying such taxes would not, after all, necessarily encourage them to leave areas where such taxes were high. It is arguable that those who pay taxes for redistribution programmes derive satisfaction from the relief of poverty.[22] This satisfaction may stem partly from altruism, but it may stem also from a feeling that a reduction in poverty could lead to falls in the crime rate and to increased consumption of services, such as education perhaps, which might be thought to generate external benefits. If there were these gains from giving, then the migration problem might be much less acute. No doubt an area adopting a particularly progressive policy would stand to lose some rich citizens who were required to pay more than they would pay voluntarily, but, as Grewal (1981a, p. 29) has pointed out, they might also attract in some rich citizens from other areas where they might be required to pay less than they would like to pay.

Indeed, it could even be argued that such migration of the rich reflected only their desire to vote with their feet until they found an area where the redistribution (and other) policies most closely matched their

preferences. However, there are two objections to the migration of citizens who move in response to different redistribution policies that would not apply to the migration of citizens who move in response to different policies in the provision of other services. First, there is the question of the migration of the poor. Even if an area with a particularly progressive policy were to find that there was no significant tendency for the rich to emigrate and, thereby, was able to maintain the level of its tax receipts, it could still find a significant degree of immigration by the poor from other regions and so find it was able to give less money to each recipient; Foster *et al.* (1980, p. 51) have noted that a problem with local authority provision of relief for the poor in sixteenth-century England was the tendency of the vagrant poor to wander in search of more generous areas. Secondly, while individual rich people may genuinely want a high level of redistribution and vote for high taxes, and while they may derive pleasure from the easing of poverty that occurs when these high taxes are introduced, nevertheless they may later attempt to become free-riders by moving to other low-tax areas. Each rich person will realise that his own contribution has an insignificant effect on the redistribution programme, and he may feel that he can derive at least some of the pleasure from its effects wherever he lives. It follows that an area seeking a particularly progressive policy could drive many of the rich away, even if they derive pleasure from generous anti-poverty policies. This problem does not arise with other local services; those seeking a high level of, say, street lighting, would not enjoy the benefits if they were to emigrate when the necessary high taxes were imposed.

The objections to subcentral participation in the redistribution function appear much weaker, however, if Pauly's (1973) view is accepted that redistribution is essentially a local public good. It can be argued that redistribution is a public good on the grounds that the satisfaction derived by those who pay the necessary taxes for the relief of poverty is both non-rival and non-excludable. For redistribution to be seen as a local public good, though, it is necessary to feel that a rich person benefits chiefly, if not only, from the redistribution undertaken by his own subcentral authority. This view might be supported by pointing out that the altruism of the rich is most likely to be inspired by the sight of poverty, and they are most likely to see the poverty in their own areas; also, they might derive most benefit from the relief of local poverty if this leads to a reduction in the crime rate or external benefits from the consumption of, say, education.

Pauly considered the implications of this view by supposing that taxpayers contributing to an anti-poverty scheme regard the optimal degree of redistribution as that where the marginal cost to them (in terms of extra tax payments needed) equals the marginal benefit to them (derived from further raising the incomes of the poor) and by supposing initially that all citizens are immobile between areas and that the rich care only about the poor in their own areas. In such circumstances, Pauly has shown that the optimal degree of redistribution will be greater in those areas where the tastes of taxpayers favour more distribution, and that it will also be greater in those areas where taxpayers are richer, provided redistribution

is a normal good. Consequently, optimal redistribution will require different amounts in different areas; in this case there would be a good argument for complete decentralisation of the redistribution function.[23]

However, if the rich were to care also about the poor in other areas, then each authority's anti-poverty measures would provide external benefits to citizens elsewhere, and, left alone, individual authorities might provide too little help to the poor. It is possible that the authorities would negotiate with each other in an attempt to persuade each other to take these external benefits into account, but it is more likely that the central government would step in and do some redistributing itself, or use specific grants (as explained in Chapter 4) to try and get authorities to increase their aid to the poor. If migration were allowed, then there might be further external effects; an area with a particularly progressive scheme would attract poor people from other areas (and possibly send rich people to them) thereby benefiting the rich in those areas. These externalities would further strengthen the case for a central government role. Nevertheless, the facts that the tastes of the rich and incomes vary from area to area mean there is a case for some variations in redistribution policy between authorities. This argument is somewhat weakened by the fact noted by Grewal (1981a, p. 31) that, even with a uniform national policy, those taxpayers preferring more redistribution can always top up their payments to the poor by voluntary private donations; however, if poverty relief is a public good, then it is possible that such private donations would never result in adequate finance for redistribution.

Overall, it seems that it may be desirable to have some variation in the degree of redistribution in different areas; and it is arguable that the migration problems that might result can be overstated, so that variations in the degree of redistribution might be feasible, particularly, perhaps, if the variations were small and the areas were large. This suggests that there should be a basic national redistribution policy, and that subcentral authorities should be allowed to alter the degree of distribution in their areas within specified limits. In view of the problems that might arise if an attempt was made to get the subcentral authorities to agree on the basic national policy, it would seem that this should be determined by the central government, though it could be implemented either by the central government itself or by the subcentral authorities on an agency basis. Either way, there would be a case for giving subcentral authorities some explicit redistributive weapons, such as, perhaps, a share in the transfer payment system and the right to levy a non-proportional income tax, in order to achieve their desired final patterns of net incomes.

It should be noted, however, that the beneficiaries of a redistribution policy can be helped in ways other than cash handouts, for example by providing them with merit goods like low-cost housing or free school lunches financed from tax revenue. It has been suggested by Musgrave and Musgrave (1976, pp. 65–6) that merit goods such as these are often provided in preference to cash handouts because donors like to choose the way in which their gifts are used even though that might not be the way that recipients would themselves choose. If such merit goods were provided by subcentral authorities, then they would be able to moderate

the degree of redistribution in their areas without needing to operate transfer payments.

In general, the consumption of merit goods can be stimulated by free or subsidised public provision, by subsidised private provision, or by supplying intended beneficiaries with vouchers (that is, earmarked cash handouts that can be presented only to the suppliers of a certain specific service and used in part or full payment, the suppliers being able to redeem the vouchers for cash). The relative merits of these different possibilities are considered in Chapter 7. In order that the overall degree of redistribution should not vary greatly from area to area, it is necessary that the stimulus for different services should not vary greatly. This means that, if subsidies to private producers or vouchers were used, then the levels of subsidies and values of vouchers should be broadly determined by the central government, and that subcentral authorities should be able to have only a small effect on the levels and values operating in their areas. Likewise, if there were subsidised or free public provision, then the value of the benefits provided should not vary greatly from area to area. In this situation, however, there would be a case for subcentral authorities actually to provide the service; the case has much in common with that for allowing subcentral authorities to provide local public goods. With merit goods, subcentral provision will enable the detailed specification of services to vary from area to area in accordance with local tastes, it may help citizens to signal their preferences, it may facilitate closer democratic control of bureaucrats, it may encourage experiments, and it enables citizens to migrate to areas where the service specifications appeal to them most.

The stabilisation function

There is undoubtedly least support for subcentral participation in the stabilisation function; indeed, it is widely believed that subcentral authorities have no part to play here at all. It must be said at once, though, that the reason for this is not a feeling that views on macroeconomic policies do not vary from area to area. It would certainly be accepted that opinions and preferences on stabilisation do differ between one part of a country and another. In some areas there may be, for example, a greater willingness to increase unemployment in an effort to curb inflation, or a greater willingness to make any environmental sacrifices that economic growth might demand, or a greater willingness to borrow from abroad (or other regions) to finance an international (or interregional) balance of payments deficit, than there is in other areas. Thus there is, in principle, reason to see if macroeconomic policy-making could be handled, at least in part, by subcentral authorities. In this way, it could be hoped that variations in tastes regarding stabilisation might be met, at least in some degree, in the same way that subcentral authorities can seek to cater for variations in tastes for local public goods, and, perhaps, distribution policies. Indeed, the possibility of making a case for decentralising stabilisation policy becomes clearer if it is appreciated that items such as the rate of growth and the rate of inflation are not unlike public goods;[24]

the benefits produced by a rise in one or a fall in the other would generally be non-excludable, and some might be non-rival.

The exclusion of subcentral authorities from the stabilisation function stems partly from considering the side-effects that would result from their using the instruments necessary to control macroeconomic variables, and partly from considering the direct internal and external effects these instruments would have.

The major macroeconomic weapons are covered by the terms 'monetary policy' and 'fiscal policy'. Consider the possibility of decentralising monetary policy. It must be supposed that each of the subcentral authorities would have the power to create money, and also that, as parts of a single nation, the same currency was used in each area. In this situation, there would clearly be an incentive for each area to expand the money supply in its own area, possibly by inviting the official subcentral reserve bank to increase the deposit of the subcentral authority in return for some of the authority's bonds, and then using that deposit to purchase goods and services from other areas. The benefit to each area would be that goods and services could be purchased without the need for taxation.[25] The cost, however, would be that, with each area liable to raise its money supply rapidly, the rate of inflation could soon rise out of control. Grewal (1981a, p. 21) has pointed out that the authorities would have an incentive to co-ordinate their monetary expansion at low rates in order that all would benefit from lower inflation, but he suggests that in such a case monetary expansion might as well be a central responsibility; it is worth adding that, while each authority would have an interest in securing agreements on modest credit expansion, each would also have an interest in circumventing any agreements that were made.

Next, consider the possibility of decentralising fiscal policy. The main way in which fiscal policy is used for stabilisation policy by a government is for it to run up a budget deficit when it wishes to increase aggregate demand, and a surplus when it wishes to contract it. Now, deficits and surpluses respectively are created when public expenditure exceeds or falls short of taxation. Insofar as subcentral authorities are likely to have taxation and expenditure powers, they may be, in principle, in a position to run deficits and surpluses; in practice, however, virtually all subcentral authorities are in receipt of grants from higher levels of government, and it may be wondered whether surpluses could be sustained for very long without those grants being cut.

A further problem with subcentral fiscal policy is that sustained and substantial differences in tax and expenditure levels could lead to the migration of households into fiscally more attractive areas and yet result in their members taking new jobs where their labour was less productive. This problem could also arise, of course, if tax and expenditure levels were to vary when areas chose different levels of public services when tackling their resource allocation functions, but the problems could be much more acute in the present context. This is because low tax rates would not now be necessarily offset by low expenditure levels since areas might be anxious to secure budget deficits.

Apart from these conventional monetary and fiscal instruments,

subcentral authorities might wish to consider influencing macroeconomic variables by some of the other weapons used by central governments. Unfortunately for them, few of these would be available. As parts of a country using a common currency, they could not change their exchange rates, and it is unlikely that they would be permitted to use tariffs or quotas because citizens in other areas would probably demand that the central government forbade their use. They might be permitted to switch their own purchases of goods and services in favour of home suppliers, thereby attempting to cut local unemployment, although (as explained in Chapter 9), the central government might seek to curb this power on the grounds that purchases should always be made from the cheapest or most efficient suppliers. They might also be allowed to use a prices and incomes policy, although nationwide union bargaining occurs in many countries and would make independent policies for each area hard to pursue effectively.

It seems, therefore, that subcentral authorities would have recourse to fewer instruments than a central government, and that the major instruments would have worrying side-effects when used in a subcentral context. Nevertheless, a value judgement needs to be made about whether or not these disadvantages of decentralised stabilisation policy are outweighed by the advantages. In principle, the advantages would focus on the possibility of the authorities tailoring their stabilisation policies to meet the desires of local residents. In practice, however, there is great doubt about how much effect any one authority could hope to have on its own area. This doubt arises partly because of the difficulties areas would face in trying to set different budget surpluses or deficits and interest rates: differences in budget surpluses or deficits could lead to problems with migrating households, labour, trade and capital, while differences in interest rates would probably be impossible if there were free movement of capital within the country.

In addition to these problems, it can be argued that most of the effects of any area's macroeconomic policies will spill over into other areas. This at once makes life harder in other areas, thereby adding a further disadvantage to decentralising stabilisation, and simultaneously means that the desired internal advantages are unlikely to be fully realised. One reason that much of the effect of any policy would spill over is that subcentral authorities generally have high marginal propensities to import (either from other areas of the same country or from other countries) and this in turn means that the internal multiplier effects of any change in demand produced by monetary policy, fiscal policy or switches in purchases in favour of local suppliers will be small.[26] A prices and incomes policy would also be likely to be ineffective: if one area were to succeed initially in holding price rises at a rate lower than those elsewhere, then it would be likely to find demand rising and the situation untenable.The spillover effects of subcentral macroeconomic policies can also be viewed as spill-ins. It is often pointed out, for example, that the origins of a high level of unemployment in an area may lie chiefly in forces outside that area.[27] This means that the source of the problem is beyond the area's jurisdiction. The area could attempt to bribe other areas to take appropriate action,

but there is clearly a case for avoiding this probably imperfect process by making the central government responsible for such problems.

Regional development

It is perhaps not surprising that the general view is that stabilisation should be left to the central government. It is nevertheless becoming clear that the matter is not as simple as that. In any country, it is likely that some regions will have relatively high unemployment rates or relatively low income levels, and frequently the disparities are sufficiently large for there to be widespread support for action to be taken to try and reduce them; the most clearcut action is to try to make the regions concerned more attractive to industry in the hope that they will attract a greater proportion of new industry (and lose a smaller proportion of their existing industry) and so have higher growth rates than would otherwise be the case.[28] It is also possible that in some countries certain regions will appear too congested, in which case there may be widespread support for action that attempts to encourage industry there to move elsewhere, and that encourages new developments to take place elsewhere. This section focuses on the question of how far subcentral authorities should participate in the promotion of such regional development, which is arguably a branch of stabilisation policy.

One possibility is that regional development policies should be wholly operated by subcentral authorities. In this case, the need or desirability of development in any area could be made a factor in determining its receipts from the central government of equalisation or revenue-sharing grants and, perhaps, grants earmarked for use in attracting industry. A common criticism of this possibility is that it could lead to a tax competition[29] – that is, a situation in which subcentral authorities tried to make their areas attractive to industry by keeping their tax rates below those needed to secure efficient service levels. This line of reasoning deserves some further analysis.

Before presenting this analysis, it should be mentioned that a study conducted by the Advisory Commission on Intergovernmental Relations in the United States (ACIR, 1981a, p. 4) concluded that state and local tax differentials seemed to have little effect on business location there, thereby perhaps discouraging authorities from embarking on tax competitions. The limited effect arose partly because business location was influenced by more important factors than differences in state and local taxes, and partly because those differences were to some extent offset by federal tax concessions for payments of state and local taxes; the first of these factors may well apply in most other countries, while the second may not. Accordingly, maybe not all individual countries would be concerned about the effects of tax competition or about ways of avoiding it; but for those that are so concerned these matters are discussed below.

A critical question is whether or not the 'low' tax rates being set are, in the circumstances, the ones that the electors prefer. Discussions of this topic give great emphasis to the behaviour of local officials. It is alleged

that it is the officials who want to retain or attract businesses, and that it is they who set the low tax rates. This emphasis on officials suggests that the tax rate is set by the officials in an attempt to satisfy their own preferences, in which the level of business activity is a key factor, rather than those of electors, who are implicitly held to be more interested in service levels. Now, if the suboptimal service levels are caused because officials do not set tax rates wholly in accordance with electors' preferences, then the appropriate action is to improve either the processes by which elected representatives control the officials or the processes by which the representatives are selected and elected; the choice between these two possibilities depends on whether the representatives are inclined to side with the electors or the officials. Of course, if the electors do approve of the tax rate and the consequential levels of services, then it is not so immediately apparent that those service levels can be regarded as suboptimal.

However, a case can be made to show that electors might, paradoxically, choose suboptimal service levels. Suppose, for instance, that a system of subcentral government is established in two areas (A and B) and suppose, for convenience, that their electors choose to levy identical taxes at common rates and use the yields to finance similar optimal levels of services. Assume, too, that the distribution of businesses between A and B is unaffected by the introduction of subcentral government. As time passes, area B's citizens may come to realise that a lower tax rate would, eventually, increase the number of businesses in their area, and hence, perhaps, raise their incomes. B's citizens may accordingly demand a lower tax rate and claim that the new, lower levels of services were now optimal given their latest views on the desirable levels; in other words, they might claim that the lower service levels came in response to a change in tastes brought about by increased knowledge.

In fact, B's new budget can be judged non-optimal because B's citizens are not confronted with the whole cost of their benefits. Those benefits include the advantages accruing from attracting new or existing businesses that would otherwise locate in A. Part of the cost of those benefits is indeed borne by B's citizens when they reduce their service levels below the ones most preferred, but part is also borne by A's citizens, who lose the advantages that would have accrued to them from any diverted businesses. Indeed, it is possible that the overall level of business in a country is not very sensitive to the fiscal behaviour of subcentral authorities, so that any success B has is largely on a beggar-my-neighbour basis at A's expense; in other words, A and B may be involved in a near zero-sum game where any benefits each might gain involve a substantial external cost.

Now B's citizens cannot claim to have an efficient budget if external costs are ignored. Efficiency requires that externalities be internalised, and thus B should be required to compensate A for its losses whenever B cuts its taxes. Such compensation would discourage tax cuts in B, and it can be assumed that its tax rate and service levels would be higher if its budgets were determined in the presence of compensation requirements than if they were determined in their absence. To put it another way, an absence of compensation requirements can be assumed to lead to service levels being below the optimal levels that would arise in their presence. It follows that

introducing such requirements would, in principle, be one way of solving tax competition problems. In practice, however, it would be hard, perhaps impossible, to estimate the correct levels of compensation. A more appropriate solution would be to select taxes that impinge least on businesses and are therefore least likely to be used in tax competitions; this matter is discussed in Chapter 6.

It is worth noting, though, that the problem of suboptimal service levels could still be caused by authorities seeking to attract or retain businesses, even if they were made to use taxes that did not impinge on businesses at all. Suppose, for example, that subcentral authorities were required to levy poll taxes, and believed that the rate of tax was of no concern to businesses. In those circumstances, they might have no desire for relatively low tax rates but they might well consider cutting their expenditure on many other services in an effort to build up funds to finance industrial lures such as premises leased at rents below market levels (some other lures are mentioned in the following section). The allocation of the income available to an area's citizens between these lures and other subcentral spending (and indeed private spending) would tend to be suboptimal if the citizens ignored the effects of their lures on citizens in other areas. Unless these externalities can be internalised, there is a case for central regulation of spending on lures, just as there is a case for a prohibition on the use of taxes that might be suitable for tax competitions. In determining the appropriate expenditure on lures by each area, the central government will, or course, need to bear in mind a point discussed by many, including Walsh (1974, pp. 26–30), which is that, from a national point of view, investment should generally be encouraged to take place where it will yield the highest returns.

This analysis suggests that if the regional development function were wholly operated by subcentral authorities then the central government may well wish to curtail their choice of taxes and may well wish to control their spending on lures. At the same time, it may feel obliged to channel substantial funds to those authorities in most need of development. In these circumstances, it would not be surprising if the central government decided to explore the possibility of itself assuming the prime responsibility for tackling regional imbalances.

Even if the central government did acquire this responsibility, there would still be scope for subcentral authority participation. The central government could, of course, operate its own programme of regionally varying incentives for industry. Alternatively, it could allocate earmarked funds to the authorities themselves in proportion to its perceptions of their needs, and allow them to use these funds (and no other funds) for incentives on criteria of their own choosing. The relative merits of these two possibilities need further discussion than they have so far received. Arguments for central implementation would include the facts that it might make it easier to exclude areas that, although needy, were unsuitable for development, and that it might make it easier to include remission of central government taxes as incentives. On the other hand, subcentral implementation might make it easier to co-ordinate the incentives with any planning and infrastructure programmes that those authorities

operated, and might make it more likely that local conditions would be fully appreciated in developing criteria. Furthermore, subcentral authorities might pursue development policies with a greater sense of urgency.

Subcentral authority functions in practice

It is appropriate to conclude this chapter with a few notes on the range of subcentral functions in practice. These notes should be read with an important caveat in mind, namely that the powers that local, and even state, authorities actually have may be much narrower than they seem initially on account of various supervisory, joint or overruling powers belonging to higher levels of governments. Thus the mention of a function in the following paragraphs often indicates more the existence of some subcentral discretion rather than total subcentral autonomy.

It is interesting to begin by pointing out that subcentral authorities, particularly the states in federal countries, may have substantial law-making powers; this makes the negligible theoretical literature on this subject a particular disappointment. For example, the tenth amendment to the Constitution of the United States has been held to ensure that the states can organise their own governments without national interference, legislate for the health, welfare, safety and morals of their residents, and assume responsibility for and control of the local authorities in their areas (ACIR, 1981d, p. 17). In a similar vein, jurisdiction in Canada over education, health, welfare, property and civil rights, and the ownership of natural resources is largely entrusted to the provinces (ACIR, 1981d, p. 11). Even in a non-federal country, some decentralisation of legislative powers may occur, albeit typically on a more modest scale. The 1979 proposals for a Scottish assembly, which failed to get the requisite support in the referendum, provided for the devolution of responsibility for the general criminal law, including the right to create new offences and to abolish existing ones, although there were certain exceptions (such as the laws on treason, espionage and terrorism) where it was felt that the laws could work efficiently only if they were operated on a uniform basis; the assembly would also have had power to fix matters such as public holidays, summer time, Sunday observance and shop hours (HMSO, 1975, pp. 30–2).

Regulation of private industry by subcentral authorities may at times be permitted and controls may be used in certain cases, particularly perhaps in connection with public transport. The authorities may more frequently operate businesses themselves, however, and they may be able to prevent competition as a result. In the United Kingdom, for instance, local authorities run a number of services on a broadly trading basis, often more or less as local monopolies; these include passenger transport, harbours, docks and piers, toll bridges, tunnels and ferries, as well as markets, slaughter houses, civic restaurants, crematoria and cemeteries (HMSO, 1981a, pp. 14 and 18). As noted earlier, subcentral authorities sometimes have some powers over externalities created by industry such as setting and enforcing effluent standards.

There is a fair degree of uniformity between countries over the local public goods, or at least the items that are arguably local public goods, which are subcentrally provided. The United Kingdom's local authorities[30] provide a fairly typical range of services whose benefits are perhaps non-rival and non-excludable, including roads, footpaths and street lighting, police services, the administration of justice, coastal protection, land drainage, environmental health, town and country planning and a small amount of civil defence. They also provide a number of services where pricing would be possible but where the presence of non-rivalry suggests the efficient price might be zero, such as parks, libraries and museums. Until 1974, they provided water and sewerage services where, arguably, the costs of operating price-exclusion would outweigh the benefits, a situation that might apply also to the fire services and refuse collection and disposal services still offered, although zero-priced public provision of these could also be seen as a way of avoiding the external costs that might arise if the services were left to private enterprise and priced.

Subcentral authorities also tend to be concerned with the provision of merit goods. The *Länder* and local authorities of West Germany, for instance, provide elementary schools, secondary schools, universities and other institutions of higher education, as well as having some responsibilities for housing, hospitals and other health affairs.[31] State and local authorities often provide merit good services for disadvantaged groups, such services including facilities for the old, homeless, deaf, blind, disabled, mentally handicapped and children in need. The provision of allotments at modest rents by local authorities in the United Kingdom is also, perhaps, a merit good.

Despite the arguments noted earlier redistribution in cash is not always confined to central governments. In Western Europe, for instance, many countries arrange for social aid, one of the means-tested safety nets against poverty, to be a local responsibility, and the states in Canada, West Germany and the United States all have some social security functions. It may be noted that the present federal administration in the United States is trying to reduce its share of responsibility for income maintenance in favour of the states; if Herber (1975, pp. 88–91 and 112) is correct in saying that the failure of fiscal policy to alleviate poverty there stems mainly from over-decentralisation of policy decisions together with inadequate equalisation grants to poorer states that are unable to be generous to their poor, then this new policy, coupled with cuts in the general revenue-sharing programme, will need careful handling to ensure that the poor are not to lose as a result. The states there have a relatively high ability to affect the distribution of net incomes with tax policies: they can vary their circuit-breaker tax relief measures to try to protect low-income families from excessive property tax burdens,[32] and the forty or so states that levy personal income taxes can set their own exemptions (perhaps to give help to the old, the blind or families) and, as Maxwell and Aronson (1977, pp. 93–7) have pointed out, they generally try to have a progressive structure of graduated tax rates.

As far as stabilisation is concerned, responsibility always lies primarily with central governments. Some federations have instituted arrangements

whereby the states can make their views known – the roughly annual federal and state premiers' conferences in Australia and the same leaders' deliberations in the Loan Council (discussed a little further in Chapter 9) and the various meetings of federal and provincial ministers and bureaucrats in Canada being important examples. Perhaps a more formalised influence is available to West Germany's *Länder*, which appoint the members of the Bundesrat, the upper house of the federal parliament.

More particularly, though, subcentral authorities are often active in connection with regional development. It was suggested in the last section that, in such circumstances, the central government should, ideally, seek to control the taxes available to subcentral authorities and the amounts spent on lures for businesses in order to prevent tax competition and suboptimal service level. This would seldom be possible in federal countries, however. In the United States, for example, there is active competition for labour and capital, and this competition, which includes tax competition, is seen by Break (1980, p. 25) as a basic fact of life in a large country. Authorities there can also use such lures as industrial revenue bonds with favourable interest rates, or the purchase of facilities leased at low rates, or the provision of special water, sewerage and road facilities to incoming firms; the federal government would not be likely to seek strict controls over the amount and use of these lures, irrespective of any merits there might be in so doing.

In other federal countries, there is perhaps more central involvement with regional development, although in no way complete central control. In Canada, for instance, the federal government has long operated regional policies, and in 1969 it pulled these together with the creation of a new Department of Regional Economic Expansion. While this co-operates with the provinces, they continue to have considerable autonomy in the operation of their own development policies.

Similarly, the Australian states have substantial autonomy in operating employment-creating policies, although the federal government operates such policies too. In recent years, both the federal government and the state governments have at times allowed local authorities to act on their behalf in this field and have given these authorities specific grants. It should be stressed that the main rationale for development policies in Australia has been to ease congestion in the main urban areas; since development elsewhere confers external benefits in the urban areas, it seems reasonable for local authority initiatives to attract grant aid. In fact, though, the local authorities have often spent more than these grant receipts on their employment-creating activities (see ACIR, 1981c, p. 47) so that they can be seen as a second level of subcentral government acting with some autonomy in this matter.

In West Germany, regional development is very much a joint activity of the federal government and the *Länder*. The Constitution (article 74) lists a number of areas subject to concurrent legislation, where federal legislation may be enacted with the concurrence of the *Länder*; these include regional planning. Moreover, the 1969 constitutional amendments specify three areas for co-operative federalism, where the federal and *Land* authorities act together, and these include the improvement of the regional

infrastructure. This federal involvement may dampen subcentral autonomy, as may the fact that the individual *Länder* cannot control their tax rates and hence the total revenues available for this and other functions. Nevertheless, the federal government's contribution to the cost of regional economic policy in 1977 was just 14 per cent, somewhat higher than in previous years but still indicating a substantial role for the *Länder*.

In the United Kingdom, the primary responsibility for regional policy unquestionably lies with the central government. Compared with the sums that have been spent by the central government on capital and labour subsidies in selected areas, as well as the revenue forgone in tax concessions in those areas, the local authority contribution in financial terms has undoubtedly been modest. Nevertheless, local authorities do have a number of devices at their disposal for attracting industry. For example, they can set up units to advertise the attractions of their areas, to deal with enquiries and to try and remove any financial, legal and other obstacles to expansion; they can also give priority local authority housing to 'essential incomers' for industrial developments.[33] Similar instruments are doubtless available to state and local authorities in many other countries.

Summary

It seems that subcentral authorities are likely to be concerned chiefly with resource allocation. Government intervention with resource allocation may be undertaken for a number of reasons. For example, there may be a need to control national or local monopolies or oligopolies, either by regulation or by public ownership. There may be a need to stimulate or discourage the production of items whose production or consumption generate external benefits or costs; this may be done by subsidies and public provision, or by taxes and controls. There may be goods or services whose consumption is non-rival, non-excludable, or both; for these there are efficiency arguments for zero-price provision financed by taxes. Similar arguments arise for items where price-exclusion would generate more benefits than costs. In addition, there may be a desire to stimulate the production of merit goods, perhaps by laws, publicity campaigns, vouchers, subsidies or zero-price provision, and a desire to discourage the production of demerit goods, perhaps by regulations or taxes.

There may, of course, be many items involved in each of these various groups. For each item, the first question to arise is whether subcentral intervention would be feasible. In some cases, it could be that so many of the benefits of any one area's activities would accrue to non-residents that there could be too little intervention unless the central government undertook substantial intervention itself. In other cases, particularly where the production of goods and services is involved, technological factors could mean that there would be substantial gains from large-scale provision. On these occasions, the case for central intervention may be overwhelming.

However, there are likely to be many occasions when subcentral participation would not be ruled out on these grounds. Here, the

arguments for and against subcentral involvement need careful consideration. The main advantages of it are that any geographical variations in tastes can be catered for, that citizens can vote with their feet to live in areas with the packages of policies they most prefer, that there may be more effective control of public officials, and that there may be a greater degree of experimentation in the public sector. The first two advantages, at least, offer greater potential gains when there are many small authorities rather than a few large ones. Of these two, voting-with-the-feet seems to have attracted the most discussion. There seems to be evidence that it occurs to some extent despite various barriers such as cost; and although migrating citizens may not take account of the external effects of their moves, it is certainly possible that their actions enable the provision of public goods and services to be at levels closer to the optimal ones.

In principle, it is possible that many of the worries about decentralisation that are most frequently cited can be resolved or eased by satisfactory financial arrangements. Problems can certainly occur if subcentral authorities are given taxes that are easy to export, or that tempt them to indulge in tax competitions, or that are not perceptible to their citizens. There might also be cause for concern if they rely heavily on effort-related grants or if they are not given grants to overcome any tendency to ignore the external benefits that their activities might generate. Moreover, there is clearly a case for enforcing satisfactory expenditure control and project-appraisal techniques, and it is suggested that there is a case for regulating their use of loan finance. The existence of lower tiers is likely to generate a need for equalisation grants, revenue-sharing grants and, perhaps, a measure of control for macroeconomic purposes; all these matters will incur administration costs, and it may not be possible to handle them in ways to suit everyone. Perhaps the most publicised criticisms of subcentral authority behaviour have been those that led to the recent tax and expenditure limitation proposals in the United States. While these are seen by some as evidence that those running lower-tier authorities are inclined to overspend, it is not certain that those running the central government will necessarily behave any differently, so there seems no clear case for citing the introduction of these proposals as an argument against decentralisation *per se*.

The main problem with devolving parts of the redistribution function is that areas seeking relatively progressive policies might find them untenable if rich people left and poor people came in. Small variations in policy might, however, be possible between large areas without causing an unacceptable amount of migration, and they might be made less untenable with the help of an equalisation scheme. The case for some decentralisation rests chiefly on the fact that the desire of the rich to pay taxes to help the poor could vary from area to area. However, the need for variations in progressivity to be limited suggests that any subcentral redistribution policies, including relevant merit good provision, may need to take place within a centrally determined range of generosity to beneficiaries.

As far as the possibility of decentralised stabilisation policy is concerned, the main problems seem to be that individual areas would have recourse

to few acceptable policy instruments, and that their usually high import propensities mean that the effects of their actions could be felt largely by other areas. There are, nevertheless, some arguments in favour of decentralising regional development activities, although there may need to be some control over the instruments used, and there is a case for much of the money spent by lower-tier authorities in this context being provided by central government grants. There is also a need to control the extent to which these authorities could attempt to attract industry with low tax rates or generous lures, since these attempts might otherwise cause them to provide other services at suboptimal levels.

Notes

1 See, for example, Oates (1972, Chapter 1), Grewal (1981a, pp. 3–32) and Dafflon (1977, Chapter 1).
2 For discussions of co-operation, see Dafflon (1977, pp. 183–4) and Castles (1974).
3 This analysis derives from Foster, Jackman and Perlman (1980, pp. 70–2).
4 Samuelson's condition is based on the assumption that the taxes used are lump-sum taxes with no adverse redistributional effects. For a discussion of the modifications to his condition needed when these rules are relaxed, see, for example, Atkinson and Stiglitz (1980, especially pp. 487–99). These modifications add to the problems for policy-makers that are discussed in this text.
5 An individual's demand will also depend on his initial income and wealth. It is assumed in the text that the initial distribution of income and wealth has been fixed by the central government.
6 For an introduction to this theory, see Mueller (1979).
7 See Mueller (1979, pp. 68–89) for a review of the principal new processes.
8 In principle, externalities may also be resolved by bargaining and negotiation between areas. This is rare in practice, except perhaps where there are only a few authorities affected by each specific externality. It may be noted that grant schemes designed to encourage subcentral authorities to take external benefits into account may in fact reduce aggregate production for reasons explained in Chapter 4.
9 This example is derived from that in Sproule-Jones (1975, p. 16).
10 They also found values of around 1.0 for education and 0.9 for higher education, but it is less surprising that these items do not appear to be public goods.
11 For a number of papers on Proposition 13, see Puryear, Ross and Shapiro (1979).
12 For a discussion of this issue, see Ladd (1978, especially pp. 3–5).
13 For a formal presentation of Tiebout's model, see Pestieau (1977).
14 For an extended discussion of all the assumptions needed to obtain this result, see Bewley (1981).
15 Chapter 2 explains that there are other factors to be considered in determining the optimal size for an area.
16 For a discussion of the problems encountered in empirical tests of the Tiebout hypothesis, see Epple, Zelenitz and Visscher (1978).
17 Oates's methodology has been discussed in a number of papers, notably Edel and Sclar (1974).
18 See, for example, Buchanan and Wagner (1970), Buchanan and Goetz (1972) and Flatters, Henderson and Miezkowski (1974).
19 In Chapter 5, it is explained that to ensure that any two areas setting the same tax rate can have the same service levels, a tax effort factor has to be built in to the equalisation grant formulae. It has been noted in this chapter that such factors may encourage supra-optimal levels of local services. That may be a lesser evil than no subcentral government and so no local choice, or Pestieau's suggestion (1977, p. 184) that voting-with-the-feet be replaced by regional fixity of citizens.

20 For a discussion of how large these effects could be, see King (1973, pp. 51–6) and Gillespie (1965).
21 Foster et al. (1980, pp. 234–40) have suggested that the problems with local redistribution activities might be reduced if they were financed by a land tax. In particular, they have shown that the different tax rates that might be needed to finance similar benefits in areas with different proportions of recipients would cause migration problems with any tax except a land tax. In fact, this objection to other taxes would not arise if suitable equalisation grants were in operation to help high needs areas.
22 For a discussion of this, see, for example, Hochman and Rodgers (1969), Musgrave (1970) and Pauly (1973).
23 For a shorter exposition of Pauly's model, see Boadway (1979, pp. 411–17).
24 Head (1976, p. 278) has argued that 'economic stabilization can be regarded as a . . . public goods problem with an important spatial dimension'.
25 Formally, taxes would have to be raised to pay the interest to the subcentral bank that holds the bonds. This bank could well be owned by the authority, however, and could therefore be instructed to accept the bonds at a zero interest rate or required to hand its profits back to the authority.
26 This issue is discussed by King (1981). He has shown (p. 24) that, under plausible assumptions, the multiplier effects of an increase in spending by subcentral authorities in the United Kingdom might lie between 0.23 and 0.62 if there were an import component in the first round spending, and between 1.03 and 1.20 if there were not. These latter figures may be compared with a range of 1.19 to 1.90 suggested (on the basis of a rather simpler model than King's) for Canadian provinces by Miller (1979, p. 13).
27 See, for example, Engermann (1965, p. 53).
28 See Dafflon (1977, pp. 50–3) for a discussion of this issue.
29 For discussions of tax competition, see Browning and Browning (1979, pp. 416–17), Grewal and Mathews (1977) and Oates (1972, pp. 142–3). For a survey of studies (made in the United States) of the extent to which subcentral tax rates actually influence firms' location decisions, see Due (1961).
30 See HMSO (1981a, pp. 17–20) for a list of their services.
31 See ACIR (1981b, pp. 5–8) for a list of these authorities' functions.
32 See Gold (1979, especially Chapter 1) for an outline of these.
33 For a discussion of instruments such as these, see Lawless (1980).

2

The Optimal Size of Subcentral Authorities

Introduction

In the previous chapter, there was a discussion about what functions might most suitably be performed by subcentral authorities. Once a decision has been made to allocate some functions to lower levels of government, the question arises of what sized authorities might be most appropriate. It is this question that is considered in this chapter. Most of the chapter is concerned with developing a theoretical model to illustrate the factors that should be borne in mind when this question is considered. The final two sections give some details about sizes in practice, and also about the complete restructuring of subcentral government in the United Kingdom in the 1960s and 1970s.

Three preliminary issues need to be mentioned before the theory is introduced. First, there is the meaning of size. Size can be measured in terms of area or population, but in the literature on this subject[1] it is almost invariably measured in terms of population. The model developed in this chapter is presented on the same basis. However, it is generally assumed that a decision to have larger areas (that is, more populous ones) would be implemented by extending their boundaries, and having fewer authorities in consequence, rather than by depopulating some authorities and moving the citizens from those areas to the others.

Secondly, something must be said about the context in which these models might be used by policy-makers. Generally, the discussion will assume that a fresh start is being made, in that the entire structure of subcentral authorities and the allocation of functions between them has come up for review. In such a case, an analysis can be made for each function to consider the optimal size of unit by which it should be provided. Such analyses may be considered the first stage of the exercise. They are likely to suggest a different size of unit for each function. It will be shown, however, that a proliferation of possibly overlapping units may not be the optimal arrangement, and the second stage of the exercise consists of considering whether a smaller number of authorities, each providing several functions, should be established instead, even though most, if not all, functions would accordingly be provided by authorities whose sizes differed from the optimum to a greater or lesser extent. Such a wide-seeking review is rare, although one was conducted in the United Kingdom in the 1960s, and this is discussed in the final section of this chapter. It is important to note, however, that the model outlined in this chapter is useful on other occasions too. For example, even where a structure of authorities exists

and where a major reform is not desired, there may arise the issues of which tier would be most suitable for a new function, or whether one function should be transferred from one tier to another. The model in this chapter, which attempts to show how the optimal size of unit may be determined, also highlights the factors needed to select a second-best size from the limited range of existing sizes that may be available. Further, if boundary changes were being considered for areas in a case where the broad allocation of functions was to remain unchanged, then the model should help determine the optimal size of the new areas.

Thirdly, it should be stressed that there are a number of alternative models in the literature. At first sight, some of these appear to differ a great deal. However, the models have in common the broad proposition that, in a country divided into optimal-sized areas, the various welfare gains from any subsequent changes in size would fall short of the losses. The differences between the models tend to result from differing assumptions about what the relevant gains and losses are. This chapter will attempt to show how varying assumptions of these types should be handled.

Most discussions of the optimal size for subcentral authorities begin by assuming that the authorities will provide a single local public good, and then distinguish between two possible types. Both types are held to be consumed in equal quantities by all individuals in a particular part of the nation, and the level of consumption for each individual is held to be independent of the number of other individuals in the area. The difference between the two types is that, with one, the size of the area in which benefiting citizens can live is constrained by technological considerations, while with the other there are no such constraints. An example of the former type might be a flood control scheme designed to prevent the flooding of a coastal town in times of adverse tide and weather conditions; in this case it might be possible to protect the whole town but not feasible to protect some parts only. An example of the latter, more common, type might be a public lighting scheme, which could readily be extended or contracted to benefit a whole town or any part of it as desired.

In this chapter, both these types of local public good will be considered. However, the situation where technological factors constrain the area over which the good produces benefits is taken to be a special case and is left for later discussion, as are three other types of subcentral function. First, there are those impure public goods where consumption may vary between individuals, either because benefits depend on consumers' locations (for instance, the distance between their houses and the fire station), or on the extent to which they choose to use publicly provided facilities (such as roads). Secondly, there are those services of subcentral authorities that are not clearly public goods at all. These include any regional development functions they may have, any redistributive functions they may have, whether exercised through taxes and transfer payments or through taxes and the provision of relevant merit goods, any powers they may have over other merit goods or over demerit goods, any public enterprises they may run, any services provided in attempts to correct for externalities, and any local monopoly control functions and law-making functions that they

may exercise; they also include any functions performed on an agency basis for the central government. Thirdly, there are subcentral authorities' revenue-raising functions.

For the moment, all these special cases are set aside, and a model is now described to outline the factors that determine the optimal sizes for authorities charged with providing pure local public goods. It is supposed that subcentral authorities are to be entrusted with the provision of a number of such goods, and that the problem is to ascertain the optimal number of tiers of authorities, and the optimal size of authority in each tier. In order to build the model up, a number of further assumptions will be made to begin with, and then successively released. These are:

- all citizens have similar tastes and incomes;
- citizens incur no costs in attempting to secure the level and type of public service they desire;
- governments incur no administration costs;
- governments do not co-ordinate with each other;
- citizens are geographically immobile;
- the authorities that provide a public service also produce that service;
- only one public service, street lighting, is entrusted to subcentral authorities;
- authority boundaries are not constrained by geographical features such as rivers and mountain ranges;
- no part of the country is dominated by particular ethnic, religious or cultural groups with particular tastes; and
- the population density in the country is broadly uniform.

It will be assumed throughout that the method of accomplishing the task of determining the optimal-size unit will be to do a cost-benefit analysis of each possible size to see which produces the highest net gain. Since the bulk of the costs and benefits (or losses and gains as they will be called) are likely to be constant from year to year, the result will not be particularly sensitive to the choice of discount rate; hence the issue of what that rate should be is not discussed. It should be stressed that, in any event, quantifying the gains and losses would be exceptionally difficult, and the model is presented principally to show what factors should be taken into account rather than in the belief that it could be rigorously applied.

The effects of the major gains from collective provision

In the absence of public lighting provision, individuals might well provide some lighting outside their own homes. However, collective public provision offers two possible advantages. First, if there are economies of scale in the provision of lighting, then it may enable individuals to obtain any particular level of lighting outside their homes more cheaply. Secondly, when individuals provide lighting on an individual basis, then they are likely to ignore any external benefits to passers-by; hence individual

provision is likely to lead to underprovision. When a subcentral authority is formed to provide the service, however, then many (if not all) of these external effects will be internalised and the level of service provision is likely to become closer to the optimum as a result. It is true that, even with individual provision, bargains between individuals could, in principle, result in externalities being internalised and an optimal level of provision being attained, but in practice this is unlikely to occur as a result of the substantial organisational costs that would be incurred.[2]

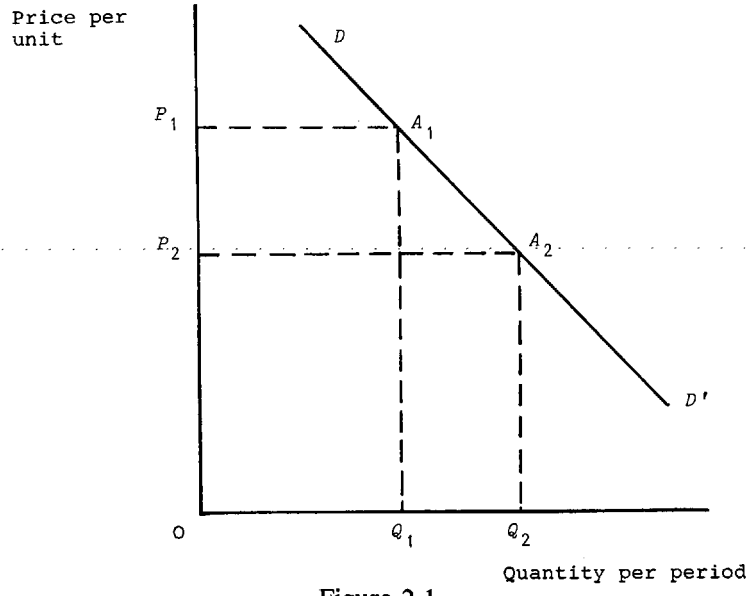

Figure 2.1

Figure 2.1 shows the first benefit as it affects any one individual.[3] The assumptions of uniform tastes and incomes mean that the situation would be the same for all individuals. DD' shows the individual's demand curve for lighting outside his home. OP_1 shows the price that a unit of lighting would have if individual provision were the rule. In that case, the individual would consume OQ_1 units. If subcentral authorities of populations, say, N_1 were formed, however, then the unit price per person might fall to OP_2. It is assumed, for simplicity, that the costs of public lighting would be shared so that each citizen faced the same price per unit, but if costs were shared in another way then OP_2 would be different for different citizens. With the new price of OP_2, the desired consumption by each citizen would rise to OQ_2. Actual consumption would now be decided by by group decisions, but since all group members in all groups would have similar tastes and incomes, and hence all would desire OQ_2 units, it is reasonable to suppose that each individual would be able to consume this quantity. The rise in the individual's welfare is the incremental consumer surplus of $P_1P_2A_2A_1$. The total rise in welfare for the country in any period would be this amount times the population. Figure 2.2 reflects the present value of this total for all future periods in the line OG_1, which

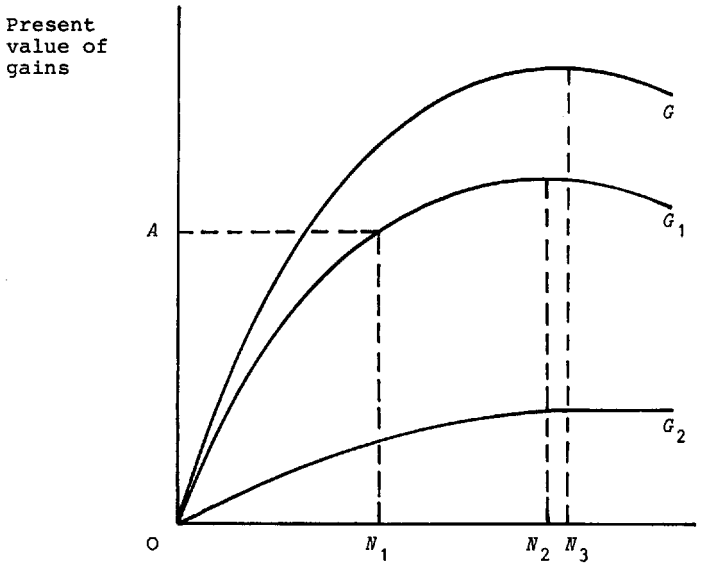

Figure 2.2

indicates the total gain resulting from unit costs falls for each possible authority size. For example, with group populations of N_1, indicated by ON_1 on the figure, the total gain for the country's population is OA.

It should be stressed that economies of scale may include both production economies and managerial economies. The latter have been discussed by Foster *et al.* (1980, pp. 512–13 and 573–4). Such economies would arise with public provision if larger authorities needed to employ proportionately fewer managers, or if they were able to attract higher-quality representatives or employ higher-quality officials. They could also arise if one function of management, namely inspection, were conducted by the central government. This inspection might include examinations of the expenditure control and project appraisal techniques used by the authorities, and it would be subject to economies of scale if the inspectors found it easier to inspect a small number of large authorities than a large number of small ones.

However, these cost reduction gains are not the only gains. Individuals also benefit as size rises because externalities are internalised and lighting provision approaches the optimal level. The present value of these gains is indicated in Figure 2.2 by the line OG_2. The total gains from both causes are shown by the line OG, which is a vertical summation of OG_1 and OG_2.

The shapes of OG_1 and OG_2 will depend on the particular service in question. In this example, it is supposed that economies of scale cease, or at least are outweighed by diseconomies of scale (possibly managerial diseconomies), after a certain point, and as a result OG_1 eventually (after ON_2) slopes downwards. Some research by McMillan, Wilson and Arthur (1981) noted in Chapter 1 suggests that economies of scale may be exhausted at low populations for some services, for example the

population served by a single fire station in the case of fire protection.

It should be noted, though, that OG_1 is derived from the gains to individuals, which depend on the price offered as a result of group provision, and it should be realised that, in reality, this price in an authority of a given size will not depend solely on the extent to which a group of that size can exploit economies of scale. It was suggested in Chapter 1 that the price offered may be further reduced as a result of experiment and innovation; it is possible that such reductions might be more significant for small areas than for large ones. The price could also depend on the effectiveness of the expenditure control and project appraisal techniques used, and it is possible that this might be greater in larger authorities. Thus the shape of OG_1 will be affected by the extent to which authority size affects these other factors.

As for the shape of OG_2, it is assumed that the benefits from internalising externalities eventually cease to increase. It might be supposed that once the lighting in any street had reached a certain high standard, then the standard would not be raised even if the authority responsible for it were increased in size. As a result, OG_2 eventually (after ON_3) becomes horizontal; OG in turn stays a constant distance above OG_1.

One important aspect of the model illustrated in Figure 2.2 should be noted, namely that the gains from collective provision could change if the general level of demand for lighting changed; consequently, the positions of OG_1, OG_2 and so OG could also change, and this in turn would alter the optimal authority size. The effects of a change in demand on OG_1 are of particular interest. Suppose, for example, that the level of service provision is measured in units, which can be defined in terms of hundreds of lampposts per 1,000 people, and that the annual costs of different numbers of lampposts are as shown in Table 2.1 (in an undefined currency unit). These costs would reflect the total cost of provision if it were assumed that the initial capital outlay was met by a loan and that the figures in the table included the interest and repayment costs of the loan as well as all running costs. The total costs could be the same each year if the loan were repaid over the whole lifespan of the lampposts in such a way that the combined repayment and interest payments were equal in each year. This system of repayment would be consistent with the borrowing rules suggested in Chapter 8 if the benefits from the lighting were equal in each year. It will be seen that it is imagined that it would cost rather more to have a certain number of lampposts, say 1,000, if they were scattered widely to serve 4,000 people, than if they were placed close together to serve 1,000.

Table 2.2 shows the unit costs per resident of various levels of service provision. For instance, if three units of output were required, that is 300 lampposts per 1,000 residents, and if the authority had a population of 3,000, then Table 2.1 shows that the total cost of the necessary 900 lampposts would be 16,200; this is 5.4 per resident, which is 1.8 per unit of lighting supplied, as shown in Table 2.2.

The desired population in terms of cost savings alone occurs where unit costs are minimised, and Table 2.2 shows that this population will decrease as service levels rise. This is because the desired size is one that results

Table 2.1 *Total lamppost costs*

Number of lampposts	*Size of authority*			
	1,000	*2,000*	*3,000*	*4,000*
100	2,500	2,600	2,700	2,800
200	4,000	4,200	4,400	4,600
300	4,800	5,100	5,400	5,700
400	5,200	5,600	6,000	6,400
500	5,500	6,000	6,500	7,000
600	6,000	6,600	7,200	7,800
700	7,700	8,400	9,100	9,800
800	10,400	11,200	12,000	12,800
900	14,400	15,300	16,200	17,100
1,000	20,000	21,000	22,000	23,000
1,100	27,500	28,600	29,700	30,800
1,200	37,200	38,400	39,600	40,800
1,300	49,400	50,700	52,000	53,300
1,400	64,400	65,800	67,200	68,600
1,500	82,500	84,000	85,500	87,000
1,600	104,000	105,600	107,200	108,800

Table 2.2 *Unit lighting costs per citizen*

Number of lampposts per 1,000 citizens	*Size of authority*			
	1,000	*2,000*	*3,000*	*4,000*
100	2.5	2.1	1.8	1.6
200	2.0	1.4	1.2	1.6
300	1.6	1.1	1.8	3.4
400	1.3	1.4	3.3	6.8

in lampposts being supplied at minimum average cost. Now, if the average cost were broadly to fall and then rise with the quantity of lampposts, then there would be some quantity at which the average cost was minimised; in the present example it is 600 lampposts, and 600 lampposts will be required by a larger area when standards and demand are low, and by a smaller one when they are high. Accordingly, OG_1 in Figure 2.2, and hence OG, would reach their maximum points at smaller populations if service standards were to rise in response to a change in residents' tastes or incomes.[4] In fact, successive increases in demand might cause OG_1 to change successively from OG_1 to OG_1' and OG_1'' as shown in Figure 2.3, thus increasing the relative attractiveness of smaller areas. If diseconomies of scale did not occur, and if, instead, economies of scale persisted for all increases in population, then OG_1 would always slope upwards. Its position would still depend on demand, however. The higher the level of demand, the greater the increase in consumer surplus would be as unit prices fell with increasing population. Thus, in this case, successive increases in demand might cause OG_1 to change successively to OG_1' and OG_1'' as shown in Figure 2.4, and the result is to make large areas even more attractive.

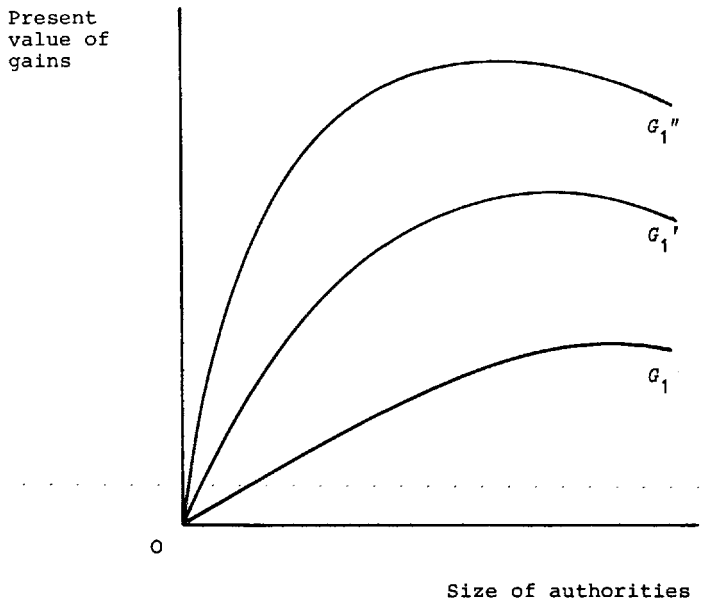

Figure 2.3

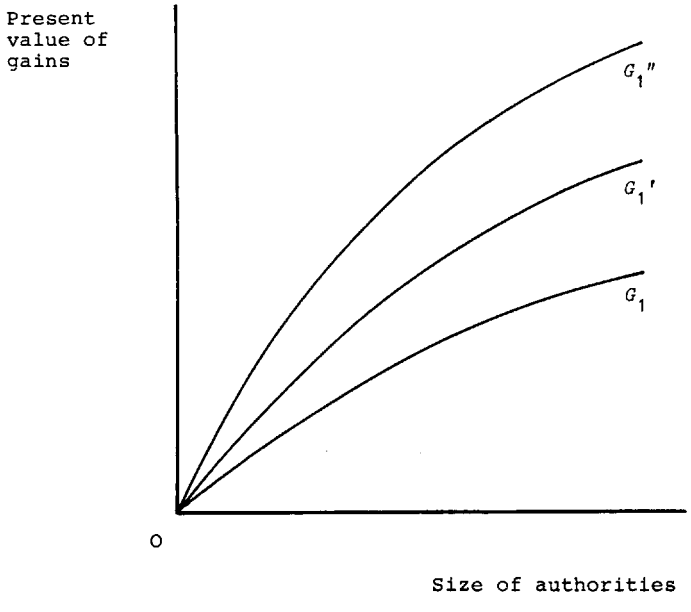

Figure 2.4

The effect of the major loss from collective provision

It is sometimes suggested that Figure 2.2 embodies all the relevant information needed to ascertain the optimal size of authority. This is true under the assumptions so far made. These will now be released in turn. Consider, first, the effects of removing the assumptions of uniform tastes and incomes for all citizens. One implication of allowing tastes and incomes to vary is that the demand curves for different individuals will now vary, so that the quantity that each individual will have to consume at the lower price made possible by association with others will not usually coincide with the quantity he would choose himself. Figures 2.5 and 2.6 show the different demand curves (each labelled DD') for two individuals who find themselves constrained respectively to consume smaller and greater quantities of lighting than they would choose. In each case, collective provision has reduced the price from OP_1 to OP_2,[5] and increased the desired consumption from OQ_1 to OQ_2. OQ_2 is different in each case, reflecting the different demand curves that stem from different incomes and tastes. However, each individual must consume the same quantity of a public good, and the authority fixes this at OQ_3. Now, if each individual could choose his own consumption level, then each would gain $P_1P_2A_2A_1$, but the actual gain for the individuals considered in Figures 2.5 and 2.6 is less than this because they also suffer losses arising from the fact that they cannot now choose their own consumption levels. For the individual in Figure 2.5, the actual gain is $P_1P_2BCA_1$, which is A_2BC less than the $P_1P_2A_2A_1$ that he might hope to get with a price of OP_2; this loss arises because he has to consume $(OQ_2 - OQ_3)$ less units than he desires at the new price. For the individual in Figure 2.6, the actual gain is also A_2BC less than the $P_1P_2A_2A_1$ he might hope to get; this loss results from having to consume $(OQ_3 - OQ_2)$ units more than desired. Naturally, the aggregate value of these A_2BC losses for the citizens in any authority will be higher the more their demand curves vary.

It should be stressed that collective provision may result in losses for a different reason when tastes vary, in that citizens will have different views over the type of service provided as well as over its level. For example, some might prefer to have sodium lamps outside their homes while others might prefer mercury lamps, yet group provision is likely to result in a uniform system, at least for each street. These losses must be included along with the A_2BC losses discussed above.

It is assumed, of course, that citizens will try to reduce the size of their own losses by taking actions to influence the level and type of service provision. These actions, which could include voting, lobbying and joining pressure groups, may be termed signalling; such signalling has been discussed at length by Breton and Scott (1978, especially pp. 31–2 and 51–8). It is assumed here that the losses stemming from varying tastes and incomes are measured with respect to the level and type of service fixed by the authority when this signalling has occurred.

These losses must be incorporated into the model. It would be possible to redefine OG_1 to be the present value of all $P_1P_2A_2A_1$s (which will now, of course, vary in size from individual to individual) less the present value

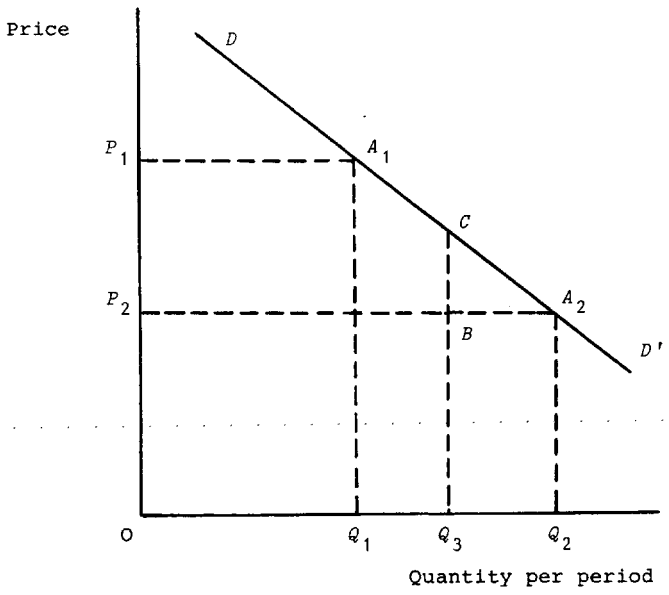

Figure 2.5

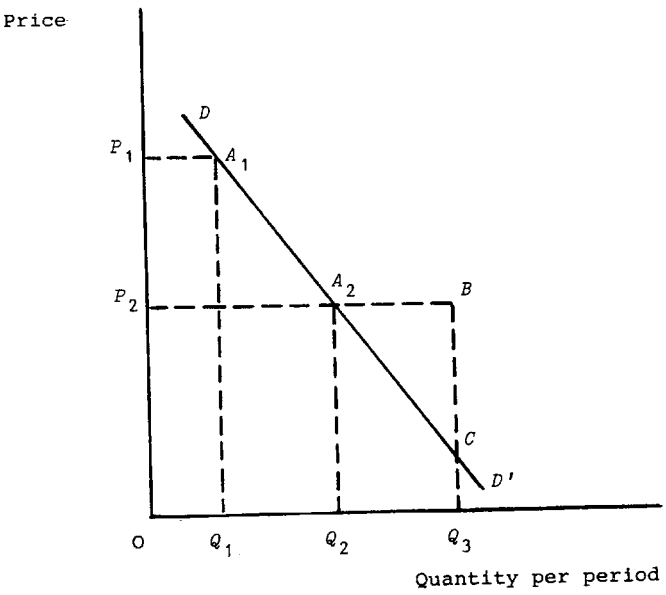

Figure 2.6

of all these losses. An alternative is to keep OG_1 as the present value of all $P_1P_2A_2A_1$s and to introduce a new line to show the present value of all the losses. This is done in Figures 2.7 and 2.8. OG_1 is the present value of all $P_1P_2A_2A_1$s and OG adds in the gains from internalising externalities. OG overstates the gains because the present value of all the losses has not been deducted. This is shown, for varying authority sizes, by the line LL. The net gain is therefore shown by the gap between OG and LL.

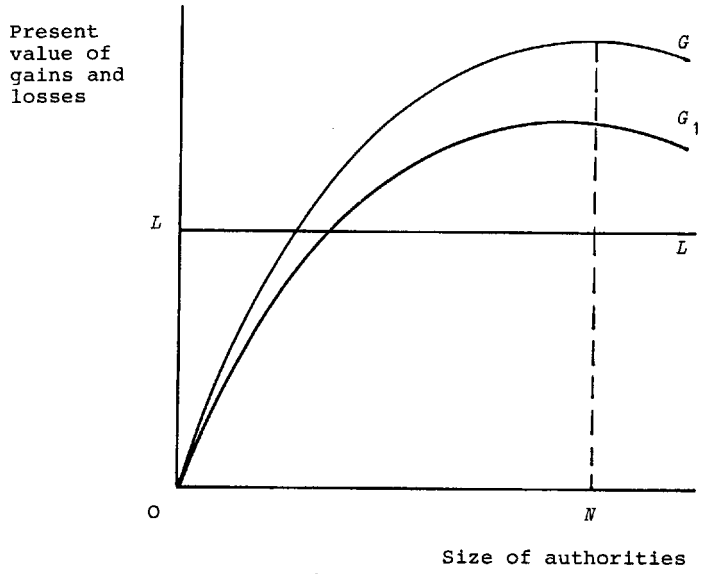

Figure 2.7

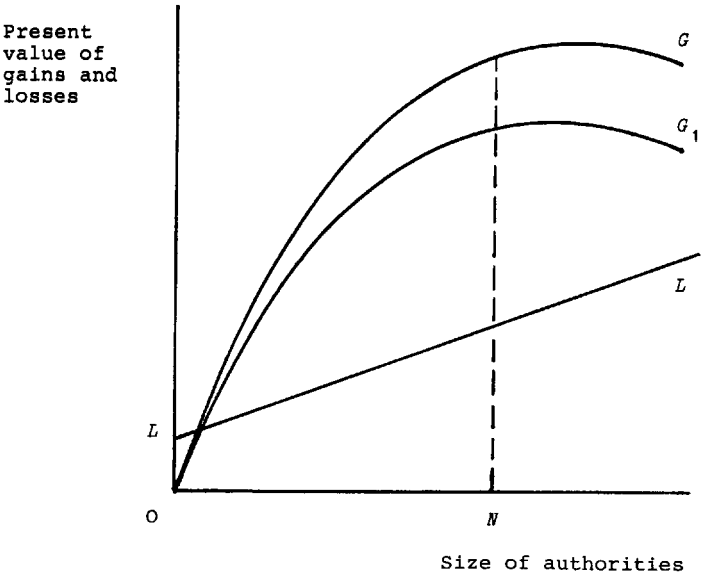

Figure 2.8

Figures 2.7 and 2.8 indicate two possible shapes for the line *LL*. These shapes demand some consideration. Suppose, first, that although tastes vary significantly between individuals the distribution of tastes does not vary significantly between different areas. To simplify the discussion, suppose that all citizens prefer one type of lamp and that only three different levels of consumption are ever desired – namely, high, medium and low – and suppose that in every street the numbers of individuals desiring each level are equal. In such a case, it is likely that the medium level will be provided in every authority no matter how large or small the authorities may be. There is a loss of A_2BC for all those seeking high levels (as drawn in Figure 2.5) and for all those seeking low levels (as drawn in Figure 2.6), but these losses, and hence the aggregate loss, will not depend on authority sizes; this situation is reflected by the horizontal line *LL* in Figure 2.7 (arguably, this line might bend and approach the origin when the authority size is very small, say 1 or 2). In this case, the maximum net gain, found where $(OG - LL)$ is a maximum, is at the population ON where OG is a maximum, and the optimal size is not affected by the losses now under consideration.[6]

Conversely, suppose that, although the numbers seeking high, medium and low levels are equal in the country as a whole, the distribution of tastes does vary from area to area, and that some areas are dominated by those desiring low levels and some by those desiring high levels. If just one nationwide authority were established and if, as is likely, it were to set a medium level, then the total loss would be as high as it would be for any authority size in the last case. If, however, smaller authorities were established, then in areas dominated by high-level seekers, high levels would doubtless be set; in those areas the aggregate welfare loss would be less than it would with a nationwide medium level. Similarly, the loss would be less in areas dominated by low-level seekers if low levels emerged. The smaller the areas became, the more closely standards could be tailored to the patterns of wishes of their inhabitants, and the smaller the aggregate national loss would be. This situation is reflected in the curve *LL* in Figure 2.8; arguably, this line too might bend and approach the origin when the authority size is very small. In this case, the maximum net gain, found where $(OG - LL)$ is a maximum, is at a population to the left of the point where OG is a maximum. This suggests forcibly that, if tastes were to differ from area to area, then the optimal size would be below the size needed to exhaust the combined benefits from economies of scale and internalising externalities. Indeed, the optimal size could well be below the level needed to exhaust the benefits from either of these factors, a situation that is reflected in Figure 2.8. Furthermore, the more tastes varied between areas, the steeper *LL* would be, and, in turn, the smaller the optimal authority size would be.

In fact, the analysis embodied in these figures and later diagrams may lead to rather bolder results than an optimal subcentral authority size. For example, the service concerned may be subject to inexhaustible economies of scale causing OG_1 to be shaped like those shown in Figure 2.4. In that case, it is possible that the gap between the resulting OG and

LL would continue to widen as the authority size rose, even if *LL* had a positive slope. If so, then national provision, or possibly supranational provision, would seem called for. Conversely, *LL* may be so high or so steep, and *OG* so low or so flat, that *OG* never succeeds in intersecting *LL*. In that case, there would always be a net loss with group provision, and individual provision would seem appropriate.

A quite different effect of allowing incomes to vary is that the distributional implications of various sizes of authority could become important. So far, the analysis has assumed that the authority size yielding the highest net gain is always the most desirable. In principle, this assumption is not necessarily valid, since the distributional implications of authorities of that size might be deemed undesirable. In practice, these implications would probably not be considered unless incomes varied, as now allowed. Perhaps the most serious possibility is that the model would suggest a small optimal size, which could result in some authorities being dominated by poor people and others being dominated by rich people. Areas of the former type would generally have lower standards or higher tax rates than areas of the latter type. It would clearly be possible to prefer a situation of larger authorities that yielded lower net gains from group provision but less unequal standards of public goods and less unequal tax rates. All the results in the rest of this chapter need to be qualified in this way. However, it should be stressed that the equity problems noted here can be overcome, whatever authority size is set, by means of a suitable equalisation scheme (as explained in Chapter 5), and there is clearly a case for ameliorating the distributional effects of the optimal size with at least some measure of equalisation rather than by compromising on the size itself.

Organisational aspects

Much of the literature on fiscal federalism assumes that Figures 2.7 and 2.8 embody sufficient information for determining the optimal authority size. However, these diagrams contain a number of assumptions that remain to be released. This section allows that citizens' signalling costs and governments' organisation costs may be positive.

Recent analysis by Breton and Scott (1978) has suggested that there are two types of organisation cost. First, there are administration costs that relate to activities internal to governments, such as the costs of setting up authorities and the costs of selecting, implementing and overseeing policies, which will include the search costs of ascertaining the preferences of citizens and selecting production methods. (The direct resource costs of producing public goods are excluded here since these have already been allowed for in OG_1.)

Secondly, there are the costs incurred in dealings between authorities. Unless authorities are large enough to internalise all externalities, there will always be scope for attempts to increase mutual welfare. It should be noted that the level of lighting provision in any area may affect citizens elsewhere, not only because some may pass through it at night, but also because some may one day end up living there. The co-ordination costs

incurred in bargaining between subcentral authorities should, of course, bring increased gains to citizens. Indeed, it is to be supposed that authorities engage in such co-ordination up to the point where the marginal benefits from increased co-ordination equal the marginal costs, a point that may occur when there are fewer outstanding externalities in the case of a few large authorities than in the case of many small ones. In practice, co-ordination costs are often borne by the central governments when they seek to encourage lower-tier authorities to take account of external benefits by the careful use of grants, as explained in Chapter 4. For the moment, it should simply be noted that, for the country as a whole, the major costs of such an exercise would be the organisational costs of the central government for the grants themselves essentially redistribute income.

To predict the influence of these costs or losses on the optimal size of authority, it is helpful to assess their effects on the *LL* curve in the diagrams. Broadly speaking, it might be expected that signalling costs would rise with the size of authorities, since the larger these bodies are, the greater the number of citizens there may be who find a divergence between their preferred output types and levels and their existing output types and levels. On the other hand, it might be expected that co-ordination costs would fall as the size of authorities rose, since there would be fewer authorities and hence fewer bodies to co-ordinate. As far as administration costs are concerned, it is arguable that any economies of scale there may be in actually establishing and running government bodies could be largely offset by diseconomies of scale in ascertaining citizens' preferences in large areas; these diseconomies may arise because preferences vary more in large areas, or because the greater distance between representatives and electors reduces the general understanding by the former of the wishes of the latter. Altogether then, allowing for the present value of these costs could result in a parallel shift upwards of *LL*, meaning that their combined addition to costs, which has to be seen as an addition to the losses accruing from collective provision, could be broadly equal irrespective of what size of assembly was established. Of course, it is possible that these costs would in aggregate rise or fall as authority size increases, in which case the new curve ($L'L'$) would be respectively steeper or flatter than the original (LL) as shown in Figures 2.9 and 2.10. (These figures assume that the original *LL* curve had a positive slope, but this assumption is not critical to the results that follow.) In the former case, there would be a tendency towards a smaller-sized optimal authority, and, in the latter case, a tendency towards a larger size. However, to ascertain the optimal size, attention must also be given to the present value of the benefits accruing from co-ordination. These benefits or gains can be included in Figures 2.9 and 2.10 by adding them to the gains indicated by the line *OG* to produce a new line $G'G'$, which indicates the combined value of all the gains. It is likely that the benefits from co-ordination will be less marked when authorities are large, since there will be fewer outstanding externalities to attend to. Accordingly, the new curve ($G'G'$) is generally closer to the original (OG) in Figures 2.9 and 2.10 when authority sizes are larger. Taking all these factors into account, the optimal size in these figures is shown as *ON*, and is smaller in Figure 2.9 than in Figure 2.10.

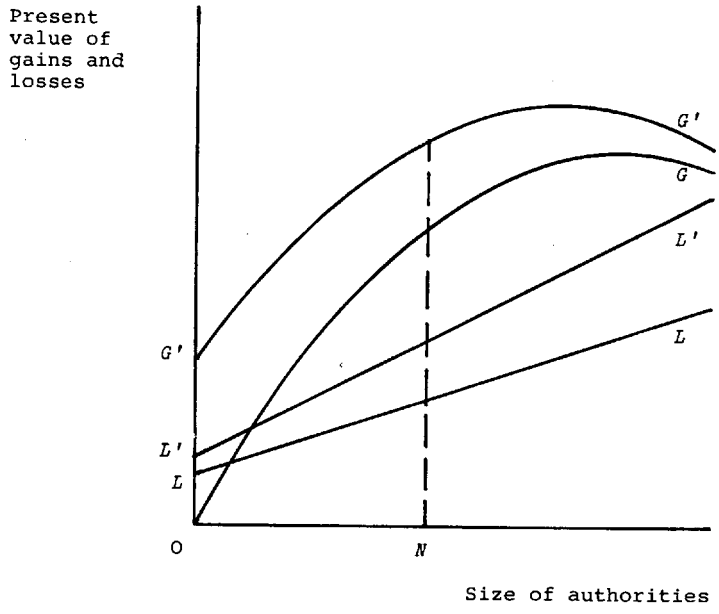

Figure 2.9

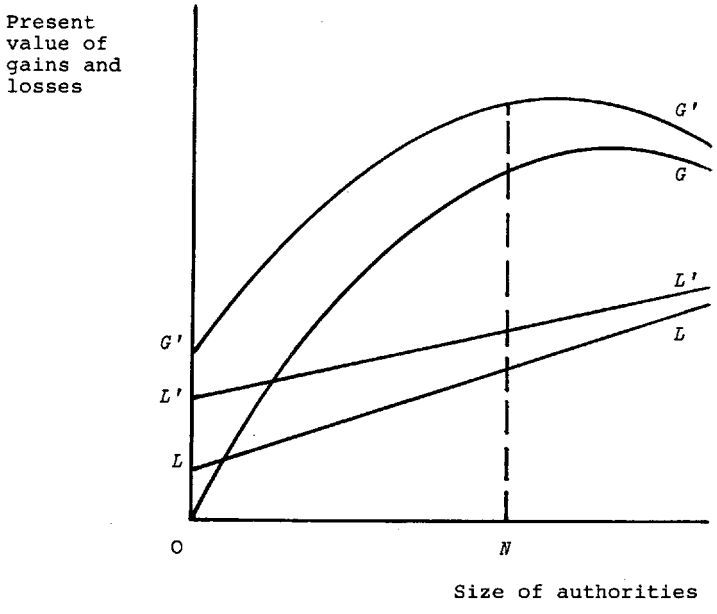

Figure 2.10

Consumer mobility

Another assumption that should be released is that citizens are geographically immobile. Allowing for consumer mobility introduces one particular conceptual problem to the analysis. The model has been designed to indicate the optimal size of authority in terms of population. It has been supposed that, once a set of authorities with optimal populations has been established, those populations will remain stable. However, if migration is permitted, and if migration can be induced by varying levels of subcentral taxes and public services, then it is clear that the initial populations may not remain stable. Accordingly, decisions about authority sizes must take migration expectations into account. It will be supposed that the object of the exercise is to divide part or all of a country up into authorities that will ultimately have optimal populations when the induced migration has ceased; the initial population of each authority may well be different from its final expected population.

It is useful to distinguish between two major aspects of subcentral government that might cause citizens to migrate. First, there could be a widespread desire for citizens to move from poor areas to rich ones in the hope of finding better services, lower tax rates, or both; it will be assumed that an equalisation grant scheme is in operation that prevents this from occurring. Secondly, citizens might migrate in the hope of finding areas whose levels and types of public services, and the associated levels of taxes, were closer to their preferences than the arrangements in their original areas. Some advantages of this Tiebout-type migration were outlined in the last chapter, and it will be supposed that such migration does occur. It effects are shown in Figure 2.11, where *GG* and *LL* represent

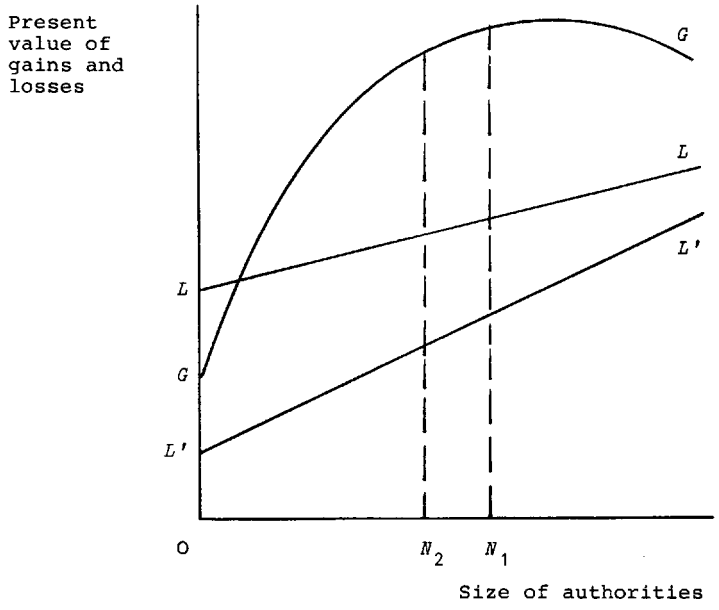

Figure 2.11

the total gains and losses from the factors so far considered (as embodied in $G'G'$ and $L'L'$ in Figures 2.9 and 2.10). The loss schedule LL has been based on the assumption that migration does not occur. The losses from collective provision could fall markedly, however, if citizens migrated to areas that offered them more attractive fiscal packages. This fall is indicated in Figure 2.11 by a downwards shift in LL to the new position shown by $L'L'$. The shift in LL should be made net of the present value of the migration costs that will be incurred, such as carriers' expenses. It should also allow for the external benefits and costs produced when citizens move (as discussed in the last chapter), which are unlikely to balance out exactly. Indeed, while citizens who move weigh up their private gains against their private losses and presumably feel they have an overall gain, it is possible that the external benefits and costs result in a net loss that could conceivably exceed the private gain.[7] It will be assumed here that there is a net gain, and hence that, overall, LL shifts downwards to $L'L'$.

As it happens, what matters for determining the optimal size is only whether $L'L'$ is steeper or flatter than LL. This is not easy to predict with certainty. It may be reasonable to suppose that migration will be more common when authorities are small than when they are large, because with smaller authorities the range of service levels might well be larger and so more people might be likely to find a move worthwhile; in turn, this suggests that the gross benefits from mobility will be greater when authorities are smaller. However, the total costs of mobility will be greater with more moves, so it is not certain that the net benefits will be higher with small authorities. Figure 2.11 reflects an arbitrary decision to assume that the average net gain from each move would be the same irrespective of how large the authorities were, and, hence, of how many moves occurred. Thus the total net gain will be higher with smaller authorities as there will be more moves. Its present value is also higher, so the new curve ($L'L'$) is further from the old (LL) for small-size authorities. In this case, then, allowing for migration reduces the optimal size of authority from ON_1 to ON_2.

Purchasing public services

A big impact on optimal size is likely if it is assumed that the authorities that provide the service do not necessarily produce it. In this case, the authorities may still determine the level of service production but are free to purchase it from a specialised producer. The possibility of purchasing replacing production has often been discussed, and it has been suggested by Tulloch (1969, p. 21) that only conservatism and organisational rigidities prevent the widespread purchase of services by governmental units of any size from organisations large enough to obtain the full benefit of any economies of scale that may exist.[8] In principle, such an arrangement should enable all authorities of whatever size to offer citizens the lowest technologically feasible unit price for, say, street lighting and hence achieve the maximum possible gain in this respect. If this were so, then the optimal size would depend only on the other factors.

This effect of purchasing can be illustrated with the help of Figure 2.12, where *gg* represents the total gains from factors so far considered (as embodied in *G'G'* in Figures 2.9 and 2.10) and *LL* the total losses (as embodied in *L'L'* in Figure 2.11). Og_1 shows the gains accruing from the lower unit prices attainable with group provision (as shown in the line OG_1 in Figure 2.2). The gap between Og_1 and *gg* reflects the benefits resulting from externalities being internalised within authorities (as embodied in OG_2 in Figure 2.2) and the benefits from co-ordination (as reflected by the gap between *Og* and *G' G'* in Figures 2.9 and 2.10). The analysis in this chapter has suggested that the benefits from the first of these factors tend generally to increase with authority size, while the benefits from the latter tend to decrease; Figure 2.12 assumes that the former factor is dominant so that the gap between Og_1 and *gg* increases with authority size, although in principle the latter factor could be dominant, in which case the gap would decrease with authority size. It can be seen that the gap between *gg* and *LL* is greatest with authority size ON_1, and this is the optimal size if authorities produce the service themselves.

Suppose, now, that authorities purchase the service from one or more private suppliers, each of which is large enough to offer it for sale at the lowest possible price. In that case, the maximum possible gain from scale economies should be available for authorities of all possible sizes, so that Og_1 is replaced by G_1G_1, and *gg* by *GG*; the gap between G_1G_1 and *GG* is equivalent to that between Og_1 and *gg*, and is therefore also assumed here to increase with authority size. The optimal size now occurs where the gap between *GG* and *LL* is greatest. In Figure 2.12, then, the optimal size with purchasing is ON_2, much less than the ON_1 with production by authorities since a major advantage of large size has now disappeared. A further

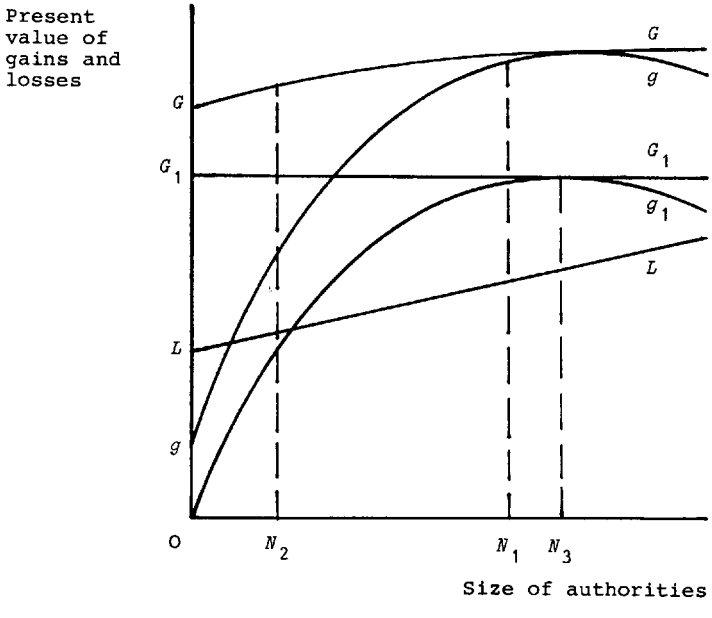

Figure 2.12

implication of Figure 2.12 is that when purchasing of services occurs then the net gain from subcentral authority activity may be less sensitive to small changes in authority size. Note that if the gap between Og_1 and *gg* decreased with authority size, then *GG* would have a negative slope and individual provision would seem to be optimal. This result could occur even with a *GG* curve that had a positive slope, if, for example, it was a straight line with a smaller slope than *LL*. The point is that, with purchasing, individual citizens might be able to arrange for a private producer to supply the service to them at the lowest possible cost, while any other gains from group provision could be more than offset by the losses.

It will be noticed in Figure 2.12 that citizens' gains from the economies of scale possible with public provision will always be greater if subcentral authorities purchase services, so that G_1G_1 is the relevant curve, than if they produce them, in which case Og_1 is the relevant curve, unless their authority has a population of ON_3. Bennett and Johnson (1980, p. 392) surveyed a wide range of studies of subcentral services in a number of countries and concluded that they showed, without exception, that the same level of output could be provided at substantially lower costs if output was produced by the private sector rather than by the public sector. They argued (pp. 365–6) that this was partly because subcentral authorities might be too small or too large to produce at the lowest costs (that is, they might not have the size ON_3) and partly because the managers of private sector firms were more likely to be stimulated by competitive forces to produce cheaply than managers of enterprises that were owned by subcentral authorities and operated on a local monopoly basis. This second point echoes that made by Borcherding (1977, p. 62), who suggested a 'Bureaucratic Rule of Two' whereby removal of an activity from the private to the public sector doubles the unit production cost. If that rule applies, then arguably G_1G_1 should be drawn much higher than the maximum point of Og_1, and *GG* in turn should also be drawn higher. G_1G_1 would nevertheless remain horizontal, so *GG* would still have the same slope at each population level and the optimal size would remain ON_2.

In practice, the relative infrequency of the purchasing of services may not always be the result of conservatism and organisational rigidities. It could be that many subcentral authorities are, in fact, large enough to produce their own services and still exhaust all, or at least most, possible economies of scale. Alternatively, it could be that there are problems with purchasing. Two possible problems will be considered here.

First, each small authority may be obliged to purchase all of its service from a single producer, as would almost certainly be the case with services such as water supply or policing, and this would put it in a position where it could suffer exploitation at the hands of a monopolist. Any price rises resulting from this sort of exploitation would shift G_1G_1 (and hence *GG*) down. For any authority of size ON_2, as in Figure 2.12, the monopolist could raise prices until *GG* intersected (an authority-producer's) *gg* above N_2; thereafter, the authority would begin to produce for itself. It would be understandable if authorities were more enthusiastic about purchasing services such as road maintenance and refuse collection, where monopoly exploitation would be harder since it would be easier for other competing

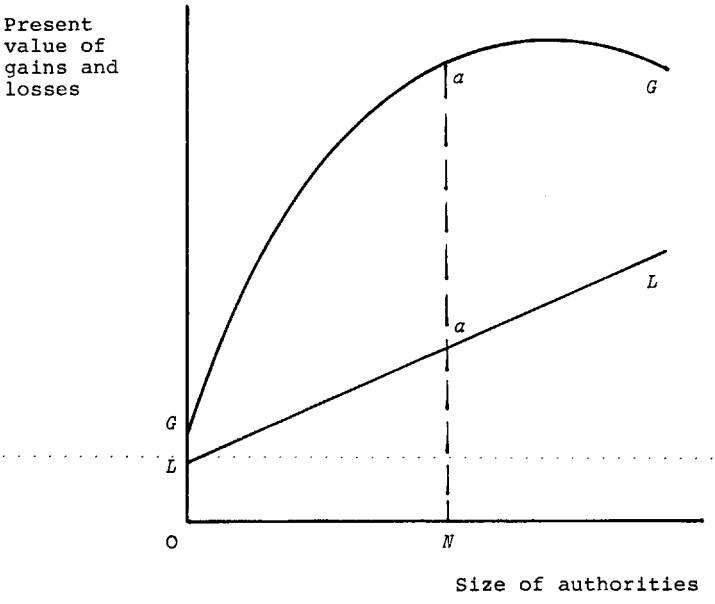

Figure 2.13

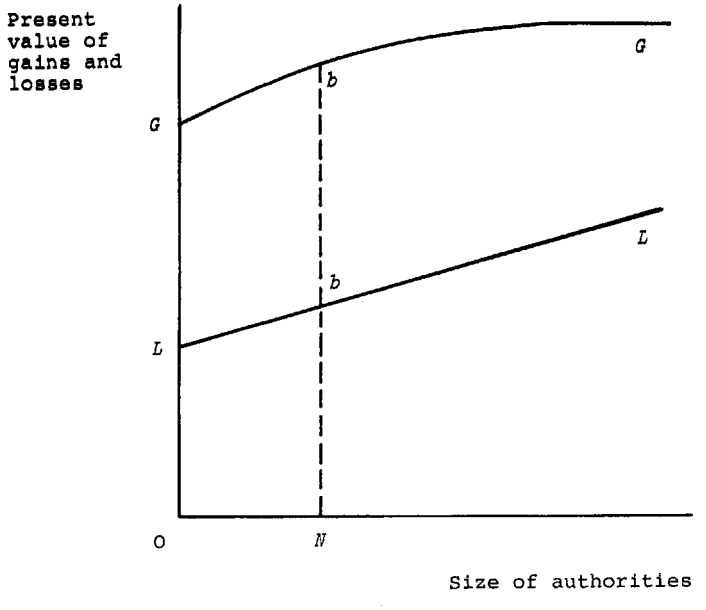

Figure 2.14

producers to enter the production industry. Of course, these are necessarily the services least susceptible to economies of scale, since such economies form a major barrier to entry; hence they are the services where the optimal size may be small even with production by authorities, and where a switch to purchasing might produce only a modest reduction in optimal size.

Secondly, it is possible that each authority might be less able to control the type or level of service consumed by its citizens. For example, a small authority purchasing water might have little or no scope for influencing the purity of the water, the additives in it, or the likelihood of restrictions on consumption during periods of dry weather. Thus, when an authority that is too small fully to exploit economies of scale considers a switch to purchasing, it will have to weigh up any gains from lower unit prices against the possible welfare losses resulting from types and levels of services being tailored less closely to the preferences of its citizens.

The possibility of purchasing should be borne in mind when authorities are established. The correct procedure is to estimate the optimal size and associated net benefits for production and purchasing respectively. Figures 2.13 and 2.14 illustrate possible outcomes, assuming no inefficient public enterprise managers in the former case and no monopoly exploitation by producers in the latter case. It is assumed that when authorities purchase, as in Figure 2.14, they have less control over services, and hence provision is less closely tailored to local preferences, so that *LL* is generally higher than in Figure 2.13. However, this problem is probably least serious when authorities are large and more influential relative to optimal-size producers. Hence the difference between the two *LL* curves diminishes as size increases; this is reflected by *LL* being steeper in Figure 2.13 than in Figure 2.14. In this example, the optimal size with production brings net benefits of *aa* in Figure 2.13, while the optimal size with purchasing brings smaller net benefits of *bb* in Figure 2.14. Thus authorities of size O*N*, as shown in Figure 2.13, should be established and should produce their own services.

In other cases, higher net benefits may result from requiring authorities to purchase services and establishing authorities of the correct size for that purpose. This will always be so if purchasing does not reduce local control over services and hence does not shift *LL*. This result can easily be seen by considering Figure 2.12. The gap between *GG* and *LL* at size ON_2 represents the maximum net gain obtainable from purchasing; the gap between *gg* and *LL* is always smaller, indicating that production can never realise comparable net benefits unless it lowers *LL*. The case against production would be further strengthened if public sector managers were less efficient than private sector ones so that *GG* in Figure 2.13 was much lower than *GG* in Figure 2.14.

Structures in cases of many subcentral services

The model developed above shows the principal factors that affect the optimal size of authority when only one service is to be provided by subcentral authorities.[9] In practice, many services may be supplied by these authorities. One approach to ascertaining the optimal system of

authorities in such a case would be to use the model for each service separately. This approach would be likely to indicate different optimal-sized authorities for most, if not all, services. It should be noted, of course, that some services may interact with others. For instance, the authorities providing personal social welfare services may be very concerned about the actions of the authorities concerned with housing. In such cases, individual authorities may wish to bargain with other authorities providing different services. The costs of this extra co-ordination, and the benefits it brings, should be included in the analysis at the same time as the costs and benefits of bargaining with authorities providing the same service in different areas (as discussed earlier).

It is possible, however, that total welfare might be increased by replacing an arrangement of a different set of authorities for each service either by one in which only a few tiers of authorities were established, and where each authority in any one tier was required to produce many services, or by one in which there were a few tiers of single-function authorities and one or more tiers of multi-function authorities; overlapping boundaries might be a marked feature of any system of the latter type. It is true that these alternative possible arrangements would mean that most, if not all, individual services would be provided in areas that were not of the optimal size. The resulting welfare losses would be least in situations where the optimal sizes were clustered around, say, two or three sizes, and where two or three tiers of authorities of those sizes were established. However, reducing the number of authorities could result in significant welfare gains. For instance, signalling costs might be very much lower, as citizens would have far fewer councillors and elections to bother about and they would never have to work out which authority was responsible for a particular service; co-ordination costs might be lower as there would be far fewer authorities to co-ordinate. Administration costs might also be lower; for example, there could be quasi-economies of scale if individual officials were able to oversee and plan different services.

On the other hand, it should be borne in mind that several reasons were noted in the last chapter for supposing that reducing the number of tiers might result in service levels moving away from those desired by electorates, thereby increasing the losses from collection provision. These reasons include the facts that fewer tiers might mean fewer opportunities for signalling of preferences by citizens, less effective control of public officials, and a reduced awareness by citizens of the costs of individual services. Thus, finding the optimal structure of authorities would, ideally, follow an analysis of each possible structure (including, perhaps, one with separate authorities for each service) to see which would offer the highest net gain. In such a context, the existing structure always has a headstart, because no initial organisation costs will be incurred. In practice, the analysis might be confined to those structures that seemed intuitively promising, for, as Breton and Scott (1980, pp. 25–6) have pointed out, analyses of this type are themselves costly. Note, too, that when deciding what sized authorities should be given a new service, or whether some existing service should be transferred to authorities with a different size, it may be found that the net gain from establishing any new set of seemingly

optimal-size authorities is, in fact, less than the net gain from giving the service concerned to existing authorities of not too dissimilar size if initial organisation costs could thereby be avoided.

Public goods whose benefits are confined to fixed areas

Throughout the development of the model, it has been assumed that there is no technological constraint that rigidly defines the area in which the citizens who will benefit from the provision of a service may live. Despite the frequency with which a situation in which there is this type of constraint is discussed in the literature,[10] it is most unusual for any services actually entrusted to subcentral authorities. Suppose, however, that a country, or part of a country, consists of a series of twelve villages, each with a population of 100, which lie along the seashore. Suppose, moreover, that the sea is prone to rise from time to time and flood these villages, and that it would be possible to build separate protection barriers for each village but not for individual parts of any village. In this example, each barrier might be held to benefit precisely 100 people. Now, it might be decided to divide the area up into authorities charged with providing barriers, and the level of service could be held to be a function of the height (and perhaps the durability) of these barriers. The question of an optimal-sized jurisdiction for such a service has been approached in two ways, and these will be considered in turn.

On the first approach, used by Oates (1972, p. 34), it is assumed that the population is geographically immobile. This means that the (average) population of the jurisdictions established can be altered only by changing the size of area over which each of them is responsible; in this sense, the approach is similar to that used for determining the optimal size of authority for public goods whose benefits are not constrained by technology to precisely defined areas. It has been suggested by Oates that, on this first approach, the optimal size of authority for providing barriers is, arguably, one with a population of 100; in this case, each authority would provide one barrier, and there would be what Oates termed a case of perfect correspondence between the citizens living in any jurisdiction and the citizens who benefited from the public good provided by that authority.

It is hard to imagine that smaller authorities would ever be set up, since if there were, say, four for each village, then there would be a situation in which four authorities existed to provide the same service for the same set of 100 people. In such a case, it is likely that the gains from mergers, which would include the benefits from internalising externalities, would far outweigh any losses that might ensue. Nevertheless, it is not in fact certain that mergers would be beneficial. It is possible that in each village the members of one group of 25 would willingly pay for some sort of barrier, even though it would benefit all other 75 citizens, whereas each of the other citizens would not be prepared to pay even 1 per cent of the cost of a significant barrier. In this case, the result of having four authorities might be a barrier in each village whose benefit to the 25

contributors far exceeded the cost to them, and which also provided some benefits to the 75 non-contributors. This would be a better result than having no barrier, which would be the likely outcome from single authorities of 100 citizens for each village.

As for larger authorities, it is likely that the only possibilities that would be considered would be ones containing all the citizens of two, three or more villages. There would be no point in this, though, since the example imagines no gains through cost saving (though in principle there might be economies of scale in respect of, say, research costs) or through internalising externalities (though such gains might be relevant if the erection of a barrier in one village were to change the flood risk for its neighbours), whereas there would be losses because each individual set of 100 people would have less control over the height of the barrier affecting them.

Of course, if the subcentral authorities in the country under discussion were entrusted with many services, then the advantages of having a few tiers, each concerned with several services, would mean that most services were provided by authorities of seemingly non-optimal size. Flood protection might well be one such service.

On the second approach, it is assumed at once that authority boundaries will be drawn so that there will be a perfect correspondence between the area of each jurisdiction and the area over which its public good creates benefits, but the assumption that the population is immobile is relaxed. In this situation, the population of any authority could change only if births and deaths were unequal over a period of time or if migration were to occur. The objective of this approach is not to throw light on what might be suitable authority boundaries, for these are taken as given, but to see what the optimal population for each area would be. The usual analysis of this question stems from the economic theory of clubs (developed originally by Buchanan, 1965) and considers the relative merits of allocating a given total population between a small number of highly populated areas, a large number of lightly populated areas, or various intermediate positions.

It is argued that the optimal population for each area is the one where the gains from adding one citizen precisely equal the losses. The main gain is that the per capita cost of public good provision will fall when extra citizens are added. The main losses are that congestion will rise, as noted by Musgrave and Musgrave (1976, pp. 616–17), and that total output might be lower if citizens were required to live and work in areas with higher populations, as noted by Atkinson and Stiglitz (1980, p. 522), perhaps because the marginal product of labour falls as more people occupy a fixed quantity of land. Thus, if the total population of 1,200 were to be allocated among twelve villages, and if the optimal population for a village was found to be 150, then 150 should be allocated to each of eight villages while four should be left empty. If the 1,200 people were already living in the villages, but not in groups of 150, then there would be a case for relocating people providing the gains from relocation exceeded the losses, which would include mobility costs. If, instead, the optimal population was 80, and if there was no space for more villages, then it would be

necessary for at least some villages to have excess populations. Now, as Atkinson and Stiglitz have observed (p. 551), an optimal distribution of population would not necessarily occur without intervention, partly because migrants would consider only the private gains and losses, and partly because a subcentral authority, unlike a club, could not easily control its size by refusing to admit new members or by offering tempting inducements to new entrants.

The principal policy implications of this second approach are that there may be in principle a case for encouraging some migration, though it is doubtful if, in practice, the gains would exceed the mobility costs. However, as far as the process of determining the number of tiers of subcentral authorities and their boundaries is concerned, this second approach offers no new advice, for it assumes that the optimal boundaries for authorities providing public goods whose benefits are confined to precisely defined areas will coincide with those areas; by contrast, the first approach sought to show that these were indeed the optimal boundaries.

Impure local public goods

The model presented in this chapter has assumed that each local public good provided by subcentral authorities is a pure one. In practice, many of the goods they provide may be impure ones. It is therefore necessary to see if the model used to determine the optimal-size authority for any one service needs modifying in such cases. Two types of impurity need to be considered.

First, there is the case where the benefits to individuals from a local public good depend on their locations. A typical example might be fire protection, since it is arguable that benefits are highest for those living near the fire station and taper off with distance. To simplify the discussion, it will be supposed that the task is to divide the country up into fire authorities, each of which will be presumed to run one fire station with one fire engine. In such a case, the analysis has to be modified at the earliest stage when welfare gains from economies of scale are considered, while tastes and incomes are assumed not to vary.

Suppose that all individuals have the same tastes and incomes, have similar residences, and make equal contributions in taxes to the costs of fire protection. Now, those living near the station will be likely to receive more effective protection than those at some distance. The situation for a citizen near the fire station is shown in Figure 2.15, and the situation for one at a distance is shown in Figure 2.16, assuming that quantity in these figures is directly related to the effectiveness of fire protection. The citizens' demand curves (labelled DD') are identical, and, in the absence of group provision, each would consume OQ_1 units of protection at a unit price of OP_1. With group provision, the citizen in Figure 2.15, who lives near the station, enjoys more protection than the citizen in Figure 2.16, who lives some distance away. The amounts they receive are shown as OQ_2 in each case, and reflect only the fact that the amount is higher in Figure 2.15. However, each pays the same tax, so the per unit price

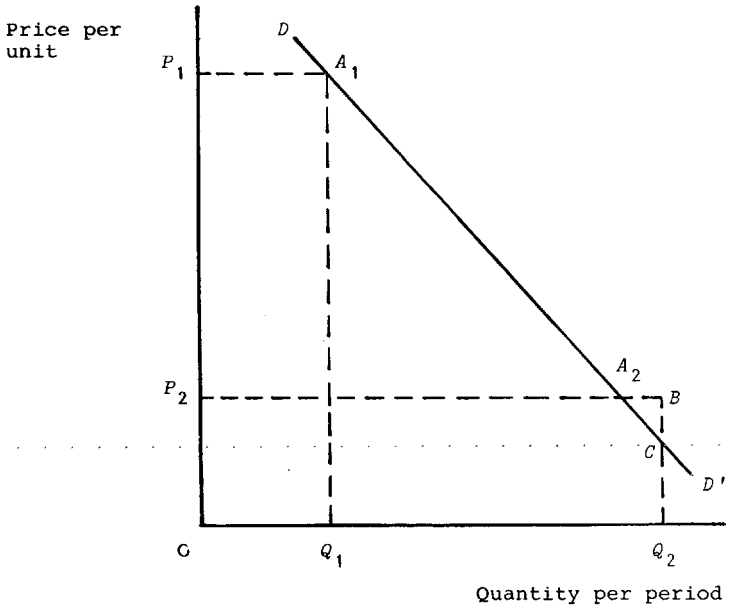

Figure 2.15

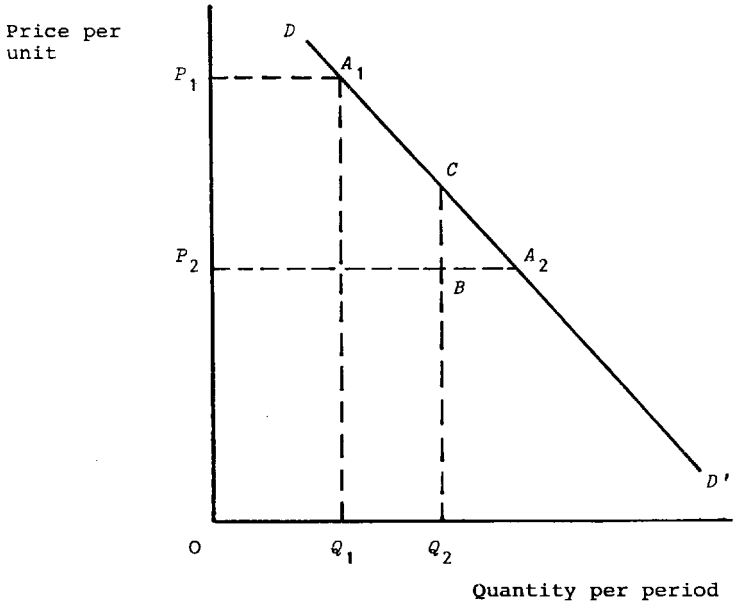

Figure 2.16

is higher in the latter case. The unit prices are shown as OP_2 in each case. It can be seen that group provision gives each citizen a gain of $P_1P_2A_2A_1$ less A_2BC. The present value of all such gains should be included in OG_1 of Figure 2.2. The rest of the analysis would proceed as before.

It is arguable that the difference in the level of fire protection might have capitalisation effects on property values, in which case the value of the property of the citizen in Figure 2.15 would exceed the value of the property of the citizen in Figure 2.16. If the authorities were to levy a property tax, then the citizen in Figure 2.15 would pay more tax than the one in Figure 2.16. This could be shown on the figures by raising P_2 in Figure 2.15 and reducing P_2 in Figure 2.16. Each area A_2BC would increase, so it appears to follow that the gains from group provision might be smaller with a property tax; it might be argued, though, that this disadvantage of a property tax would be offset by the fact that it results in tax payments being higher for those citizens enjoying higher service levels.

Secondly, there is the case where benefits to different individuals differ because they choose to consume different amounts. This situation simply reflects one effect that taste differences may produce; no modification to the model is needed, since taste differences have been taken into account. To see this, consider Figures 2.5 and 2.6 once again, and suppose that they reflect different citizens' demands for roads. The citizen represented by Figure 2.5 has the higher demand, and he may well use roads more than the citizen represented by Figure 2.6. Their demand curves reflect the different benefits they anticipate from road provision, and hence these different benefits are taken into account when the $P_1P_2A_2A_1$ and A_2BC areas for each citizen are estimated.

Functions other than local public goods

The model developed in this chapter has focused on determining the optimal size of authority for producing local public goods. In this context, the most widespread examples might be services such as street lighting, police forces, fire services and roads. However, as noted earlier, subcentral authorities might also be entrusted with some rather different types of responsibilities, and it remains to be seen whether the model needs to be modified to find the optimal sizes for these purposes.

The first of these responsibilities concerns the part subcentral authorities may play in stimulating regional development. Secondly, the authorities may be permitted to undertake some redistribution, either by a combination of taxes and transfer payments or by a combination of taxes and the provision of relevant merit goods. Thirdly, they may be allowed to stimulate the production of other merit goods, or to discourage or prohibit the production of demerit goods. Fourthly, they may be permitted either to operate public enterprises or to regulate local monopolies or oligopolies in services such as those provided by crematoria or public transport undertakings. Fifthly, they may attempt to correct market imperfections caused by externalities by themselves providing some services, such as refuse collection and disposal. Sixthly, they may have some law-making functions over matters as diverse

as public holiday dates, car parking regulations or effluent standards. Seventhly, they may perform some functions on an agency basis for the central government.

In each case, of course, the optimal size for authorities with a single function would be the one maximising the net welfare gains. The optimal sizes for each of these functions (and there could, for example, be many law-making functions and many public enterprises to run) would then be taken into account, together with the sizes for all other functions, when determining the final number of tiers and the size of authorities in each tier. Unfortunately, the literature contains little explicit discussion of the factors that should be taken into account when considering the net gains and losses for most of these items, although Foster *et al.* (1980, p. 44) have pointed out that large authorities might be best for income redistribution, since they would help to ensure a wide disparity of incomes within each area. All the factors mentioned in this chapter will doubtless need to be examined. However, it is possible that economies of scale will have a major effect on the final result only in cases where subcentral authorities actually provide merit goods themselves, or provide services intended specifically to deal with externalities, or operate public enterprises, or act as central government agents. The costs of formulating policies for investment incentives, for redistribution by taxes and transfer payments, for influencing private consumption of merit and demerit goods, for regulating local monopolies and oligopolies, or for making laws, may all exhibit economies of scale, but these costs will be relatively small; the costs of implementing the policies, particularly those involving investment incentives, transfer payments or subsidies, may be large, but will not generally be subject to economies of scale.

This hypothesis suggests that the optimal size of group for the provision of many of these items will be affected more by the other gains possible with group provision, in particular the internalisation of externalities. This, in turn, means that large authorities may have significant advantages over small ones. For example, a small authority might be reluctant to spend large sums to encourage the development of a big factory in its area if much of the benefit were to spill over into adjacent areas; it might not wish to impose high taxes on any demerit good, such as tobacco, if the only effect were that its citizens made their purchases in other areas; it might not wish to have different public holidays from adjacent areas if that meant that children at its schools had different days off from their parents working in other areas; it might not wish to adopt strict parking rules if doing so were to drive shoppers elsewhere; and it might be harder for small authorities to persuade their richer citizens to contribute more generously to income maintenance policies than their friends and colleagues nearby. Nevertheless, there does seem substantial scope for more analysis of these matters.

Size and subcentral revenue

The analysis of this chapter has concentrated on how the optimal sizes for subcentral authorities are affected by the various services they seek

to perform. The sizes may also be affected by consideration of the ways in which the authorities acquire the finance to perform those services. There are perhaps four major points in this context. These can be most simply explained in the context of the graphical analysis used earlier for a single-function authority.

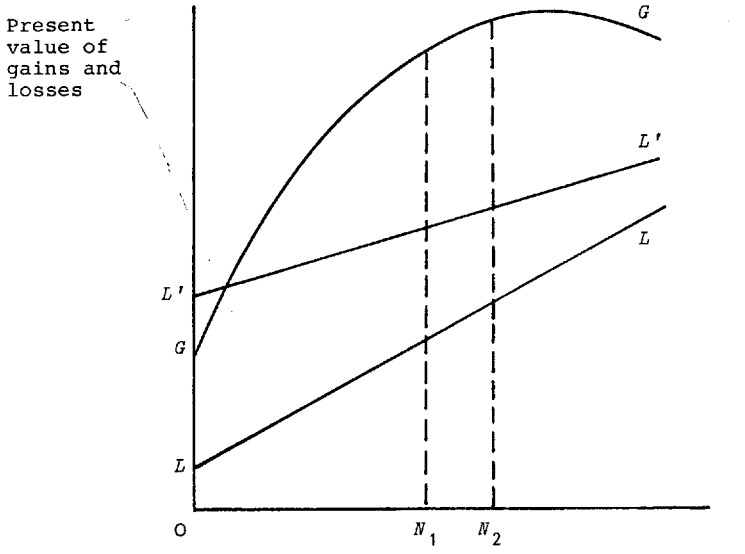

Figure 2.17

First, it is possible that there are economies of scale in the administration of tax collection. The costs of tax collection have so far been ignored, but they could be included in the *LL* schedule, which would be shifted up as a result. If total costs were to fall as area size rose, then the rise in *LL* would be more marked for small areas. This is shown in Figure 2.17, where *GG* and *LL* are equivalent to *GG* and *L'L'* in Figure 2.11; accordingly, *LL* shows the losses ignoring tax collection costs, while the new *L'L'* curve shows the losses including them. The result is an increase in optimal size from ON_1 to ON_2. Of course, if collection of subcentral taxes (even at varying rates) could be handled by the central government at a fixed total cost independent of authority size, then *L'L'* would be parallel to *LL* and the optimal size would be unaltered.

Secondly, it is possible that different tax rates in different areas could lead to inefficient moves of labour, trade and capital. For example, different rates of taxes on expenditure could encourage resource-wasting shopping expeditions, different rates of taxes on payrolls could encourage workers to seek employment in areas where their net pay was higher even if their contributions to output there were lower, and different rates of property tax could, in the long run, alter the supply of buildings in different areas. This problem is discussed at length in Chapter 6, but it may be observed here that it is generally likely to be more severe when areas are small and so migration is easier. The costs of these effects can be included in the *LL* schedule, which

will therefore (as in Figure 2.17) move up most for areas of small size. Thus consideration of this factor might also lead to an increase in the optimal size.

Thirdly, it is likely that tax exporting, which stimulates excessive levels of subcentral spending, will be more serious with certain taxes (such as a payroll tax or taxes on expenditure) when authorities are small than when they are big. The losses from inefficient spending could be included in the *LL* schedule, which again would move up further for areas of small size (as in Figure 2.17). This is, therefore, another factor that suggests an increase in the optimal size.

Fourthly, it is arguable that the model should take into account the costs incurred in negotiations over the formulae for equalisation and revenue-sharing grants, and that these should be included in *LL*. It is likely that these costs would be lower with large areas because there would be fewer parties to the negotiations. Furthermore, disparities in resources and needs might be smaller so that grantees' receipts would be marginally less sensitive to changes in the formulae and hence grantees might make less effort to have the formulae changed. Once again, therefore, *LL* might shift up most with areas of small size, and there would be another argument suggesting an increase in the optimal size.

Other factors affecting size

Notwithstanding the discussion in this chapter, it should be noted that, when a general review of the structure of subcentral authorities is undertaken, it is unlikely that equal-sized authorities will emerge. Those drawing up the constitutional boundaries may well use the points covered in this chapter to decide the ideal number of tiers, and to form a notion of the optimal-size authority in each tier. However, they will find that some boundaries have such a long history that proposals to replace them might meet stiff opposition, irrespective of any merits such proposals may have. They will also have to take note of matters such as geographical barriers (including rivers, mountain ranges and the sea around islands), which have been conveniently ignored. Further, some parts of a country may contain particular ethnic, religious or cultural groups with particular tastes; having authority boundaries coterminous with those areas may be sound in terms of reducing welfare losses from lack of control over service levels, even if, in general, authorities of such sizes are not optimal. Factors like these are likely to mean that for some individual authorities the optimal size is larger than the model suggests, and that for others it is smaller.

Another problem confronting those drawing up boundaries and determining authority populations may be that of varying population densities. This problem is important for services where the costs of a given level of service depend on population density. If the population density does vary between areas, then some will have higher tax rates or lower service levels than others unless the cost differences are fully allowed for in a system of equalisation grants; such full allowance is rare. However, those responsible for drawing up the boundaries may feel there is little they can do beyond making any boundary adjustments that could mitigate the problem, such as

giving each authority some urban and rural districts where possible. Of course, they might feel that some relocation of population to obtain areas of optimal density would be desirable. In fact, this will tend to occur anyway if tax rate or service level differences more than offset moving expenses and any reduction in wages. A related issue would arise if services were to cost more to provide in the sparsely populated rural parts of large authorities than in their densely populated urban parts. Here, gains might be achieved by encouraging a townwards population drift. However, this is more a matter for the subcentral authorities than for the boundary makers, and it is thus not a major issue in the context of this chapter.

Sizes in practice

It would be pleasant to conclude this chapter by showing how the structures of subcentral authorities in a number of countries have been determined with reference to the model outlined above. However, widesweeping reforms of these structures are rare, so that existing structures generally reflect the boundaries of history, along with the changes that have been politically feasible, more than they reflect a rigorous analysis of factors such as economies of scale and welfare losses through diminished individual control over services.

Nowhere, perhaps, is this more true than with the boundaries of the states of federal countries. These boundaries, if they change at all, change rarely, and state sizes were originally determined with limited reference, if any, to the functions that are now of most concern to their legislatures. Certainly the range of state sizes is very large.[11] In Australia, for instance, these varied in 1977 from some 4.9 million in New South Wales to 400,000 in Tasmania and 100,000 in Northern Territory; it was chiefly the smallness of the last that led to its attainment of self-government being deferred until 1979. The absolute range of sizes is even greater in Canada, varying in 1980 from 8.6 million in Ontario to 380,000 in Newfoundland–Labrador and 120,000 in Prince Edward Island; this excludes the Yukon Territory (around 20,000) and the Northwest Territories (around 40,000), which have been prevented from attaining provincial status chiefly on account of their sizes, although they have acquired a considerable degree of self-government, and their residents are expressing some desire to move towards provincial status. The absolute range of sizes is greater still in the United States, varying in 1978 from 22.3 million in California to little over 400,000 in Alaska and Wyoming, and being under 1 million in over a quarter of the states.

In proportionate terms, the largest Australian state has a population about forty-seven times that of the smallest, while a comparable figure for Canada would be sixty-nine (but a mere fifteen for the mainland provinces), and for the United States would be fifty-four. In contrast, the largest West German *Land*, North Rhine–Westphalia, with 17 million people in 1979, had a population just twenty-four times that of the smallest, Bremen, with 700,000. This smaller variation doubtless reflects in some measure the fact that the West German structure was imposed after World War II by the Allies, and hence is less the result of earlier history.

As far as the theory put forward in this chapter is concerned, these figures at least hint at the possibility that some states are above the optimal size and some are below. However, apart from the fact that geographical units such as islands often make obvious units irrespective of their size, a number of lines of defence might be raised against this proposition. First, it could be argued that at least some state services need optimal sizes expressed in terms of area rather than population; this, though, is not very persuasive, since state areas in Canada, the United States and West Germany vary even more in proportionate terms than populations.

Secondly, and more promisingly, it could be argued that many states are, in population terms, dominated by one large city (or urban area), which could not be split up for some services without a risk of losing important economies of scale or advantages from internalising externalities. The varying populations of these cities are thus a major explanatory factor of varying state sizes. This point is classically exemplified in Australia.

Thirdly, it could be argued that the *GG* and *LL* curves in diagrams analogous to Figure 2.13 would, for most services at least, be a roughly constant distance apart for a wide population range, even though both curves might have a positive slope over that range. This indicates that authorities with a wide range of sizes could still get aggregate national welfare fairly close to the maximum possible. In principle, this proposition should be empirically testable.

Fourthly, and perhaps somewhat provocatively, it might be suggested that economies of scale and other benefits are virtually exhausted by the smallest states (so that *GG* perhaps becomes horizontal at a low size), while inter-state mobility enables citizens to indulge in Tiebout migration to their preferred state, so that the aggregate loss from lack of individual control (as reflected by the line *LL*) is, in practice, negligible. In this case, there might be little advantage in merging smaller states or splitting larger ones. This point might well be strongest in countries where different states have the greatest differences in traditions or cultures, for there the range of tastes within each state might well be modest.

More generally, it could be pointed out that there is a large range of sizes for the new upper-tier authorities in the United Kingdom, even though these have been established specifically for their present functions. The reforms there are discussed in the next section, and they are evidence that any rational structure is, in practice, likely to have units of widely varying populations.

While state boundaries are largely determined by history, local government structures are perhaps rather less immutable, so it might be thought that these would offer more promising illustrations of theory being put into practice. Perhaps it could be shown, for example, that the trends, noted in the Introduction to this book, towards fewer, larger authorities in West Germany and the United States reflect either technological factors, which have led to greater possibilities for economies of scale, or a growing uniformity of tastes, which has reduced the case for small areas; or perhaps these changes have resulted from changes in the local authorities' functions. Nevertheless, it would be a challenging task to use the model to show why different states and provinces in the United States and Canada have different local government structures and widely varying styles of local authorities;

some counties in the United States, for example, have populations over 250,000 while many are below 2,500. In Australia, where local authority functions are roughly the same in each state, it might not be easy to explain why local authority populations vary from a mere 67 to 700,000, or why, in Western Australia alone, the populations range between 67 and 162,000, and the areas from 1 square kilometre to 380,000.

Another interesting task would be to see if the model could explain why there is a predominantly one-tier system of local government in Australia, a two-tier one in the United Kingdom and West Germany, while there are more complex arrangements in Canada and the United States; in the United States, for example, there are frequently three or four layers of local government and occasionally seven or eight. An interesting feature in the United States is the widespread use of some single-purpose authorities – notably school districts, but also authorities such as sanitary districts, soil conservation districts and drainage districts. The diversity of the arrangements between (and even within some) countries suggests that it is probable that the historical, geographical and other non-economic factors noted in the last section play an important part in determining local authority size, and that, as a result, detection of the economic factors discussed in detail earlier is not easy.

Reforms in the United Kingdom

The United Kindom might appear to afford a rather more promising chance to see the model in practice, for the entire structure of subcentral government there (except in Northern Ireland) was reformed between 1963 and 1975. The new structure contains a tier of counties (or regions in Scotland) and a tier of districts, and was set up with a cavalier disregard towards history, as is exemplified by the fact that virtually all the old county boundaries have disappeared, even though the counties they delineated were generally several centuries older than any of the states in the federations discussed in the previous section. Even so, the new arrangements are influenced by history to the extent that those in Scotland differ a little from those in England and Wales. The following paragraphs concentrate on England and Wales.

The new system in these areas has resulted in a diversity of populations that is probably rather less than is the case for local authorities in most countries. In 1980, the largest county, Greater London, had a population of 6.8 million, which was sixty-four times that of the smallest, Powys.[12] This proportionate range is not very different from the states of Australia or the United States, but it is much lower than comparable figures for their local authorities; in Australia, for instance, the largest authority has a population over 10,000 times that of the smallest. More revealing, perhaps, is the fact that 85 per cent of the English and Welsh counties have populations within the range of one-third to one and two-thirds of the mean, compared with only 40 per cent of Australian local authorities; and in the United States, where the mean size for counties in 1975 was just over 60,000, 11 per cent had populations exceeding 100,000 and about 50 per cent had populations under 10,000. Of

course, the relative homogeneity of local authorities in the United Kingdom may be partly explained by factors such as a relatively modest range of population densities between different parts of the country, but the recent reforms greatly reduced the diversity of size, and it is probable that there would be similar reductions in many other countries if they too undertook complete reforms of their systems.

Even so, the process leading up to the reform in the United Kingdom was perhaps less clearcut than might be expected.[13] Local government in the London area was reformed in 1963 in the wake of the report of a special commission. Two further commissions were set up in 1966 to consider, respectively, local government in the rest of England and in Scotland; these produced two different sets of proposals (HMSO, 1969a,b). The government (see HMSO, 1971a) introduced a modified version of the Scottish commission's proposals throughout the country, though it was modified rather less in Scotland than it was in England and Wales. The new two-tier system was introduced while yet another commission was considering whether some central government functions should be devolved to subcentral authorities. In the event, the functions it suggested might be devolved (see HMSO, 1973a, especially pp. 337–9) seemed to require larger authorities than the new counties; as it happens, there was little support for a new third tier of authorities, and no devolution has occurred. It is arguable that it would have been wiser to wait for this final commission's report before implementing any changes.

In terms of the model developed in this chapter, it has to be conceded that the two 1966 commissions conducted little research into economies of scale and virtually none into external benefits. They did tentatively explore the effect of size on education performance, education being the most costly local government service, and they felt that 250,000 or so was a satisfactory size for securing the highest standards. This might be taken to imply that economies of scale were exhausted at that point, which suggests that consideration of other factors, particularly welfare losses, might well have resulted in a lower size being regarded as optimal; in fact, the commissions paid little heed to welfare losses and concluded that 250,000 or so should be a minimum size rather than a maximum one. The commissions believed that housing and other social services should be provided by the same authorities as education, and, as these services between them accounted for over half total local authority expenditure, proposals might have been expected for some 200 authorities of about this size in England and Wales. These authorities could have formed a lower tier, for some other functions, such as police and fire protection, main roads and refuse disposal, were generally felt to require larger areas. In the event, the government introduced smaller lower-tier authorities with little explanation (HMSO, 1971a). It gave housing to these small districts, but felt, understandably, that they were generally too small (except in six major metropolitan areas) to be given education and the other social services. Thus these three services are not generally handled by a single tier, and education, arguably the most important service, is provided by about 100 large authorities instead of 200 medium-sized ones. Accordingly, it is far from clear that the existing arrangements are optimal in terms of the model developed in this chapter.

Summary

This chapter develops a model for showing how the optimal-size authority for any function may be determined. Broadly speaking, the approach is to assume that the function is initially provided by individuals, or individual households, at their own expense. It is then supposed that there would be both gains and losses to the country's citizens if individual provision were replaced by subcentral authority provision. The present value of the future gains and losses to all citizens would depend on the size of the authorities into which the country was divided, and the optimal size is the one yielding the highest net gain.

One major gain is that group provision could exploit the possibility of economies of scale and so enable services to be provided at a lower unit cost. Another is that many external benefits arising with individual provision, but very likely ignored when spending levels are determined, would become internal to the group and be taken into account by the authorities. The major loss is that individuals would be likely to have less control of service levels and specifications. There are further losses in the form of organisation costs, and possibly further gains if authorities co-ordinate in an attempt to have outstanding external effects taken into consideration.

Depending on the service concerned, the maximum possible gains from group provision may be attainable with relatively small areas, or unattainable even with national provision. However, the model outlined in this chapter suggests forcibly that, if tastes in a country vary from area to area, then the highest net gain occurs with an authority size below the one needed to secure the maximum gross gain. The optimal size occurs where any further gains resulting from an increase in size would be more than offset by the losses caused by different areas being less able to do things in their preferred ways. One important implication of this result is that authorities may be of optimal size even if they are too small fully to exploit economies of scale. Another implication is that, the more tastes vary between areas, the smaller the optimal size will be. An interesting feature of the model is that it can be used to show the circumstances in which individual, national or supranational provision may be better than subcentral provision.

The model shows also that the optimal size will alter, possibly upwards and possibly downwards, if demands for the service concerned grow; it suggests that the more people vote with their feet, the smaller the optimal size is likely to be. The possibility that authorities could purchase services from private suppliers is also considered. It appears that if this arrangement does not affect the ability of each area to choose the level and specification of its services, then the optimal size may be much smaller than with production by the authorities; if it does, then it may so reduce the ability of authorities to cater for varying tastes that production by the authorities themselves is the better option.

The chapter acknowledges that when subcentral authorities provide many services it may be unwise to have a separate tier of optimal-sized authorities for each function; rather, it may be best to have a few tiers,

each with an authority size reasonably suitable for a number of major services. It acknowledges, too, that authority boundaries will, in practice, be influenced by diverse matters such as history, geographical barriers, and ethnic and cultural groups.

Notes

1 See, for example, Oates (1972, Chapter 2), Tiebout (1961), Tulloch (1969), and, for an unorthodox approach, Breton and Scott (1978).
2 For a discussion of such a bargaining process, see Coase (1960).
3 The analysis at this stage closely follows that of Oates (1972, pp. 38–49). However, two important differences should be noted. First, the gains and losses in this chapter relate to all the country's citizens, and vary according to the size of groups into which the country is divided, while the gains and losses in Oates's model relate only to the citizens of one particular area. Secondly, the model in this chapter, unlike Oates's expresses gains and losses in present values.
4 This analysis derives from that given by Tiebout (1961, pp. 82–7). It would be possible to derive smooth average cost curves from Table 2.1 corresponding to those on Tiebout's diagrams. Tiebout's own table contains numbers that produce scalloped cost curves and, in his (arguably spurious) example, optimal size rises with service levels.
5 As explained in relation to Figure 2.1, a different OP_2 could apply to each individual. Note, too, that the price in any one area could be lowered by certain types of grant.
6 Although the discussion closely follows Oates (1972, pp. 43–4) at this point, it is not entirely clear whether his model yields this particular result.
7 It will be recalled that the principal external effects result from the possibilities that a migrant will alter the taxes that have to be paid in the areas he exits and enters, and might respectively ease and aggravate congestion in them; if these effects were internalised in the way noted in the last chapter, then they could be ignored.
8 For a further discussion of purchasing, see Breton and Scott (1978, p. 6), Grewal (1981a, pp. 13–15) and Foster *et al.* (1980, pp. 569–72).
9 Some other factors are discussed in later sections of this chapter. Factors not discussed in the text include the suggestion made by Foster *et al.* (1980, pp. 564–5) that small authorities might give more encouragement to able but shy citizens to stand for election. This latter advantage of small areas might be enhanced if those citizens subsequently become effective central government representatives.
10 See, for example, Atkinson and Stiglitz (1980, pp. 519–51), Musgrave and Musgrave (1976, pp. 615–23), Oates (1972, pp. 33–8) and Olson (1969).
11 Most of the information in this section is taken from ACIR (1981d, Chapter 2); that relating to Australian local authorities is taken from C. P. Harris (in ACIR, 1981c, pp. 39–43), and that relating to local authorities in the United States is taken from the United States Department of Commerce (1981, p. 298).
12 The United Kingdom figures are taken from HMSO (1982a, pp. 12–28). The Australian and United States figures come from the sources listed note 11.
13 For a discussion of this process and for proposals that embody more explicitly the theory in this chapter, see King (1980a).

3

Grants: Their Types and Effects

Introduction

In this chapter and those that follow, it is supposed that decisions have been made to allocate some functions to subcentral authorities, and that one or more tiers of authorities have been established. It is now necessary to decide how these authorities will be financed. Revenue sources for subcentral authorities can doubtless be classified in a number of ways, but one fundamental distinction is between grants, which are received from other governments, and independent sources of revenue, which are raised directly. The three principal independent sources are taxes, charges and loans, and these are examined in Chapters 6–8. Grants form the subject of the present chapter and of Chapters 4 and 5.

There are perhaps two major policy issues in formulating a satisfactory system of grants. First, it is necessary to determine the purposes for which grants are to be given, and, secondly, it is necessary to decide which type of grant is most appropriate for each purpose. These two issues are discussed in the next two chapters. However, it is not possible to select appropriate grants without a knowledge of the principal types available and without some understanding of how subcentral authority spending is likely to be affected by the different types. The present chapter is concerned with these matters.

It should be noted that grants are usually paid by a higher-tier authority to a lower-tier authority. The most frequent situations are grants from central governments to states and local authorities, and grants from states to local authorities. Other situations are possible, however. Sometimes grants may be made by some authorities in one tier to other authorities in the same tier. This occurs, for example, with the London Rate Equalisation Scheme in the United Kingdom. On other occasions, grants may be made by lower-tier authorities to higher-tier ones. Mathews (1980a, p. 23) has noted that in the People's Republic of China the budget of the central government is supported by financial transfers from the provinces. The effects of grants on the grantee should not depend on whether the grantor is a higher, equivalent or lower-tier authority. Indeed, some of the purposes of grants can, in principle, be satisfactorily achieved by grants that move upwards or across as well as by grants that move downwards, though others are unlikely to be achieved except by downwards grants.

Types of grants

The principal types of intergovernmental grants can be usefully described

with the help of Figure 3.1. It will be seen that the first major distinction is between specific grants and general grants. A specific grant is one whose proceeds must be spent on some activity stipulated by the grantor. Such grants are sometimes termed 'selective grants', since the grantor selects the purposes for which they may be used, and they are sometimes termed 'conditional grants', since the grantor makes conditions about how they may be spent. General grants, by contrast, are ones whose proceeds may be spent at the discretion of the grantees. These grants are sometimes termed 'non-selective grants' or 'unconditional grants'.

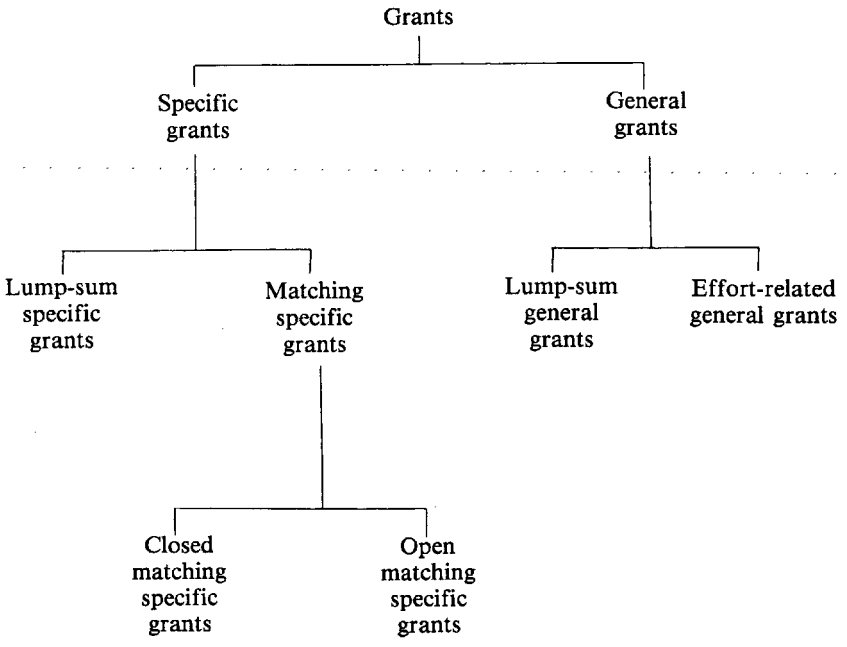

Figure 3.1

It will be appreciated that the conditions attached to a specific grant necessarily limit the manner in which it may be spent. In some cases, the limits may be narrow; in others they may be wide. Grewal (1981a, p. 33) has suggested that the term 'block grant' should be used where the limits are broad, as might be the case with a grant that was designed to assist school education, and that the term 'specific grant' should be reserved for cases where the grant was designed for a narrow purpose, such as improving the status of libraries in schools. However, there could be some confusion from using the term 'block grant' in this context, for the term has also been used to describe general grants (HMSO, 1975, pp. 21–2 and

44–5).[1] The term 'categorical grants' has sometimes been used to describe cases of specific grants where the limits are narrow (Mathews, 1980a, p. 26), but it is not always so used (Musgrave and Musgrave, 1976, p. 652). Perhaps less ambiguous terms would be 'wide specific grants' and 'narrow specific grants', though it should be recognised that there is really a spectrum of spending conditions rather than two polar positions.

Specific grants can be subdivided into lump-sum grants and matching grants. In the former case, the grantee receives a fixed sum of money to spend on the designated service, whereas in the latter case the sum received is variable and depends on how much of its own revenue the grantee spends on the service concerned. Lump-sum grants are sometimes referred to as non-matching grants, while matching grants are sometimes referred to as percentage grants. The term 'percentage grants' has arisen because the value of a matching grant may be defined as a certain percentage of the area's own expenditure on the relevant item: thus, a 50 per cent matching grant indicates that the grantee receives 50 pence of grant for every pound it spends from its own revenue. Matching grants can themselves be subdivided into closed or closed-ended grants and open or open-ended grants. In the latter case, the grant might be set at, say, 50 per cent of the grantee's relevant expenditure, no matter how high that is; in the former case, the grant might be established in the same way subject to an upper limit, so that the grant might be 50 per cent of the grantee's expenditure if that were below £1 million, and £500,000 if expenditure equalled or exceeded £1 million.

General grants may also be fixed in amount or they may depend on the actions of the grantee. The ambiguous term 'block grant' has been used to describe both these situations.[2] In this volume, the term 'lump-sum general grant' will be used where the amount of grant is fixed. Where the amount of grant is to depend on the actions of the grantee, it might be thought, by analogy with matching specific grants, that it would depend on the grantee's total expenditure from its own funds, which could include revenue from taxes, charges and loans, not to mention sales of assets and changes in balances. In practice, it is usually related to the amount of revenue the grantee secures from just one of these sources, namely taxes (though occasionally some allowance is made for revenue from charges). Since these grants depend on the grantee's revenue effort, the term 'effort-related grant' is used here. It is possible to have an effort-related grant where receipts are defined as a fixed percentage of the grantee's tax revenue, and possible to make such a grant open or closed. However, the relationship between grants and tax effort can be defined in a number of other ways. These are not shown in Figure 3.1, but two of them are discussed in Chapter 5.

The framework of the traditional models of grant effects

In order to determine a suitable type of grant for any particular purpose, it is important to know what effect each type of grant will have on the grantee, in particular on its expenditure. Traditionally, orthodox economic

models have been used for this purpose, and the analysis based on them forms the substance of this and the next four sections. However, there is some evidence that, in practice, grantees do not always behave as these models suggest, and some alternative models have recently been put forward that, it is claimed, make predictions that do square with the evidence. Some of these models are considered later in the chapter.

It should be stressed that all these models are primarily concerned with the initial direct effects of grants on the grantee. The subsequent indirect effects could, however, be very varied.[3] An increase in grants may, initially, cause an increase in spending by the grantee and a reduction in its tax rates. These responses will lead to changes in consumers' expenditure and saving, and also, perhaps, encourage higher- or lower-tier authorities to alter their tax rates and expenditures. Further, there may be changes in, for example, land values, household and business location choices, investment decisions, and traffic flows. Some of these induced changes could feed back to both grantor and grantee governments and cause new changes in their behaviour. Up to the present time, little work has been done to analyse the effects of grants in a general equilibrium framework.[4]

Economic models of the effects of grants are typically couched in an indifference curve framework in which indifference maps reflect preferences between units of one or more grantee functions and residual income to be spent on other items. However, there is some diversity in the indifference maps used. Sometimes (for example, Wilde, 1968) the indifference map is taken to reflect the preferences of the decision-making body of the grantee authority; sometimes (for example, Romer and Rosenthal, 1980) it is taken to reflect the preferences of the median voter;[5] and sometimes (for example, Boadway, 1979, p. 424) it is taken to reflect the preferences of all voters in a set of community indifference curves. Using the preferences of the grantee authority's decision-making body is, in principle, the procedure likely to give the best predictions of the effects of grants, but economic models tend to assume that this body will try to reflect the preferences of local citizens, so that this model becomes more like the other two. The positive predictions from the other two options are similar. As Romer and Rosenthal (1980, p. 459) have noted, the advantage of using the median voter's preferences is that the only restrictive assumptions needed are that the identity of the median voter is unaffected by changes in grants, and that the median voter is always the pivotal or decisive voter. Using community indifference curves, in contrast, requires the assumption noted by Wilde (1971, p. 144) that grants have no redistributional effects at all in order that the indifference map should be stable. Oates (1972, p. 75n) has observed, however, that the advantage of using community indifference curves is that, by reflecting the welfare of the community as a whole, they enable some normative conclusions to be reached as well as positive ones. For this reason, community indifference curves are used below.

A number of further assumptions need to be made in order to proceed with the analysis:

- community indifference curves are assumed to be convex;
- it is assumed that no corner solutions occur along the axes;[6]
- it is assumed that the prices of all grantee functions are unaffected by the quantities purchased;[7]
- grantees are assumed to be unable to export any part of their taxes, so that increases in their taxes must be wholly financed by their own citizens;
- grantees are assumed to provide goods and services alone, and not subsidies or transfer payments, so that increases in their taxes necessarily reduce the funds available for their citizens to spend on private goods and services;
- the reactions to grant changes by any one authority are assumed to have no external effects on others, so that, in turn, no feedback problems arise; and
- it is assumed that the introduction of any grant, whether specific or general, has no demonstration effect that would alter tastes in the direction of a particular grantee function or of grantee functions in general.

Generally speaking, a grantor may wish to use a grant to affect grantee expenditure on a particular function, or it may wish to use a grant to affect grantee expenditure in general. These two cases will be examined separately. In each case, it is initially assumed that grantee citizens pay no taxes to the grantor to help finance its grant to their authority.

Lump-sum grants

One way of stimulating provision of a particular service[8] is to use a lump-sum specific grant. This is illustrated (for a period in time) in Figure 3.2, where the horizontal axis measures quantities of the particular public good or service concerned (X). The vertical axis measures in pounds the resources available to the citizens of the grantee authority to spend on other public and private items or to save. These resources include all its citizens' incomes, less taxes paid to authorities other than the grantee, together with receipts of grants. In the absence of grants, the authority faces the budget line AA', whose slope reflects the exogenously determined price of X. The authority opts for the point J on indifference curve I where it purchases OX_1 units of X and retains OY_1 pounds for individual or authority expenditure on other items. Now, if the grantor were to introduce a lump-sum specific grant of value AB pounds, which was sufficient to purchase $A'B'$ or AC units of X, then the authority would face the budget line ACB'. It would opt for point K on indifference curve II and purchase OX_2 units of X, leaving OY_2 pounds for expenditure on other items. The grant would stimulate purchases of X so long as it was a normal good – that is, so long as the grantee's income consumption curve $OICC$ had a positive slope. It should be noted that the grant would not generally give the grantee a budget line of BCB'. If the authority were to purchase fewer than AC units of X, then it would probably not be eligible for the

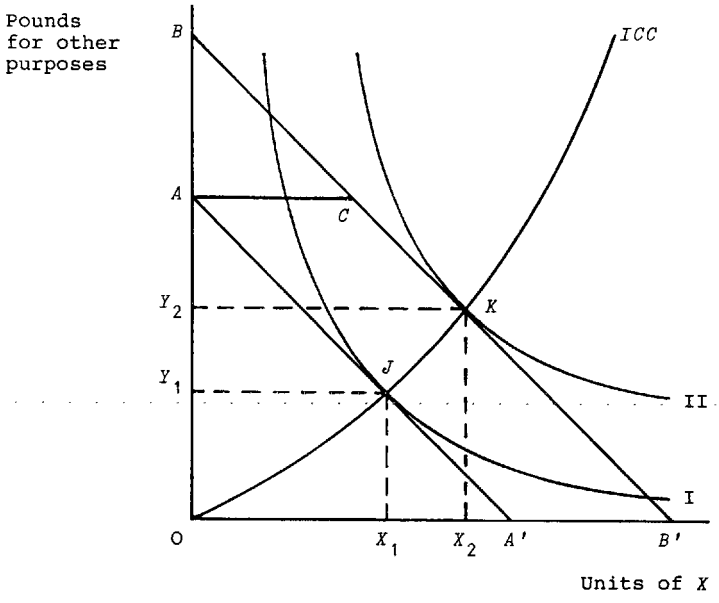

Figure 3.2

full grant; if it were to purchase no units of X at all, then it would probably receive no grant and would be at A rather than B.

It will be noticed that the lump-sum specific grant of AB pounds shown in Figure 3.2 would, in fact, have precisely the same effect as a lump-sum general grant of the same amount. Such a general grant would give the grantee the budget line BCB' rather than ACB', but equilibrium would still be at K. However, lump-sum specific grants and lump-sum general grants do not always have the same effect, as is shown in Figure 3.3. Here the initial equilibrium is again at J on the budget line AA' and indifference curve I. A lump-sum general grant of AB pounds, sufficient to purchase AC units of X, would result in equilibrium at K on the new budget line BCB' and indifference curve III. A lump-sum specific grant of AB pounds, on the other hand, would result in the budget line ACB', which would not allow the grantee to reach indifference curve III. The grantee would opt for point C on indifference curve II; it would therefore purchase AC units of X and receive AB pounds of grant, which is sufficient to finance its entire purchases of X. The grantee would thus acquire AB pounds with either grant, but it would prefer the general grant (which would enable it to reach indifference curve III) to the specific grant (which would

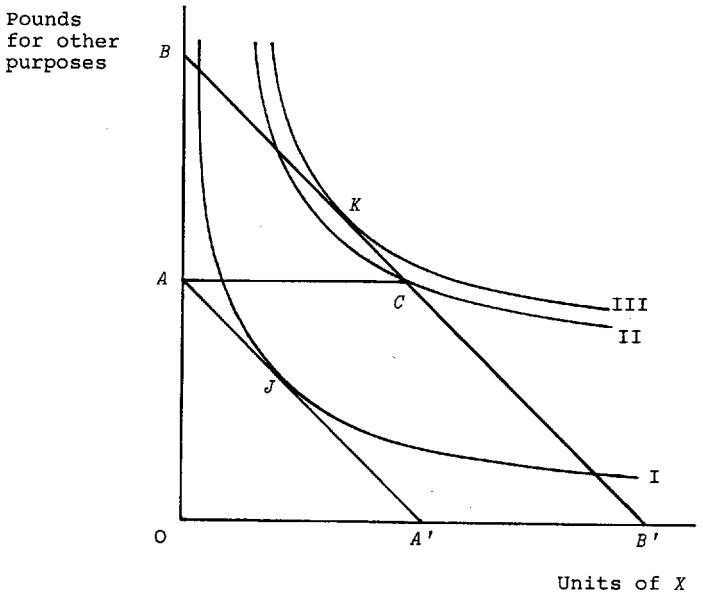

Figure 3.3

enable it to reach only indifference curve II). The grantor, however, whose aim is to stimulate purchases of X, would prefer the specific grant, since more units of X would be purchased at C than would be purchased at K. Wilde (1968, p. 341) has used the term 'deflective effect' to describe situations where specific lump-sum grants result in more expenditure on a particular item than general lump-sum grants. It is interesting to note that, in a study of ten central-city governments in the United States (between 1962 and 1970), Follain (1979, p. 498) concluded that specific grants did not, as a rule, stimulate expenditure more than general ones – that is, that the deflective effect was seldom operating.

Overall, then, these two types of grant may (as in Figure 3.2) have the same results, but where the deflective effect occurs (as in Figure 3.3) grantors will prefer a lump-sum specific grant to an equal value lump-sum general grant, while the converse applies to grantees. There seems no point in having a scheme of lump-sum specific grants as opposed to lump-sum general ones unless the deflective effect is expected to operate for at least one grantee authority (and for which the grantor will end up financing X in full).

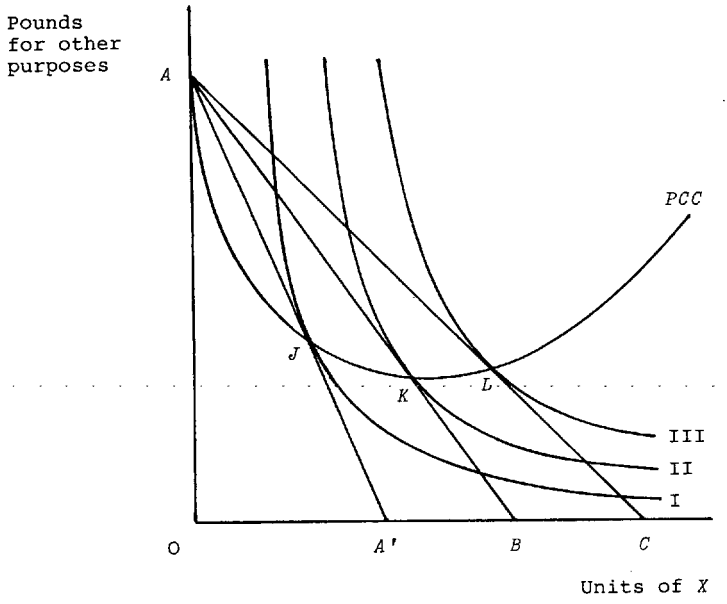

Figure 3.4

Matching grants

Consider, next, the possibility of stimulating expenditure on a particular function by introducing an open matching specific grant. This is illustrated in Figure 3.4, where the initial equilibrium is at J on budget line AA' and indifference curve I. If a matching grant were introduced that met the fraction $A'B/OB$ of the cost of providing X, then the budget line would become AB and equilibrium would move to K. If the grant were increased to meet the fraction $A'C/OC$, then the budget line would become AC and equilibrium would move to L. Points A, J, K, and L trace out the price consumption curve $APCC$. Provided X was not a Giffen good, each increase in grant would raise purchases of X. It may be noted that successive grant increases would reduce pounds spent on other items, and hence increase the grantee's own expenditure on X, until $APCC$ reached a minimum at K, so that demand for X is price elastic along AJK; thereafter, grant increases would raise spending on other items, thus decreasing the grantee's own expenditure on X and reflecting inelastic demand for X.

Open matching specific grants are compared with equal value

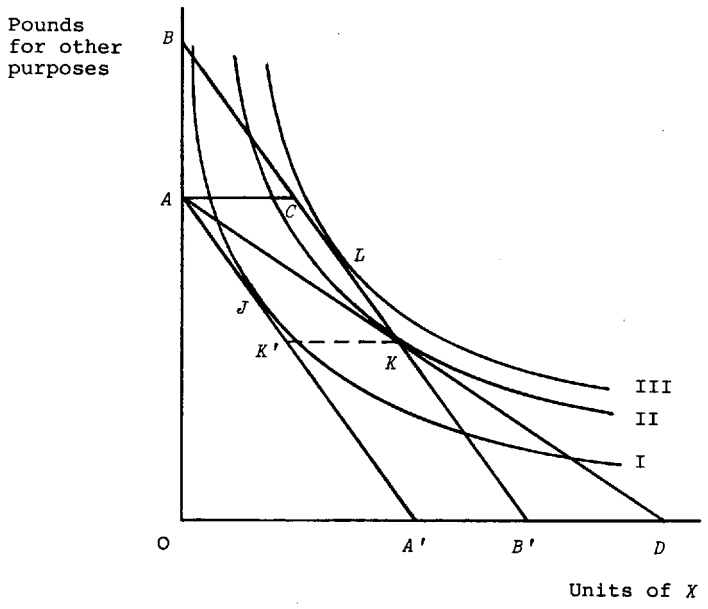

Figure 3.5

lump-sum grants in Figure 3.5. Here, an open matching specific grant that meets $A'D/OD$ of the cost of providing X has taken the grantee from J on indifference curve I to K on indifference curve II. The grantee uses its own income to finance the bundle of goods and services denoted by K', and finances $K'K$, which equals $A'B'$, units of X from its grant receipts. Thus an equal value lump-sum general grant would give the grantee the budget line BB'. This would be preferred by the grantee, as it would enable it to reach L on indifference curve III; the grantor, on the other hand, would prefer the matching grant, since more units of X would be purchased at K than at L. A lump-sum specific grant of equal value would result in the budget line ACB', and would also result in equilibrium at L. It, too, would be preferred to the matching grant by the grantee, but not by the grantor; it can readily be shown that these preference orderings would still apply even if the deflective effect was in operation.

The case of a closed matching specific grant must be considered next. There are three possible outcomes, two of which are illustrated in Figure 3.6, where the initial equilibrium is at K on AA'; indifference curves are not shown in order to simplify the diagram. It is supposed that the grant meets the fraction $A'F/OF$ of the cost of X, subject to some upper limit.

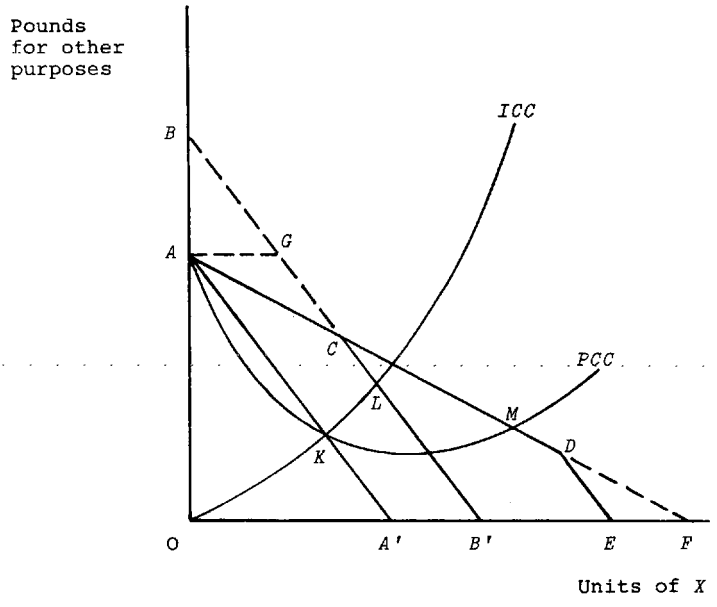

Figure 3.6

If, in fact, there were no upper limit, then the grantee would face the budget line AF, and the grantee would, in principle, be able to secure a grant adequate to meet the full cost of $A'F$ units of X if all the grantee's own resources were also devoted to purchasing X. Now, if the upper limit were equivalent to the full cost of $A'B'$ units of X, then the budget line would have a kink and would be ACB'. As a result, even if all the grantee's own resources were devoted to purchasing OA' units of X, it would receive enough grant to purchase only $A'B'$ units extra. On the other hand, if the upper limit were equivalent to the full cost of $A'E$ units of X, then the budget line would again have a kink, but would be ADE.

The effects of a closed matching specific grant depend on where the kink lies in relation to the income consumption curve $OICC$ and the price consumption curve $APCC$. If the kink were above both curves, as at C, then the equilibrium would be where the budget line cuts $OICC$, in this case at L. This is because the relevant part of the budget line, CB', is parallel to AA', and so it will touch an indifference curve where it meets $OICC$. At L, the maximum amount of grant would be secured; this is AB

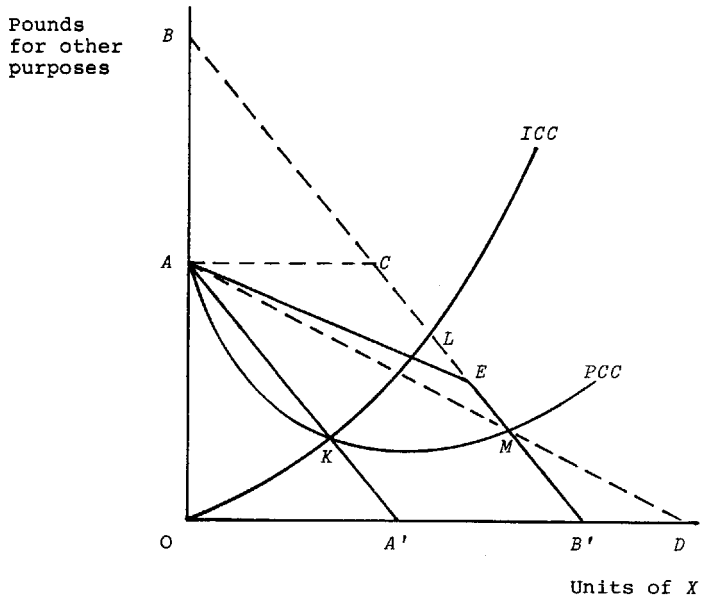

Figure 3.7

pounds, which would be sufficient to buy $A'B'$ units of X. Moreover, the grant would be exactly equivalent to an equal value lump-sum general grant, which would produce the budget line $BGCB'$, or an equal value lump-sum specific grant, which would produce the budget line $AGCB'$ (and with which, incidentally, the deflective effect would not be in operation); in each case, equilibrium would be at L. Both grantor and grantee would be indifferent between all three grants.

Conversely, suppose the kink were below both O*ICC* and *APCC*, as at D. The equilibrium would then be where the budget line cuts *APCC*, in this case at M. This is because the relevant part of the budget line, AD, starts from the same point on the vertical axis as AA', and so it will touch an indifference curve where it meets *APCC*. At M, the maximum amount of grant would not be secured because the equilibrium position is above the kink on the budget line. The grant this time would be exactly equivalent to an equal value open matching specific grant, which would produce the budget line AF, for this would also result in equilibrium at M. Thus grantor and grantee would be indifferent between these two sorts of grant.

The third possibility arises when the budget line kink is between O*ICC* and *APCC*, as in Figure 3.7 where the kink is at E. In this case, it can

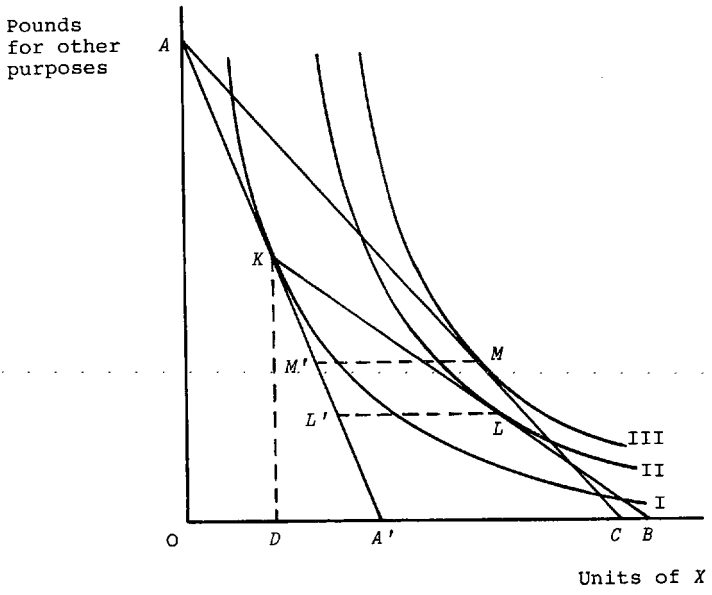

Figure 3.8

be shown that the equilibrium point must be at E itself. There is no tangency solution along AE, since all (unkinked) budget lines starting at A have tangencies on $APCC$, which does not cut AE; and there is no tangency solution along EB', since all budget lines parallel to AA' have tangencies on $OICC$, which does not cut EB'. Accordingly, the equilibrium must reflect a corner solution. By assumption, no solutions lie along the axes, and so the equilibrium must be at E. It follows that the grant paid would be the maximum available and equivalent to the full cost of $A'B'$ units of X, which is AB pounds. The grantee would prefer an equal value lump-sum general grant or an equal value lump-sum specific grant, giving budget lines respectively of BCB' and ACB', for each would enable it to reach L, which was unobtainable before; the indifference curve touching CB' at L must pass outside E, and this means L reflects a superior position to E for the grantee. The grantor would prefer an equal value open matching specific grant or an equal value closed matching specific grant, giving budget lines of AMD and AMB' respectively, for each of these would produce an equilibrium at M, where more X is purchased than at E.

This analysis implies that there is little point in introducing a closed matching specific grant for any one authority, or in introducing them

generally when the conditions of matching rates or maximum amounts can be varied between grantees. Whenever the equilibrium is expected to be above *OICC* or below *APCC*, then identical effects can be obtained by other grants at equal cost. Where the equilibrium is expected to be between *OICC* and *APCC*, then a number of equal cost alternatives are available, some preferable to the grantor and some to the grantee. Nevertheless, there clearly is a case for closed grants when it is felt that the grant conditions should be similar in each area, for closed grants with a high matching rate might encourage many grantees to spend more on *X* than they would with lump-sum grants, and yet be economical in the cases of those areas that responded the most.

All this analysis suggests that grantees generally prefer lump-sum specific grants to other specific grants (and may like lump-sum general grants even more), while grantors will find open matching specific grants most reliable in increasing purchases of *X*. It is sometimes suggested that open matching specific grants would have even more effect if they were made conditional upon a certain level of expenditure on the item concerned.[9] Figure 3.8 illustrates a case where the grantee receives a fixed proportion of the amount by which its expenditure on *X* exceeds the expenditure it would devote to *X* in the absence of grants, as reflected by the pre-grant equilibrium at *K* on *AA'*. The budget line thereby becomes *AKB*, where the fraction *A'B/DB* of extra expenditure is met by the grant. The equilibrium shifts to *L* on indifference curve II, and grant receipts are represented by *L'L*. An equal value open matching specific grant would result in a budget line *AC*, giving equilibrium at *M* on indifference curve III and (equal) grant receipts of *M'M*. This latter grant would be preferred by the grantee as it ends up on a higher indifference curve, but the former would be preferred by the grantor since it results in higher purchases of *X* for the same grant value.

One problem with this type of grant is determining the level of spending on *X* beyond which grants apply. This need not necessarily be the level that it would be without grants; indeed, that level might not be known when the grant has operated for some time. If the level were set too high, then the grant might have no effect. This is shown in Figure 3.9, where *AA'* shows the pre-grant budget line on which the authority selects the point *K* on indifference curve II. In this case, the grant changes the budget line from *AA'* to *ABC*, but the grantee prefers *K* to the best position available on *BC*, namely *L*, for that is on the lower indifference curve I. Accordingly, the grantee will continue to opt for *K* and so will receive no grant.

The effects of grantor taxes

It remains to consider the effects of the taxes that grantee citizens will be liable to pay to the grantor towards the cost of its grant programme aimed at stimulating the provision of a particular function. Such taxes will move the post-grant budget line towards the origin. Assuming that the grantor raises in taxes all the revenue needed for grants, and cannot

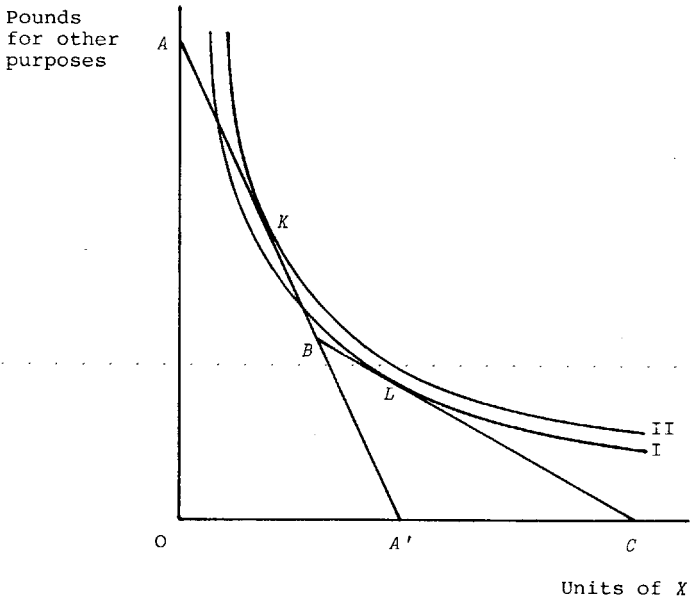

Figure 3.9

export its taxes on to citizens of other countries, then the citizens of many authorities are likely to find themselves paying roughly as much in grantor taxes as their authorities receive in grants. For simplicity, assume that tax payments precisely match grant receipts in each area.

In this case, equilibrium with a lump-sum general grant would be in precisely the pre-grant position. The same would apply to lump-sum specific grants, unless the deflective effect were operating, when there would still be some shift into X from other items; the inwards shift of the budget line caused by the grantor's taxes makes it more likely that this effect would be operating. An open matching specific grant would also continue to encourage purchases of X, for its relative price would be lower and there would be a pure substitution effect since income would be the same with and without the grant. A closed matching specific grant might have no effect, or a pure substitution effect, or an intermediate effect, depending on which situation applied. If conditional open matching specific grants (considered in Figures 3.8 and 3.9) have an effect at all, then they should continue to have a substitution effect in favour of X.

General grants

Little need be said about general grants designed to increase aggregate spending on grantor goods and services. Two of the most commonly discussed types are lump-sum general grants and general grants whose values are proportional to grantees' own tax-financed spending on such goods and services. These can be compared with the aid of Figure 3.10, where the horizontal axis shows spending on all grantee functions, and the vertical axis shows pounds left to grantee citizens for spending on private goods and services or saving. The initial equilibrium is at J on AA'. A lump-sum grant worth AB pounds actually produces the budget line ACB', with CB' parallel to AA', and equilibrium at L; the budget line would be BCB' only if some grant receipts could be used for transfer payments to enable the money available for private purposes to exceed the initial maximum of OA. An equal value proportional grant, meeting the proportion $A'D/OD$ of spending on grantee functions, results in equilibrium at K. The grantor would prefer the latter because it has more effect on spending on grantee functions. The grantee would prefer the

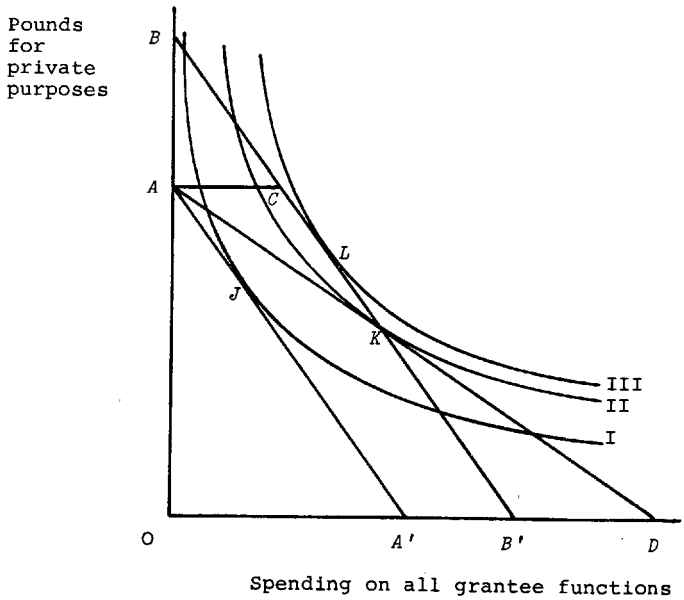

Figure 3.10

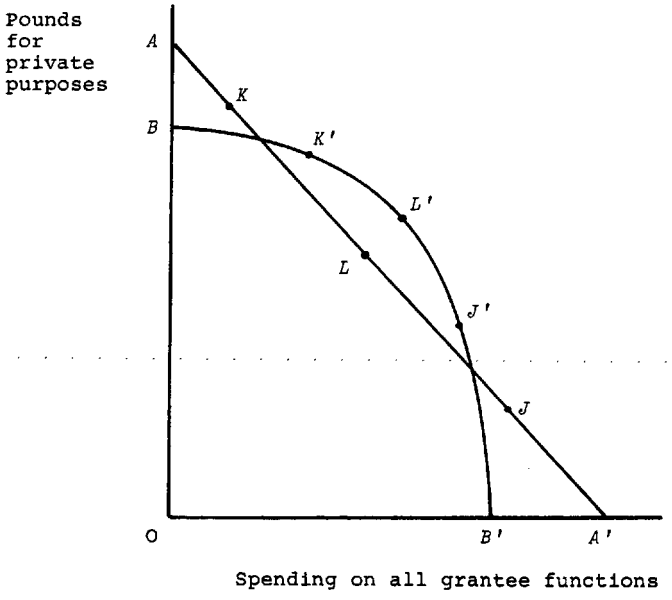

Figure 3.11

former because L is on a higher indifference curve, III, than K, which is on curve II. Allowing that grantee citizens have to pay taxes to the grantor to help finance the grants, and assuming that in the area concerned in Figure 3.10 these tax payments precisely equal grant receipts, then the lump-sum general grant would have no effect while the proportional grant would induce a substitution effect from private spending to grantee functions. Note that introducing a lump-sum grant and equal value tax payments would raise spending on grantee functions if the grant's value exceeded the initial spending on those functions (at J) and could be spent only on those functions and not on transfer payments.

Finally, mention may be made of a general grant specifically designed to discourage both high and low spending on grantee services.[10] This could result in negative grants when spending was low or high, and positive ones in between. The effect is shown in Figure 3.11, where the pre-grant budget line, AA', is replaced by the curve BB'. High public spenders might move from J to J' and low ones from K to K'. In each case, the new position could be on a lower or on a higher indifference curve. Medium spenders might move from L to L' on a higher curve, and some of them might find that their new equilibriums were in precisely the same places

that they would be with equal value lump-sum or proportional grants. BB' shifts downwards when tax payments to the grantor are allowed for; this might accentuate the tendency for grantee authorities to move towards medium amounts of public spending.

The background to the new models

The traditional model described at length in the previous sections predicted the effects of grants in the context of community indifference curves. As noted earlier, similar models can be built on the basis of the median voter's preferences, and these yield very similar results. It should be stressed that the predictive power of these models depends crucially on the assumption that the politicians and bureaucrats attempt to fix the pattern of grantee services in a way that maximises community welfare or the satisfaction of the median voter.

Now, one particular feature of these models is that the effect of a general lump-sum grant can be predicted by considering a parallel shift in the budget line (as from AA' to ACB' in Figure 3.10). This results in a rise in spending on grantee functions if these are normal goods, and also a rise in the money devoted to private purposes, indicating that the grant leads to some reduction in grantee taxes. A parallel movement of the budget line would also occur if the incomes of grantee citizens rose, as they would, for instance, if the grantor authority reduced its taxes. The effects of a lump-sum grant and an equal value cut in grantor taxes might not be identical because the two policies might have slightly different distributional impacts.[11] Nevertheless, it would be expected that the two policies would usually have comparable effects on grantee spending. However, in a review of the empirical evidence on grants in the United States, Gramlich (1977) concluded that lump-sum grants from the central government stimulated much more grantee spending in the long run than did cuts in central government taxes. The studies showed (pp. 227–30) that a \$100 increase in its citizens' net incomes might raise a subcentral authority's spending by \$5–\$10, while a \$100 lump-sum grant might raise it by \$40–\$100. This result has become known as the 'flypaper' effect – a term Gramlich (p. 236) says was coined by Okun – because it appears that the grantor's money tends to stick in whichever sector (public or private) it arrives.

It is often concluded from this evidence that, on receipt of a grant, politicians and bureaucrats fail to reduce grantee taxes as much as they would if they genuinely tried to maximise community welfare or the welfare of the median voter. Gramlich himself concluded (p. 234) that there was a disharmony between the interests of bureaucrats and voters. The rest of this chapter is concerned with five models in which the authority officials operate in ways that result in lump-sum grants having a bigger effect on budgets than similar increases in incomes, although, in the last two at least, this is not because of any disharmony in interests.

The Niskanen model

The discrepancy between what voters would like and what they get when grants are increased is blamed squarely on bureaucrats in the model of bureaucratic behaviour put forward by Niskanen (1968) and developed by Breton and Wintrobe (1975). Niskanen argued that bureaucrats maximise their own welfare, rather than that of voters. Each one's welfare depends on many variables including salary, perquisites of the office, public reputation, power, patronage, output of the bureau, ease of making changes and ease of managing the bureau, all of which, with the possible exception of the last two,[12] are held to be positive functions of the total budget of the bureau during the bureaucrat's tenure. Thus the bureaucrat's maximand is essentially the size of the bureau's budget.

The bureau is alleged to be a monopoly that sells its services to the governing politicians. The model is illustrated in Figures 3.12 and 3.13, which assume that the bureau produces a single service (X) whose quantity (per period of time) is measured along the horizontal axis. The governing politicians are assumed to know the preferences of their citizens, and the sum of the citizens' marginal valuations of the bureau's services is shown in the curve labelled MV; associated with this is the sum of the

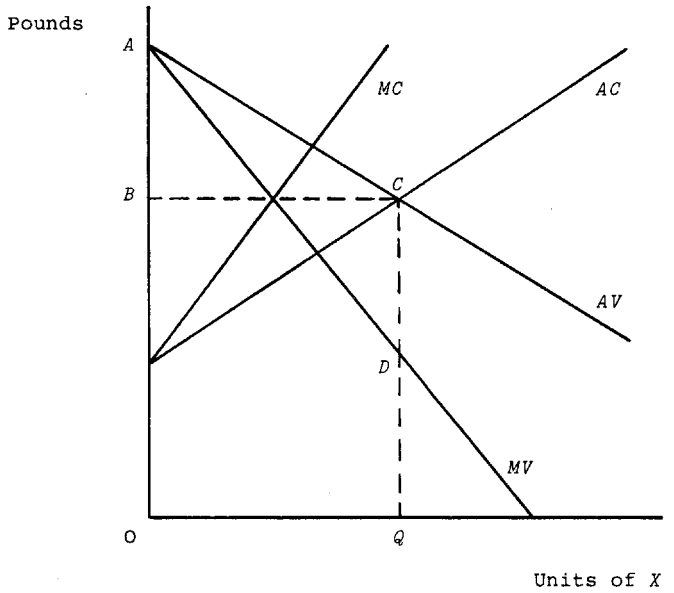

Figure 3.12

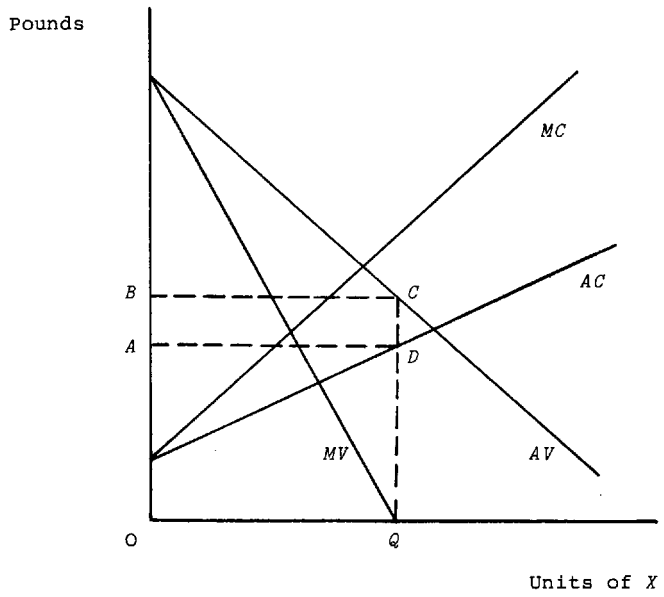

Figure 3.13

citizens' average valuations shown in the curve AV. The minimum average cost at which each possible output level can be produced is shown in the curve labelled AC, and the curve MC shows the related marginal costs. The bureau exploits its monopoly position by offering its service on an all-or-nothing basis. This means it can raise its budget up to the point where the total benefits just equal (or marginally exceed) the total costs to the politicians, and hence citizens, since at that level the politicians and citizens are indifferent between all and nothing (or marginally prefer all).

In the situation depicted by Figure 3.12, Niskanen hypothesises that this budget will be the one associated with producing output OQ at the average cost OB; at this output AC equals AV so that total benefits and costs are equal. The bureau will not set a lower output, for this would generate a smaller total benefit and therefore require it to raise less revenue and so have a smaller budget; it will not set a higher output, because this could not be done without total costs exceeding total benefits and so exceeding the revenue that could be raised. It will be noticed that the bureau produces OQ at the minimum possible average cost OB, and hence is X-efficient. It will also be seen that the total costs and benefits equal $OBCQ$; the total benefits can be derived instead from the area under the MV curve as $OADQ$. Finally, it should be stressed that output at OQ exceeds the socially optimum level, which is

where MC equals MV[13] (if the prices of other items equal their respective marginal costs).

Figure 3.13 illustrates a special case. Here, output is expanded up to OQ where the marginal valuation equals zero. At this output, AV exceeds the minimum possible average cost shown by AC, but there is no point expanding output until AV equals AC, for total benefits will not increase, and thus the total revenue that can be extracted from politicians will not rise. The bureau will produce the output OQ that generates total benefits valued at $OBCQ$ and sell it for the same sum, which thus defines the budget level. This output could be produced for a total cost of $OADQ$, which is less than the budget. The bureau will therefore have some fat, which is absorbed by X-inefficiency, so that its total costs rise to equal its budget. Once again, output exceeds the socially optimum level where MC equals MV.

Suppose, now, that the bureau produces its service, X, for a subcentral authority, and that an open matching specific grant is introduced for this service to meet one-third of its cost. This reduces the average costs that the bureau must recoup from the authority's politicians and citizens by one-third, as shown in Figures 3.14 and 3.15, where the AC, AV and MV curves are as in Figures 3.12 and 3.13 respectively, and marginal costs are omitted for simplicity. The grant produces the new average cost curve,

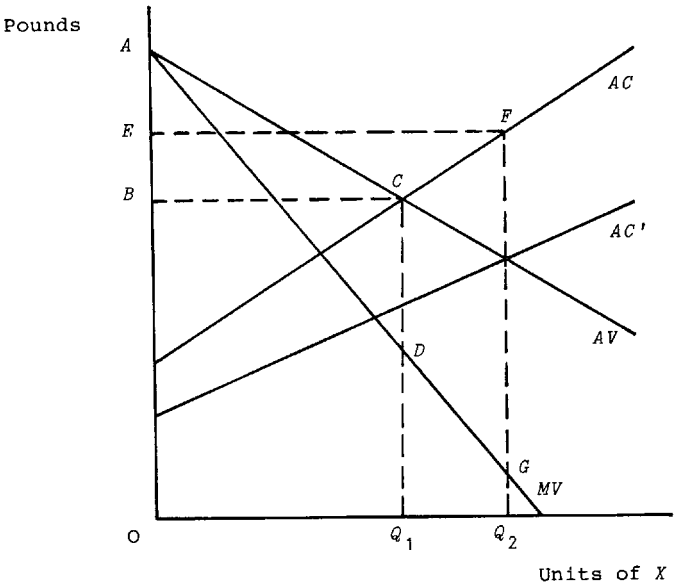

Figure 3.14

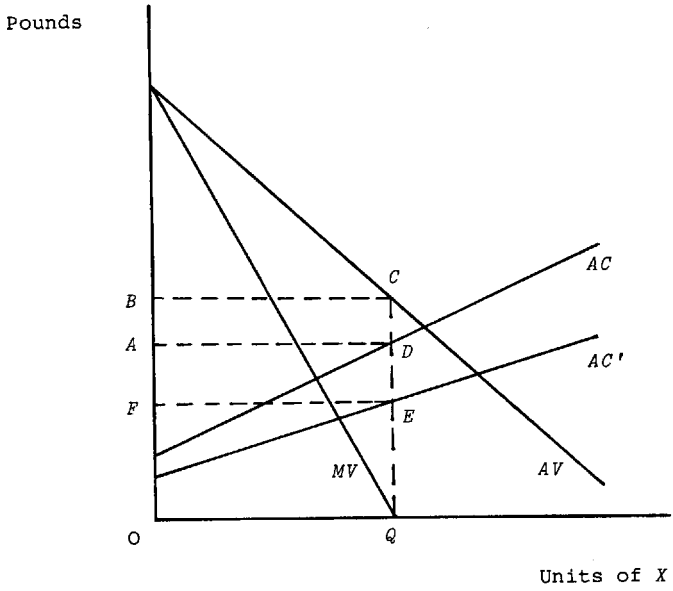

Figure 3.15

AC', whose height above the horizontal axis is two-thirds that of AC. In the case shown in Figure 3.14, output rises from OQ_1 to OQ_2 and the budget rises from $OBCQ_1$ to $OEFQ_2$. The higher output increases total benefits from $OADQ_1$ to $OAGQ_2$, that is by Q_1DGQ_2, and so the bureau is able to extract this amount more money from politicians and citizens.[14] Thus its tax revenue rises, and hence its budget rises by more than the amount of the grant (to equal $OAGQ_2$ plus the grant).

In the case shown in Figure 3.15, by contrast, output does not rise at all, because an output rise would not create any extra benefit and would not enable any extra tax revenue to be raised. The bureau continues to raise $OBCQ$ from the politicians, but adds the value of the grant to its budget and absorbs this in further X-inefficiency; the grant's value is $ADEF$, so the budget equals this amount plus $OBCQ$. In this case, then, the budget rises by the full amount of the grant alone, and output is constant.

It should be noted that the effects on the bureau's output and budget would be identical for any grant of equal total value. In the case of Figure 3.14, any grant of equal total value would cut average costs at the output level OQ_2 by the same amount, so the new average cost curve would intersect AV at that output, ensuring that it was still the equilibrium level.[15] The budget would equal $OAGQ_2$ plus the grant as before. In the

case of Figure 3.15, no grant will result in a higher output, and an equal value grant will raise the budget by that value.

Niskanen's theory shows how a grant could lead to a grantee's budget rising by the value of the grant or by even more. For his theory to explain the flypaper effect, it is sufficient for it to show that the effect of a rise in citizens' incomes could be to raise the budget by less than that rise. This is easily done. A rise in incomes will affect the situation only through its effect on citizens' valuations and hence their MV and AV curves. If these remained unaltered, then there would be no effect on the output and budget levels shown in Figures 3.12 and 3.13. Indeed, if the bureau's service were inferior, then these schedules could shift to the left to produce lower outputs and smaller budgets. Conversely, if it were normal, then the schedules would shift to the right, and outputs and budgets would rise, but the increase in the budget need not equal the value of the grant, let alone exceed it.

This theory has been subjected to criticism (for instance by Breton and Wintrobe, 1975). It has been claimed that bureaucrats are not necessarily budget maximisers, and suggested that politicians may not be as passive as the model suggests. Furthermore, it has been argued by Peacock (1979, pp. 235–42) that in the United Kingdom (at least at the central government level) it is politicians who exert the power ascribed by Niskanen to bureaucrats, and that bureaucrats may be more interested in maximising on-the-job-leisure than budgets; Peacock concedes, though, that politicians in the UK have some interest in securing high budgets. Niskanen's formulation is perhaps extreme, but to the extent that his model has some validity it may help explain the flypaper effect, which is not, in any case, as marked as his model might suggest.

The Romer and Rosenthal model

A related theory to explain the flypaper effect has been developed by Romer and Rosenthal (1980). This assumes that the subcentral authority is budget maximising, though no particular distinction is made between bureaucrats and politicians. In this model, the authority is obliged to set its annual expenditure at some exogenously determined reversion level unless there is a referendum in which citizens choose between that reversion level and a proposed new level and choose the latter. Proposals for new levels are determined by those operating the authority (known collectively as the setter), who are assumed to know voters' preferences. The reversion level could be, for instance, the previous year's level; Romer and Rosenthal have suggested (1980, p. 454) that the reversion level could be zero, but their exposition and the present discussion assume a positive level.

Figure 3.16 illustrates the situation (for a period of time) for an authority producing a single service (X) whose output is measured on the horizontal axis; the vertical axis shows pounds left for private consumption or saving. The indifference map is that of the median voter, who is assumed to be the pivotal and decisive voter. The budget line, AA', shows the combinations of units of X and pounds available for spending on other

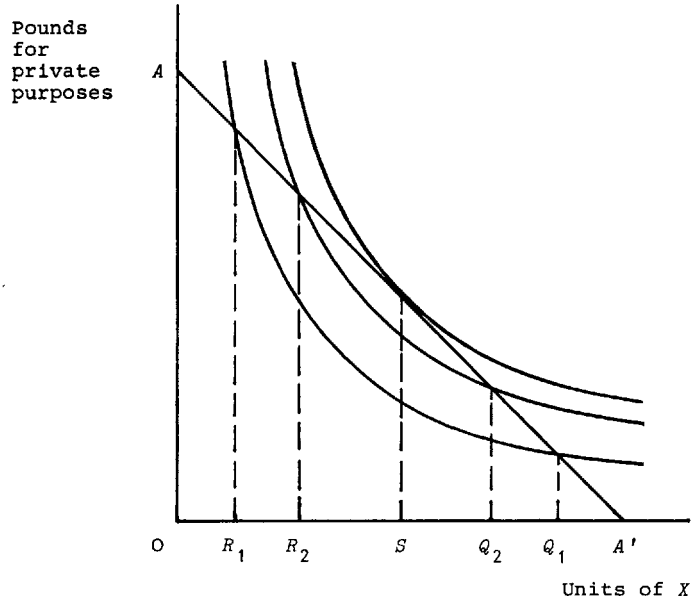

Figure 3.16

goods and services or saving that are available to the median voter when a given single tax is levied by the authority.

If the reversion budget level were sufficient to finance an output of OR_1, then the setter would propose the level OQ_1 (or, rather, one just below it) for the voter is indifferent between OQ_1 and OR_1, and a level just below OQ_1 will be ratified in the referendum; this indicates the highest output and budget that the setter can obtain. If the reversion level were OR_2, then the setter would propose a level just below OQ_2, and this would be passed in the referendum. If the reversion were OS or above, then the setter would be best advised to spend at the reversion level and not bother with a referendum, for any proposal for higher spending would be rejected.

Suppose, for simplicity, that the budget line and the median voter's preferences are stable for a period of years. Suppose, also, that the reversion level is defined as the previous year's spending, and that in one year it is below OS. In this case, the setter will have a referendum and obtain a level of spending above OS. It follows that in future years the reversion level will exceed OS and referenda will not be held. It also follows that the reversion level will generally exceed the sum needed to finance OS (that is, the median voter's preferred level) and that the spending level will equal the reversion level.

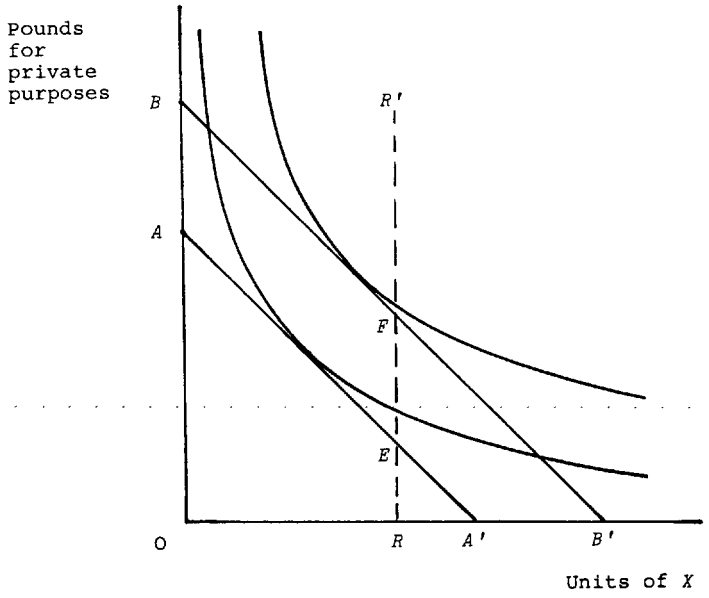

Figure 3.17

Figures 3.17 and 3.18 show respectively the results of a rise in incomes and of an equal value lump-sum grant in a case where the median voter's new preferred level is below the original reversion level. In each figure, the median voter's initial budget line is shown by AA' and the reversion level, reflected by OR, exceeds the median voter's initial optimal level. The initial output of X will equal OR so that the median voter ends up at E on his initial budget line. In Figure 3.17, a rise in incomes shifts the budget line to BB', but, provided the rise is insufficient to take the new optimum to the right of RR' (which indicates the reversion level), the output will still equal OR. Thus the income change will have no effect on either the output or the budget, and the median voter will end up at F. The same result would occur with the grant, except that Romer and Rosenthal assume that the grant must be spent along with the old reversion expenditure. In this case, the effective reversion level rises by the amount of grant; in terms of units of X, this would now be OS, and is indicated by SS' in Figure 3.18. In subsequent years, the reversion level must be presumed to remain at OS unless the grant alters. This must be above the median voter's new optimum in a case where that is to the left of RR', so no referendum will be held, but output of X will rise from OR to OS. This means that spending on X rises by the full amount of the grant, and the median voter will end up at G.

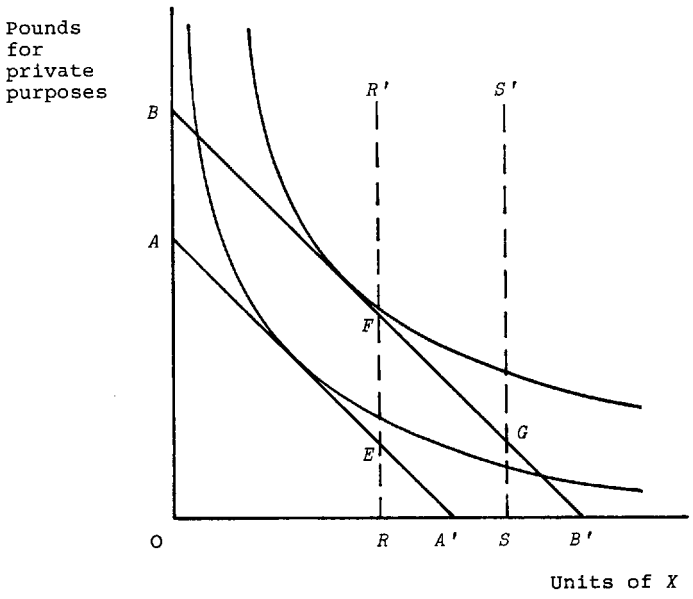

Figure 3.18

It follows that the particular situations shown in Figures 3.17 and 3.18, where the new optimal level is below the original reversion level, offer another explanation of the flypaper effect. Romer and Rosenthal (1980, pp. 456–7) have shown that, if the reversion level is below the median voter's new optimum, then the flypaper effect may not occur. It should be noted that any grant, such as an open matching specific grant, whose value was equal to the lump-sum one just considered would raise the level of spending on X by the same amount as that grant, provided the authority's new reversion level had to equal its old one plus the amount of the grant.

The Oates model

Some attempts have been made to explain the flypaper effect in terms of fiscal illusion by voters. In one such model, Oates (1979a,b) imagined that the subcentral authority officials set the budget in accordance with the median voter's preferences, but give this voter incomplete information when a grant is received, and therefore deceive him into accepting an excessively high level of output.[16] The model is illustrated (for a period of time) in Figures 3.19 and 3.20. It is assumed that the authority provides

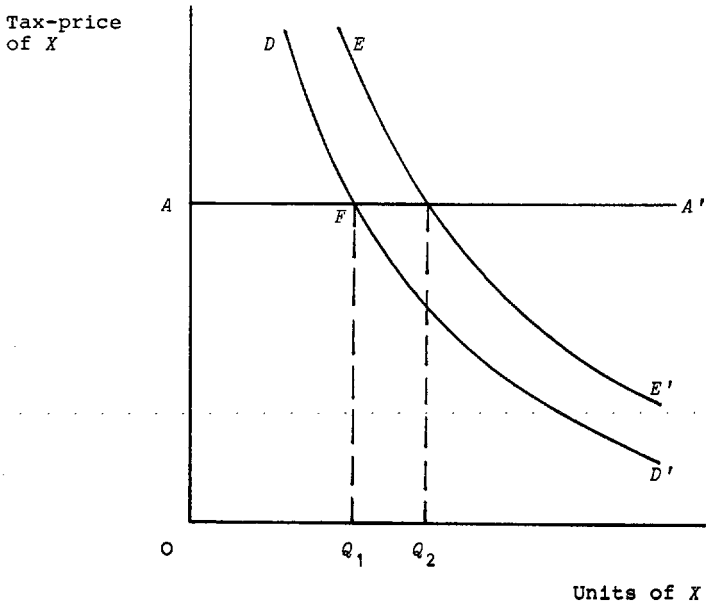

Figure 3.19

one service (X) whose average and marginal costs are constant, and it is assumed that the median voter pays a fixed proportion of these costs. In a situation without grants, therefore, the tax-price to the median voter – that is, the amount he must pay in taxes for each unit of X – will be constant, and this is shown in each figure by the line AA'. His demand curve is shown by DD'. The authority provides the quantity he would select, OQ_1, and collects a tax payment of $OAFQ_1$ from him.

In contrasting the effect of an increase in income with the introduction of a lump-sum grant, it is assumed that the median voter remains the same person, and that an increase in his income is accompanied by an increase in the incomes of other citizens. If incomes in general, and the median voter's in particular, were to rise, then his demand curve would shift to the right, say to EE' in Figure 3.19, provided that X was a normal good. The quantity demanded and thus provided would rise from OQ_1 to OQ_2 and the rise in the budget would be proportional to this, though its absolute size would also depend on the initial level of spending on X. The rise in the budget would therefore depend on the income elasticity of demand. For the budget to rise as much as the rise in income, though, the voter would have to wish to spend all the increase in his income on X, suggesting a zero income elasticity of demand for all other items; this is clearly an unlikely situation.

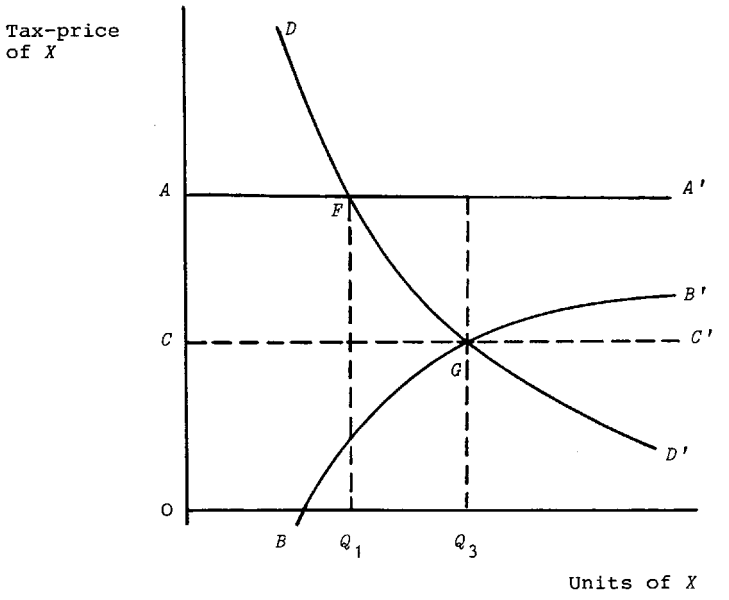

Figure 3.20

On receipt of an equal value lump-sum grant, the authority could make transfer payments to its citizens, and it could do so in such a way that the median voter ended up once again with the demand curve EE' and desired an output OQ_2 of X. In fact, though, the authority is regarded as offering to produce X at a lower tax-price. The new tax-price for each possible output level is shown in Figure 3.20 by the curve BB'. The median voter therefore selects the quantity OQ_3 and pays taxes of $OCGQ_3$. The effect on the quantity and budget depend on his price elasticity of demand and the initial level of spending on X. If price elasticity of demand were to exceed zero, then the output, and hence the budget, would rise; if it were to exceed 1, then his tax payments would also rise, with the result that the budget would rise by more than the grant. It should be noticed that if the grant were presented in this way, then the voter would be deceived into believing that the price of X had fallen in relation to the prices of other items, and would therefore act on the false assumption that the grant was equivalent to a fall in the price of X rather than equivalent to a rise in income. Thus the voter responds to the grant in a state of fiscal illusion.

Since the effects on the output of X depend respectively on the income and price elasticity of demand and the initial level of spending on X, it

cannot be said *a priori* whether the grant would have a greater or smaller impact. However, Oates has shown (1979a, p. 13; 1979b, p. 28) that, on plausible assumptions about elasticities, the effect of a \$100 rise in incomes would be to raise the budget by \$10, while a similar value lump-sum grant would raise it by \$40. Thus this model, too, offers an explanation of the flypaper effect.

Oates did not investigate the effects of an open matching specific grant, but this would presumably result in offers to produce X at a constant price below OA. To persuade the voter to opt for quantity OQ_3 with a matching grant, it would be necessary for the grantor to pay one that resulted in the grantee giving the median voter the offers indicated by the dashed line CC' in Figure 3.20. The median voter would pay taxes of $OCGQ_3$, as with the lump-sum grant, and the same total grant would be paid towards the cost of the same output level OQ_3. Thus, once again, lump-sum and matching grants of equal values appear to have similar effects.

The Break model

In another attempt to explain the flypaper effect, Break (1980, p. 91) has pointed out that local voters and taxpayers are also voters and taxpayers for any higher authority that distributes grants. They might desire higher local spending but fear the adverse effects of any rise in local taxes. In this situation, a rise in incomes might not lead to much pressure for higher local budgets, whereas an equal rise in grants might willingly be allowed to be devoted chiefly to local public spending. In this case, too, it seems possible that equal value grants of any sort would have similar effects.

A problem with the new models

The first three new models just outlined all assumed, for simplicity, that subcentral authorities provide a single service. Each model showed how the receipt of a lump-sum grant (which, in the context of a single-service authority, could be seen as a lump-sum general grant or as a lump-sum specific grant) could stimulate a greater increase in grantee spending than an equal value rise in incomes. The models thereby offered explanations for the flypaper effect, that is for lump-sum grants to have more stimulus on grantees than equal value central government tax cuts. The fourth model (Break's) also offered an explanation for this effect. Thus, all four models are consistent with one piece of evidence for the United States, noted by Gramlich (1977, p. 234) in his review, that a significant flypaper effect occurs. In this respect, the models are all superior to the traditional model, which suggests that a significant flypaper effect is unlikely.

The first three models also seem to suggest, however, that a lump-sum (specific) grant would have as much effect on grantee spending as an equal value matching (specific) grant, and it appears that Break's model would produce a similar result. Accordingly, all four models seem inconsistent

with another piece of evidence for the United States, noted in Gramlich's review (p. 234), for he found that lump-sum grants had less effect than equal value matching grants.[17] Ironically, the models are inferior to the traditional model in this respect, for that model makes predictions consistent with this conclusion of Gramlich.

There is clearly a case for constructing models that square with the evidence in both respects, so that they would show how a flypaper effect might occur and yet also show why matching grants might have a greater effect on grantee spending than equal value lump-sum grants. The rest of this section puts forward a new model with these two properties. This model also assumes, for simplicity, that grantees provide a single service (X).

The new model is, in essence, a version of the traditional model described earlier with two modifications. First, it is supposed that subcentral authorities act in accordance with the median voter's preferences. Thus it is supposed that it is his indifference curves that are relevant, that his identity is unaltered when incomes change or grants are received, and that he is the pivotal voter. This modification is often made, and it does not change the principal predictions of the traditional model.

The second and novel modification is that it is supposed that the median voter selects the amount of subcentral authority spending that enables him to reach his highest level of indifference between subcentral authority spending and private spending and saving subject to the constraint that the tax rate set must not leave the authority's poorest citizens with a net private income below some level specified by the median voter. The effects of such a constraint are shown (for a period of time) in Figure 3.21. The budget line AA' shows the initial possible combinations of units of X and pounds left over for citizens' private spending and saving among which the median voter can choose. The median voter constrains himself to choose combinations between A and F on the grounds that were grantee spending to equal the level represented by F, requiring a sacrifice of possible private spending equal to AD, then the grantee would have to set tax rates that would leave its poorest citizens with the minimum income level he regards as acceptable; higher grantee spending is vetoed because it would entail higher tax rates and unacceptably low net incomes for the poorest citizens.

Now, introducing an open matching specific grant might produce the budget line AB. In this case, the median voter's constraint would be at H, because that position would be obtained with the same tax rate that previously produced position F (since it represents an equal sacrifice AD of possible private spending). A lump-sum grant would also place the constraint on the same horizontal level as F. Such a grant might produce the budget line ACC'; the budget line is kinked at C on the assumption that the grantee would not be eligible to receive the full grant (worth AC in terms of units of X) unless it devoted at least that amount to public spending. The constraint is at G because this position would be secured if the same tax rate were set that originally produced position F (since it again represents an equal sacrifice AD of possible private spending).

A cut in central government taxes that produced a rise in grantee citizens'

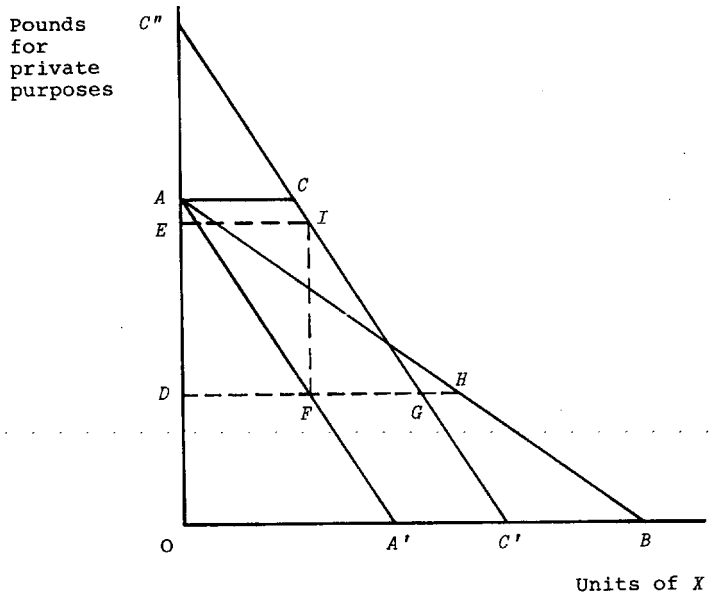

Figure 3.21

aggregate incomes equal to the lump-sum grant just considered (whose value could be measured as AC or $A'C'$ in terms of units of X, or as AC'' in pounds) would produce the budget line $C''CC'$. The position of the constraint in this case would depend on the nature of the central taxes that were cut and the type of subcentral taxes that were levied. It is sufficient for the present text to consider one likely possibility that would produce a flypaper effect. Suppose that the central government cut its income tax, and that this was of no benefit to the grantee's poorest citizens, whose incomes were previously too low for them to be required to pay this tax; suppose also that the grantee levies a property tax alone, and that the median voter believes that the incidence of any increase in this tax on domestic property in his area would fall on resident occupiers, even if they were tenants rather than owner-occupiers.[18] In such circumstances, the median voter would not want the subcentral authority's property tax rate to exceed the previous constrained level since any higher rate would be felt to leave the poorest citizens worse off than they would have been with that level. Thus the constraint would be at I, vertically above F, for here the tax yield would be the same as that which originally produced position F (since it represents a sacrifice of $C''E$ in possible private spending, which is equal to the original sacrifice AD). If the poorest citizens had received some benefit from the central tax cut, then the median

voter would allow some increase in grantee tax rates, and the constraint would be below I on $C''CC'$, but it could well be above G.

It should be observed that the constraint on $C''CC'$ could, in principle, be below G. For example, if the initial benefits from the reduced central tax payments accrued for some reason wholly to the grantee's poorest citizens, then the median voter might well permit a rise in the grantee's tax rate that took the entire increase away from those citizens to leave them with their initial net incomes. These citizens would therefore have as much money for private expenditure as before while others would have less, since they derive no benefit from the central tax cuts but would have to pay a higher rate of tax to the grantee. In this case, the median voter's constraint would reflect a lower minimum permitted level of aggregate private spending than in the initial situation. Hence his constraint on $C''CC'$ would be lower than F and so lower than G. In practice, it is unlikely that the cut in central taxes would benefit the poor alone, but their net incomes might rise by a significant proportion if, for example, income tax thresholds were raised (assuming on this occasion that the poor were initially paying some income tax), thereby permitting a significant rise in grantee tax rates, and making it possible that the constraint on $C''CC'$ would be below G.

The final outcome of each possible event depends, of course, on the

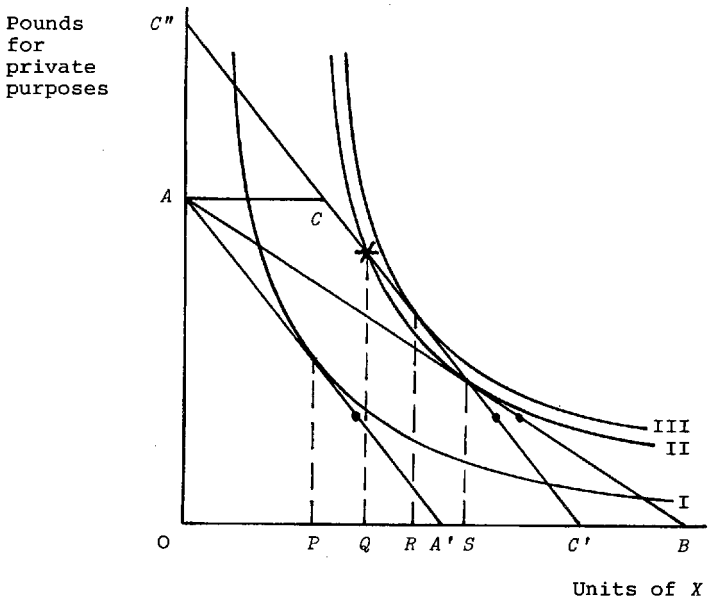

Figure 3.22

median voter's indifference map. Various different outcomes are possible. For example, the median voter's preferred positions might always be such that the constraint never operated, in which case the predictions of this model would be identical to those of the traditional model, and would not suggest a flypaper effect. However, it is possible for the indifference map to predict a flypaper effect and (unlike the other new models) to predict also that effort-related general grants would have more effect than lump-sum ones.

This possibility is shown in Figure 3.22, where AA' represents the original budget line with the constraint marked on it (with a dot). The median voter's preferred position is on indifference curve I and is, in fact, above the constraint on his budget line, so this position is selected, and the grantee provides OP units of X. A matching (specific) grant might produce the budget line AB, which has the appropriate constraint marked on it (with a dot). The median voter's preferred position would be on indifference curve II, and, as shown, is again above the constraint on his budget line, so this position would now be selected and OS units of X would be provided. To show the effect of an equal value lump-sum (general or specific) grant, the budget line ACC' is drawn (with CC' parallel to AA') passing through the preferred and chosen point on AB. The constraint on ACC' is marked in appropriately (with a dot) at the same horizontal level as the constraints on AA' and AB. The median voter's preferred position would be on indifference curve III and as, once again, this is above the constraint, this position would be selected and OR units of X would be provided. Note that this grant has less effect (by $OS - OR$) on the output of X than the matching grant, as observed in the United States.

A cut in central taxes that raised aggregate grantee citizens' incomes by the same amount as the lump-sum grant would produce the budget line $C''CC'$. The constraint is marked on this line (with a cross) and is drawn slightly to the right of the constraint on AA' to show that, in this example, the poorest citizens did receive some benefit from the tax cut. The highest indifference curve reached by $C''CC'$ is indifference curve III, but the tangency point is below the constraint (marked with a cross). The median voter therefore selects the best position attainable subject to the constraint being met. This will inevitably be at the constraint position. For simplicity, it is shown as being on the same indifference curve (II) that touched AB, though it could be on a slightly higher or lower one. The crucial point is that the stimulus to grantee spending is less (by $OR - OQ$) with this rise in incomes than it is with the equal value lump-sum grant. Thus the model shows how the flypaper effect could occur.

It should be noted that other outcomes are possible, depending on the circumstances. For example, Figure 3.23 shows a situation where, as before, a lump-sum grant produces the budget line ACC' with the constraint (marked with a dot) at the same horizontal level as the initial constraint on AA' (also marked with a dot). An equal value aggregate cut in grantee citizens' tax payments to the central government produces the budget line $C''CC'$ (as before) but, this time, the constraint (marked with a cross) is below that on ACC', suggesting that much of the initial

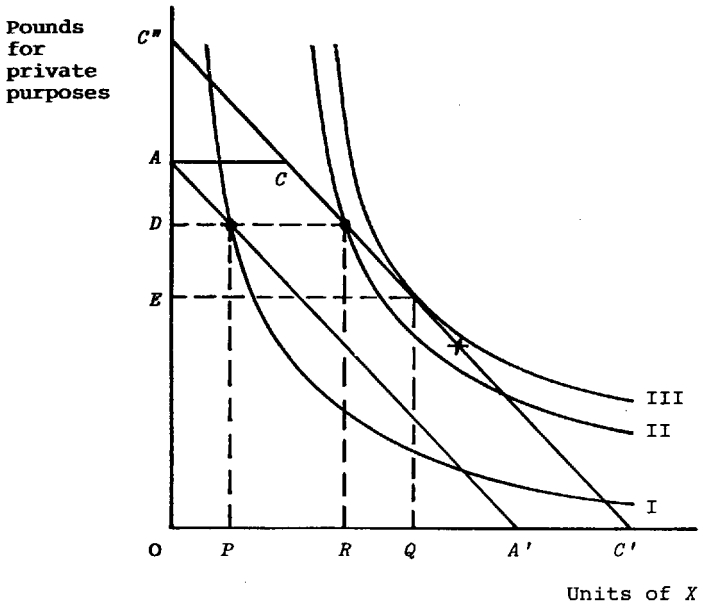

Figure 3.23

gain from the tax payment cuts accrued to the grantee's poorest citizens.

In this example, the initial position selected is on indifference curve I with OP units of X. The median voter could reach a higher level of satisfaction from more units of X, but opts for the highest quantity, OP, permitted by the constraint, which this time is in operation at the outset. The lump-sum grant places him on ACC', and he opts for the maximum quantity of X, OR, permitted by the new constraint, though he would prefer more. The tax cut places him on $C''CC'$, and enables him to select the preferred output level of OQ units, since this is above the relevant constraint (marked with a cross).

It will be noticed that, in this case, output of X would rise more (by $OQ - OR$) if taxes were cut than if lump-sum grants were raised. In other words, public (grantee) spending actually rises more if central government money is initially directed at the private sector rather than at the public (grantee) sector itself. Conversely, private spending turns out to be higher, at OD, when grants are raised than it is, at OE, when taxes are cut. It is thus possible that the final spending in either sector would be higher if money were initially directed towards the other sector. This possibility could, perhaps, be termed the 'fly-repellent' effect.

This model suggests that various effects are possible. It is, however, able to set up a situation (as in Figure 3.22) consistent with both the major

empirical findings for the United States; none of the other new models seems able to do this. Of course, that does not prove that this final model is correct; there could be other explanations that are also consistent with these two findings and that would prove more satisfactory in other ways. Nor does it prove that the four other new models are quite wrong. Taken individually or collectively, those four models do not seem to give a complete picture of grant effects, for they appear to suggest that lump-sum grants must have as much effect as matching ones; nevertheless the flypaper effect may well be aggravated by the factors that they suggest.

Summary

The principal division of grants is between specific grants, which may be spent by grantees only on specified services, and general grants, which may be spent at the grantees' discretion. Typically, the former will be used where the grantor wishes to expand spending on particular services, and the latter where it wishes to expand grantee spending overall. However, an important conclusion is that an individual lump-sum specific grant may well have no more effect on spending on the item concerned than an equal value lump-sum general grant.

In general, the stimulus to spending is greatest where grant receipts vary according to how much effort the grantees make, either in spending on a specified service or in overall spending from taxes, as the case may be. However, grantees would prefer lump-sum grants that do not depend on their own efforts. Any stimulus given by grants is likely to be at least partly offset by the taxes that the grantor must raise to finance its grants, for these reduce the disposable incomes of grantee citizens, who may respond by seeking lower subcentral tax and spending levels. These offsets will be of most concern with lump-sum grants, for they could in principle mean that systems of lump-sum grants have very little net effect on subcentral spending.

In practice, though, at least in the United States, the evidence is that subcentral spending is affected much more by changes in lump-sum grants than by equal value changes in the tax payments of grantee citizens to the grantor. This phenomenon has become known as the flypaper effect. Previous attempts to explain the flypaper effect have rested on factors such as self-interested behaviour by bureaucrats and politicians or fiscal illusion by voters. These factors may well be relevant, but models based on them seem to suggest that lump-sum grants will have as much effect as any other sort, which contradicts both the traditional theory and the available empirical evidence.

This chapter puts forward a new model to explain grant effects that is similar to the traditional model except that it assumes that grantee citizens may temper their demands for subcentral services if the necessary subcentral taxes impinge too heavily on their area's poorest citizens. This model shows not only how the flypaper effect might occur, but also why lump-sum grants could have less effect than those that depend on grantee spending levels.

Notes

1 It could be argued, somewhat pedantically, that the proposed block grants for the Scottish and Welsh assemblies were not wholly unconditional since they could be spent only on functions granted to those assemblies by the central government. However, the limits were much wider than those suggested by Grewal in his definition of block grants.
2 See, for example, HMSO (1975, pp. 21–2 and 44–5) for its use in the former case, and Jackman (1981, p. 25) for its use in the latter.
3 See Break (1980, pp. 99–101) for a discussion of this issue.
4 A model developed by Inman (1977, pp. 251–4) and applied to school spending in the New York metropolitan area (for the year 1968–9) suggested that the long-run effects of a change in grants might be 20–60 per cent below the initial effects.
5 See Break (1980, p. 90n) for a list of other median voter discussions.
6 See Wilde (1971, p. 146) for a discussion of a case where the grantee spends no money on a particular service in the pre-grant equilibrium.
7 See James (1973) for a discussion of non-constant prices.
8 The analysis here is based on the material in Wilde (1968, 1971). See also Waldauer (1973) and Thurow (1966).
9 See, for example, Waldauer (1973, pp. 219–20), and Wilde (1971, pp. 147–8).
10 A rationale for such a grant scheme is presented in Chapter 5 in the discussion of Figures 5.7, 5.8 and 5.9.
11 For discussions of how transfers could be made to citizens to produce precisely the same effects as the median voter model would predict, see Bradford and Oates (1971a,b) and Oates (1972, pp. 105–18).
12 Niskanen noted all these variables in 1968 (pp. 293–4) except for the output of the bureau, and he argued that all were positive functions of budget size. Later (Niskanen, 1973, p. 22), he inserted the output of the bureau in the list of variables, but then argued that all except the last two were positive functions of budget size.
13 See Niskanen (1968, pp. 295–6) and Breton and Wintrobe (1975, p. 197). Note that if both the AC and AV curves were straight lines, then it could be shown that OQ would be double the optimal output, though it would be less than double in the case illustrated in Figure 3.13.
14 It has been argued by McGuire (1973, pp. 209–11) that the citizens' gains from the increased bureau output precisely balance their losses from the reduced consumption of other goods and services that follows the increase in taxes.
15 With a lump-sum grant, the fall in average costs would be more marked at low output levels and less marked at high levels.
16 For another fiscal illusion model, see Courant, Gramlich and Rubinfeld (1979).
17 It may be noted that some models have been constructed to show how, in principle, the opposite could occur! See, for example, Goetz and McKnew (1972) and Jurion (1979).
18 As explained in Chapter 6, this belief is very plausible for owner-occupiers, and may well be substantially true for tenants, at least in the long run.

4

The Purposes of Grants

Introduction

The main purpose of this chapter is to consider the reasons why grants may be given. It is suggested that these reasons fall into four groups. First, there are grants designed to encourage subcentral authorities to take account of the external effects of their services. Secondly, there are grants designed to enforce grantor preferences. Thirdly, there are grants, termed 'revenue-sharing' grants, designed to correct for a fiscal imbalance between the various tiers of government. Finally, there are equalisation grants. These four groups are considered in turn.

Once any grant objective has been determined, the problem arises of designing a suitable grant, or scheme of grants, to achieve that objective. Accordingly, the discussions of the first two groups of grants are each developed to consider the relevant issues of design. As it happens, it is possible to design composite grant schemes that deal simultaneously with both revenue-sharing and equalisation objectives, and so it is appropriate to consider the design of such grant schemes after discussing those two objectives. However, so many factors are involved in the design of these schemes that it is necessary to devote a whole chapter to them; consequently, the discussion of these design issues is deferred until Chapter 5.

Broadly speaking, it appears that the first two objectives of grants are generally best tackled with specific grants, while the last two are perhaps best tackled with general grants. It seems appropriate, therefore, to follow the discussion of the first two groups with some observations about the use of specific grants in a number of selected countries. The later sections of the next chapter contain details of a number of general grant programmes.

It should be stressed that most of the discussions of grant design in this chapter and the next are couched in terms of the traditional model of grant effects discussed in Chapter 3. Thus, it is generally assumed that lump-sum general grants have similar effects to equal value rises in electors' incomes. To the extent that the former have greater effects on subcentral spending (that is, to the extent that there is a flypaper effect), the grants needed to stimulate subcentral spending may not have to be as large as the traditional model suggests. Indeed, the arguments for stimulating subcentral spending, which are considered in some of the sections of this chapter, might be weakened if there were a tendency for subcentral authorities to spend beyond optimal levels for the reasons given in some of the new models of grant effects examined in the previous chapter.

The discussions in this chapter and the next also embody the traditional view that matching grants have more effect than equal value lump-sum grants. Substantial modifications would have to be made to the analyses if this were not so, as some new models of grant effects imply; fortunately, the evidence suggests that it is.

Grants to internalise externalities

This section considers grants designed to encourage subcentral authorities to take account of any external effects that their services may generate. The situation most frequently considered is where one of an authority's services creates external benefits for citizens of other areas, and where the authority ignores these external benefits when deciding on the amount of service to provide. It is clear that its provision is unlikely to be optimal as it is not based on a consideration of all the benefits. It is clear, too, that an attempt could be made to design a grant that would encourage the authority to provide the quantity of service that it would provide if all the benefits were taken into account. It is tempting to conclude that ignoring external benefits typically leads to underprovision by the authority, and so, in turn, to conclude that the aggregate provision of subcentral services would be too small if externalities were ignored. These conclusions about underprovision have been much debated, however, and they will be considered in the next section.

It was noted in Chapter 1 that, in the absence of external benefits, an optimal quantity of a public good (Y) would be provided by a subcentral authority if provision were such that the marginal rate of transformation between that good and any private good (X), that is MRT_{xy}, equalled the sum of the citizens' marginal rates of substitution between the two goods, that is $\sum_{i=1}^{m} MRS_{xy}$ where the authority's population is m; thus an optimal level occurs where:

$$MRT_{xy} = \sum_{i=1}^{m} MRS_{xy} \tag{1}$$

So long as the MRT is less than this sum, then marginal switches of production from X to Y could enable all citizens to be made better off, while as long as the MRT is greater than this sum, marginal switches from Y to X could have a similar effect. If those running each subcentral authority sought to maximise the economic welfare of their citizens, then they would try to produce an optimal quantity of Y in the sense just explained. It is important to note that the quantity of Y thus provided would depend on both the level and the distribution of income in each area, since changes in those would alter the citizens' preferences for Y. In practice, the distribution of income would itself depend on the structure of taxes used to finance Y, and each authority would have to choose one

of the possible structures with its associated optimal quantity. Furthermore, if another level of government were to alter its level of taxes, possibly to finance a grant scheme, then the level (and probably the distribution) of residual incomes would alter, and each subcentral authority would face new circumstances. However, provided each authority were to create a new situation in which, once again, equation (1) was satisfied, then it would not be possible to alter the amount of Y provided and simultaneously make all citizens better off, and the new output pattern would still be Pareto-optimal.

Suppose, next, that the provision of Y in one area does confer benefits on the citizens in the others. For simplicity, assume that there are just two areas (A and B) and assume that all spillovers are public. This means that when area A, for example, produces Y, then its spill-outs raise the level of consumption of Y in B, but have no adverse effect on the level of consumption of Y in A. Now, if each authority were concerned to maximise the welfare of its citizens alone, then it would still set its output of Y at a point where equation (1) was satisfied. This would not result in an optimal provision of Y since benefits to non-residents would be ignored.

In principle, however, the situation could be rectified with the use of grants.[1] For simplicity, again, assume that X and Y are produced under conditions of constant cost, and that the respective costs are the same in each area. Suppose, further, that each unit of Y provided in A (or B) increases the total consumption of Y by each citizen in B (or A) by a constant fraction equal to ß (or α) of the increase in consumption that would result from an increase in production of Y by one unit in B (or A) itself. For example, an extra unit of street lighting in A might increase the consumption of lighting by each of A's residents by one unit, but increase the consumption by each of B's residents by the fraction ß of a unit. As a result, a citizen of B with an MRS_{xy} of, say, 1 would be prepared to give up one unit of Y to obtain one extra unit of X provided by B, but would be prepared to give up only ß units of Y to obtain one extra unit of X provided by A, so his MRS_{xy} in respect of A's output of X equals ß times his MRS_{xy} in respect of his home authority's output of X. In this case, optimal levels of provision of Y by each area would occur when the MRT_{xy} equalled the sum of all the relevant MRS_{xy}s, so that:

$$MRT_{xy} = \sum_{i=1}^{m} MRS_{xy(A)} + ß \sum_{j=1}^{n} MRS_{xy(B)} \tag{2}$$

and:

$$MRT_{xy} = \alpha \sum_{i=1}^{m} MRS_{xy(A)} + \sum_{j=1}^{n} MRS_{xy(B)} \tag{3}$$

are required in A and B respectively, where m and n are their respective populations. Note that, by assumption, each area has the same MRT_{xy},

and note also that if there were no externalities (so that $\alpha = ß = 0$), then each expression would reduce to that given in equation (1).

Since each area is held to consider its own citizens only, it is necessary to provide each with a grant that will reduce its effective MRT_{xy} to encourage increased production of Y. Suppose the grants met the fractions g_A and g_B of the costs of producing Y in A and B, then the respective new MRTs, which would generally be different, would be given by $MRT_{xy}(1-g_A)$ and $MRT_{xy}(1-g_B)$. The respective values of each grant per unit of Y, which will be designated G_A and G_B, would equal $g_A.MRT_{xy}$ and $g_B.MRT_{xy}$ (in terms of the cost of X) respectively.[2] Thus the new MRTs would equal $(MRT_{xy}-G_A)$ and $(MRT_{xy}-G_B)$. Each area would adjust its output so that its new MRT_{xy} equalled its own $\Sigma\ MRS_{xy}$ so that:

$$MRT_{xy} - G_A = \sum_{i=1}^{m} MRS_{xy(A)} \tag{4}$$

and:

$$MRT_{xy} - G_B = \sum_{j=1}^{n} MRS_{xy(B)} \tag{5}$$

would hold.

Now, it is necessary to adjust G_A and G_B so that (2) and (3) also hold. To find the required values for G_A and G_B, the four equations (2) – (5) must be solved for G_A and G_B; the equations also determine $\sum_{i=1}^{m} MRS_{xy(A)}$ and $\sum_{j=1}^{n} MRS_{xy(B)}$. From (2) and (3) it can be shown[3] that:

$$MRT_{xy}(1-ß) = \sum_{i=1}^{m} MRS_{xy(A)}(1-\alpha ß) \tag{6}$$

and:

$$MRT_{xy}(1-\alpha) = \sum_{j=1}^{n} MRS_{xy(B)}(1-\alpha ß) \tag{7}$$

By substitution into equations (4) and (5), it can be shown[4] that:

$$G_A = MRT_{xy}. \quad \frac{ß(1-\alpha)}{(1-\alpha ß)} \tag{8}$$

and:

$$G_B = MRT_{xy}. \quad \frac{\alpha(1-ß)}{(1-\alpha ß)} \tag{9}$$

It will be seen that $G_A/G_B = [ß(1-\alpha)]/[\alpha(1-ß)]$. Thus the proportion of grant going to A would rise if ß rose or if α fell. In other words, the areas generating relatively high spillovers acquire relatively large shares of the grant funds.

To understand the action of these grants, it is simplest to suppose that there is an external effect in one direction alone. Suppose that the flow were from A to B, so that there was no flow from B to A, and α equalled zero.[5] In this case, equation (9) shows that area B would need no grant since it would produce no externality. Equation (8) shows that A would need a grant (of G_A for each unit of Y produced such that $G_A = ß.MRT_{xy}$. From equation (3), with α equal to zero, it follows that $G_A = ß \sum_{j=1}^{n} MRS_{xy(B)}$. Now $G_A = g_A.MRT_{xy}$, so the fraction of the cost of Y met in A by the grant, that is g_A, equals G_A/MRT_{xy}, that is $ß \sum_{j=1}^{n} MRS_{xy(B)}/MRT_{xy}$. Using the expression for MRT_{xy} given in (2), this means that:

$$g_A = \frac{ß \sum_{j=1}^{n} MRS_{xy(B)}}{\sum_{i=1}^{m} MRS_{xy(A)} + ß \sum_{j=1}^{n} MRS_{xy(B)}} \qquad (10)$$

or, in other words, that g_A equals the fraction of the total marginal valuation of Y attributable to outsiders. Thus, if one-third of the benefits accrued to B, then A would require a grant paying one-third of its costs. This means it would set production of Y at a point where the net marginal costs (two-thirds of the total) equalled the internal marginal benefits (two-thirds of the total), so simultaneously equating total marginal costs and total marginal benefits. If α were positive, then more complex formulations for g_A and g_B would be derived from equations (8) and (9); these formulations would allow for the feedback effects generated in cases of reciprocal externalities.[6]

It is clear from this analysis that ascertaining the right level of grants to correct for externalities is no easy matter, and yet the analysis has made three important simplifying assumptions. First, it has assumed that the costs of producing X and Y are constant, so that MRT_{xy} is also constant. If the costs were to vary, then MRT_{xy} would also be liable to vary, though it should always be equal in each area if input costs were identical for each: if the total output of Y were to rise and that of X were to fall, and if MRT_{xy} rose as a result, MRT_{xy} should still be the same in both A and B. It follows that MRT_{xy} would depend on the final outputs of X and Y in each area, which cannot be predicted *a priori* unless each area's

preference map is known by the grantor. In practice, it may need a lengthy trial and error process to fix grant levels so that equations (8) and (9) are satisfied.

Secondly, it has been assumed that α and ß are constant. Suppose that ß falls when α's output of Y rises. In this case, the grant per unit paid to A, calculated as shown in equation (8), must also fall as its output of Y rises. Moreover, the grant paid to B would alter when A's output and ß alter, even if B's output of X were to remain constant, for equation (9) shows that B's grant depends on ß.

Thirdly, it has been assumed that there are two areas only. The model can be extended to include more areas. The result of adding one extra area (C) is to add two further equations, and two further unknowns such as $MRS_{xy(C)}$ and G_C. However, the resulting expressions for grants will be much more complex as there will be $a(a-1)$ parameters such as α and ß, where a is the number of areas; and all of these must be estimated before the grant sizes can be determined. Considering the last two problems alone, Oates (1972, p. 102) understandably concluded that specifying an efficient set of subsidies 'could become a monstrously complicated problem although still conceptually soluble'; and, of course, the problem needs solving for every subcentral service that generates externalities.

It should be noted that the grants determined in equations (8) and (9) meet a fraction of the costs of producing Y. They are, therefore, open matching specific (to Y) grants. It would be possible to design a set of lump-sum specific grants to have the same effect on each area's output of Y although, for reasons explained in the last chapter, a lump-sum specific grant would have no more effect on the output of Y than a lump-sum general grant, unless the deflective effect was in operation. However, there are two reasons for not using lump-sum grants. First, as noted in the discussion of Figure 3.5, they would cost the grantor more if they were to have the same effect. Secondly, each grantee would regard the lump-sum grant as a rise in income that did not alter relative prices. Thus A and B would produce X and Y in proportions that resulted in equation (1) being satisfied by each area, but not in proportions that satisfied equations (2) and (3). In other words, at their new income levels, they would not be producing Pareto-optimal levels of Y.

While the use of grants is consistent with Pigou's prescriptions (1932, pp. 172–203) for dealing with externalities, it has more recently been suggested by Coase (1960) that externalities may be resolved satisfactorily by the use of voluntary collective bargaining between the affected parties. The gainers from external benefits have an incentive to bribe the producers to produce more until a state of Pareto-optimality is produced, though optimality would be reached only if there were no costs involved in getting gainers and producers together to reach decisions on payment levels; and if participants were genuinely to reveal their preferences and did not indulge in strategic behaviour. Now, when externalities involve only a few subcentral authorities, it is possible that the grantor could encourage a bargaining process such as this to help it secure an optimal allocation of resources, and hence obviate the need for grants. Where many authorities are concerned, however, the decision-making costs could be very high,

and individual authorities might be tempted to understate the benefits they receive in order to become free-riders, and so the case for grants is stronger; indeed, a system of grants is almost inevitably the only solution in large number cases.

The effects of externality correction on subcentral service provision

Suppose that grants are used to encourage subcentral authorities to take external effects into account, and suppose that these grants follow the prescriptions indicated in equations (8) and (9). It follows that each beneficiary area will find the cost of Y lower than it would be without grants, provided only that grants are positive, which they will be in the two-area case if $0 < \alpha < 1$ and $0 < \beta < 1$. It therefore appears to follow that, so long as neither area treats Y as a Giffen good, each will produce more Y than it would otherwise. This, in turn, suggests that the introduction of grants, to achieve an optimal output of Y, necessarily raises the output of Y, and so implies that there would be underprovision of Y in the absence of grants. This conclusion ignores two important points, however.

First, the grantor may be expected to raise taxes to finance its grants, and its taxes will reduce incomes in grantee areas and so have a negative effect on the output of Y. Thus, in a given area, the effect of introducing a scheme of matching specific grants financed by higher taxes leads simultaneously to a price effect, which will encourage the production of Y, and an income effect, which will discourage production if Y is a normal good. Now, since grants essentially redistribute income, the overall effect of the scheme may be taken to leave the country's aggregate income constant. However, some areas will have, overall, a rise in income combined with a fall in the price of Y, and these should produce more Y, while others will have, overall, a fall in income combined with a fall in the price of Y, and these may produce more Y or less. If output were to fall in some areas and rise in others, then total output would not necessarily rise and might fall.[7] It should be noted that, if the grant scheme were to redistribute income between areas, then, although it might change the production pattern from a suboptimal one to an optimal one, the change itself would not be a Pareto-improvement, and a social welfare function would be needed to see if the move actually increased social welfare. As an alternative to raising funds by taxes, the grantor might prefer to raise them by charges imposed in each area that would relate to the value of the spill-ins received, a suggestion made by Musgrave and Musgrave (1976, p. 627). However, these charges would also be likely to result in some redistribution of income, in particular from areas receiving large spill-ins to areas generating large spill-outs, and once again aggregate output of Y could fall.

Secondly, some areas such as A or B that set their output levels of Y so that their equation (1) was satisfied – that is, in A's case, so that

$MRT_{xy} = \sum_{i=1}^{m} MRS_{xy(A)}$ – might cut their outputs of Y when grants were introduced even if their outputs of Y generated external benefits (so that they received grants) and even if their incomes (net of taxes levied by the grantor to finance the scheme) did not fall. This might appear surprising since their grant receipts would reduce their effective MRT_{xy}s. The reason is that other areas might respond to the grant scheme by substantial increases in their outputs of Y, thereby raising spill-ins and consumption levels of Y in A and B, and substantially reducing the MRS_{xy}s of A's and B's citizens. To ensure that equation (1) still held, A and B might need to cut their output of Y, and such cuts might exceed the rises elsewhere.

Consider a case where there are three areas only (A, B and C) and where the pre-grant production levels, spill-ins and per capita consumption levels are as shown in Table 4.1. It will be seen that C's production generates the highest relative externalities; C will therefore receive the highest percentage grants and will doubtless raise its output of Y to the benefit of A and B who may consequently contract their own outputs of Y. The final result could be as shown in Table 4.2. Consumption rises in each area, yet the total production of Y falls from 300 to 290.[8] In effect, the grant scheme switches production to those areas generating the highest relative spill-outs; in the example given in Tables 4.1 and 4.2, it appears

Table 4.1

	Area A	*Area B*	*Area C*
Production of Y	100	100	100
Spill-in from A	–	10	10
Spill-in from B	10	–	10
Spill-in from C	80	80	–
Consumption of Y	190	190	120

Table 4.2

	Area A	*Area B*	*Area C*
Production of Y	70	70	150
Spill-in from A	–	7	7
Spill-in from B	7	–	7
Spill-in from C	120	120	–
Consumption of Y	197	197	164

that consumption was suboptimal before the grants were introduced while production was, in a sense, super-optimal.

Grants to enforce grantor preferences for merit goods

If subcentral authorities were concerned to maximise the economic welfare of their citizens, and if they were given suitably designed matching grants to ensure that external benefits were taken into account when their budget decisions were reached, then each of those authorities would seek to provide each service at a Pareto-efficient level. There might be a case for further grants, however. For example, the authorities might be thought to need revenue-sharing grants on the grounds that they had insufficient income of their own to provide services at a reasonable level, and there might be an equity case or an efficiency case for equalisation grants that would benefit poor areas; these two matters are discussed in later sections. The present section and the two that follow concern grants that are made by the grantor in order to encourage grantees to adjust the pattern or overall level of their budgets so that they accord more closely with the preferences of the grantor. Mathews (1980a, pp. 27–8) has noted that such grants represent a coercive use of the grantor's power, and argued that when grantees accept the grants then they probably feel that their own tax resources are inadequate. The grantor's motives have been interpreted in various ways in the literature.

One interpretation, suggested by Musgrave and Musgrave (1976, p. 628), is that they arise in situations where one, some, or possibly all, subcentral services are viewed as merit goods by the grantor – that is, the services in question are considered meritorious by the grantor, who believes grantees should provide them at a higher standard than they would in the absence of grant aid. The grantor might well use open matching specific grants for each individual service that it wanted expanded, though it might use closed grants if there was some relatively high level of consumption that it felt was adequate. Alternatively, the grantor might use general grants if it viewed all subcentral services as merit goods. In this case, it might use effort-related general grants rather than lump-sum general grants, for the former will have bigger effects on the services concerned than equal value lump-sum grants. It is possible, though, that the grantor would choose lump-sum specific or general grants combined with rigorous conditions designed to ensure compliance with its preferences.

At first sight, any such grants might seem wholly beneficial from the grantee's point of view, since they appear to confer an increase in income. However, the grantor must raise funds in order to disburse grants and, unless part of its taxes can be exported to other countries, grantees' citizens overall will pay as much in extra taxes as their authorities receive in grants. For the citizens of many authorities, the effect of a matching grant favouring one service (X) could be close to that shown in Figure 4.1, which relates to a period in time. Initially, the authority faces the budget line AA' and selects point J on the community indifference curve II. The introduction of an open matching specific grant for X reduces the

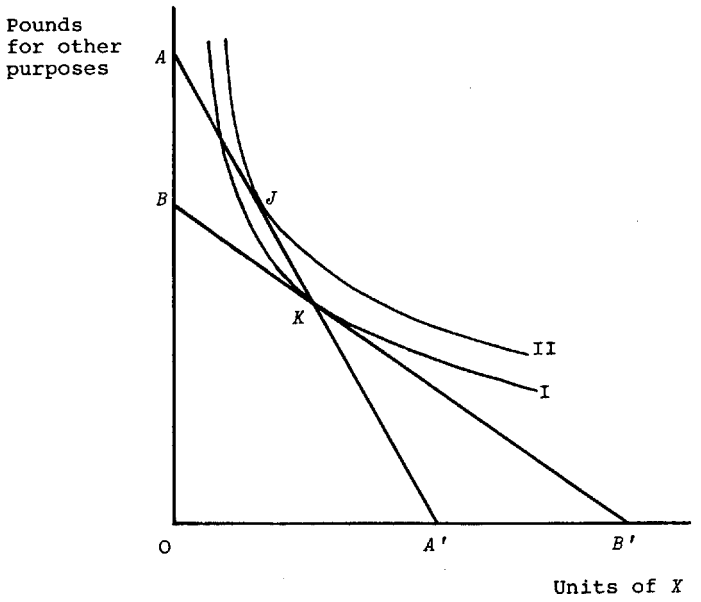

Figure 4.1

absolute slope of the budget line, but the grantor taxes needed to finance the grant shift the budget line towards the origin. The authority represented in the figure faces the new budget line BB' and opts for point K on the inferior indifference curve I. K is, by chance, on AA', showing that the authority's citizens could have afforded that position initially and hence have the same total level of expenditure as before. Some authorities, in particular those paying the most in grantor taxes, will end up inside AA', while others may end up outside and could reach higher indifference curves than initially. It is clear, however, that an area could reach a higher curve only if it received net funds at the expense of other areas and so could spend more than before; if there were no change in income, as in Figure 4.1, then it would be on a lower curve.

A similar result would follow if the grantor used lump-sum grants. These would not alter the slope of the budget line, and the grantee could reach a higher indifference curve than its initial one only if its grant exceeded the grantor taxes paid by its citizens to finance the grant programme. In fact, it could end up worse off even if its net income position were unaltered, or, for that matter, increased. This situation would arise if the conditions attached to the grant receipts forced it to buy more than its preferred number of units of X.

It follows that the introduction of such grants could leave many recipients, and possibly all of them, on lower indifference curves. A number of justifications for the grants can nevertheless be advanced. First, they may apply to services about whose benefits the grantee citizens are partially ignorant. In such cases, Musgrave and Musgrave (1976, p. 65)

implied that a temporarily imposed consumption choice may be desirable as part of a learning process, in order to permit a more intelligent choice thereafter.[9] Wiseman (1981, p. 374) has noted that if a specific grant is successful in encouraging grantee provision of a service, then the grantor may come to believe that it is no longer necessary to retain the specific grant, and it may then absorb it into any general revenue-sharing grant it gives. He argued that this type of progression from specific into general grants has occurred in the United Kingdom, and helps to explain the interest in (general) revenue-sharing in the United States. One problem with this argument is why the grantor is more likely to be persuaded about the merits of the service concerned than the grantee. Grewal (1981a, pp. 38–9) has offered an explanation in the case where the grantor is the central government and where grants are being used to encourage the provision of a wholly new service; he has defined such a purpose as the seed-money function. He suggested that interest groups are often instrumental in stimulating governments to provide new services, and that they may have more success in dealing with central governments, which will be pressurised by the entire nation's pressure group, than they have with subcentral governments, each of which may be pressurised by only a segment of the group. As examples of grants encouraging new ideas, Mathews (1980a, p. 31) has cited the school grants from the Australian government to the states for special educational projects, and grants by the Indian government to states for pilot agricultural projects.

Secondly, merit good grants could be applied to services where grantee citizens (or, more precisely, grantee voters) ignored benefits accruing to non-voters. Such grants could be used to justify grants for education in order to protect school children from the underprovision that might be preferred by adult voters. It must be observed, though, that grantor governments are also answerable to adult electors, so such grants might be rare. The most plausible scenario for this argument is one where the majority of the nation's adults do take school children's interests into account, but where only a minority do in some areas; in this case the median central voter may well demand action for the benefit of school children in areas whose median voters ignore them.

Thirdly, Musgrave and Musgrave (1976, p. 628) have argued that grants could be offered to encourage the provision of services that benefit only a minority of the population and that are therefore likely to be offered at low levels, if at all. Such services might include coastal flood defences, museums, libraries and the arts. Musgrave and Musgrave have contended that the grants may be needed to secure reasonable levels of services. This sort of grant, too, may be rare, since there is once more the problem that if a majority of grantee voters were unwilling to subsidise a service, then the same might well be true of a majority of grantor voters.

Fourthly, as Dafflon (1977, pp. 41–2) has pointed out, specific grants for some services might be used for revenue-sharing or equalisation purposes, even though later sections of this chapter suggest that there are strong arguments for using general grants in each case. The reason why general grants might be eschewed is that, if a country's citizens were obliged to subsidise some (or all) subcentral authorities, they might insist that those

authorities received an array of specific grants for specific services, to encourage the provision of those particular services, rather than general grants to be spent entirely at the discretion of the grantees. Musgrave and Musgrave (1976, p. 66) have suggested that this situation could arise if donors are more willing to make gifts when they are earmarked.[10] This argument suggests that the existence of merit goods is linked to voluntary redistribution. Any grants made for these reasons would have to depend on a grantee's taxable resources, and might depend also on its needs or its spending on the functions concerned.

Grants to enforce grantor preferences for minimum standards

A different but related interpretation of grants to enforce grantor preferences is that they are used to help secure nationally determined minimum standards, though it must be pointed out that legislation may be a more certain and, for the grantor, a much cheaper method of achieving such an end. A number of reasons have been suggested why minimum standards might be sought by a grantor. For example, Dafflon (1977, pp. 39–40) has suggested that minimum standards may be needed to protect minorities who 'may not have a proper migration alternative'. In such cases, it must be supposed that a national majority of voters are concerned with a problem that does not concern a majority in every area. Also, it should be stressed that, while attempts to secure minimum standards may well help those minorities in areas with low standards who would prefer higher service levels but cannot, for some reason, vote with their feet, they do so only by preventing the majorities in those areas from having the service levels they would prefer. Dafflon has also suggested (pp. 40–1) that minimum standards might be desired in order to discourage the underproduction that could occur if externalities were not fully internalised by the appropriate grants discussed in the previous section, a point made too by the Layfield Committee in the United Kingdom (HMSO, 1976, pp. 59–60).

Another case could be made for pursuing minimum standards if two tiers of government provided related services, such as schools and universities, roads and railways, or homes for the aged and pensions. Here, low spending on one service by one authority might well impose a need for more spending by another. In these cases, co-ordination is required, and it may have to be backed up by attempts to secure minimum standards.

A further possible reason for seeking minimum standards was noted in Chapter 1, namely that many citizens in a country expect to have to to leave their present areas at some time in the future, and so have some concern over levels of services in all other areas. It could perhaps be argued, however, that the real problem here is that areas with low service levels overlook the external benefits that higher levels would give to potential immigrants, and that this problem should be treated along with other externalities. It could also be argued that if subcentral authorities were relatively large, then the range of chosen service levels might anyway be small, while if the authorities were relatively small then enforced migrants

might well have a reasonable choice of areas in which to reside whenever they made a move.

A final case for minimum standards arises in connection with subcentral redistribution policies, whether implemented directly with cash handouts or indirectly with merit good programmes. It was explained in Chapter 1 that there could be migration problems if the degree of redistribution in different areas varied by more than modest amounts. In fact, if recipients were in a minority, then voters in some areas could be tempted to make their programmes less attractive than those elsewhere, for this would involve lower tax rates at once, and could result in further cuts if the lower benefits drove recipients away. Accordingly, there could be a widespread desire for minimum standards to be set by some higher-tier authority. Again, though, it could be argued that the real problem is that areas would be inclined to ignore the external effects of changes in their redistribution policies.

Where an attempt is made to secure minimum standards with the use of grants, the Layfield Committee (HMSO, 1976, pp. 60–2) suggested that, in principle, the appropriate grants would be lump-sum specific grants paid by the central government and would be equal to the difference between what an authority would choose to spend on a service and what the grantor wished it to spend. In practice, the Committee recognised that formidable problems existed in defining standards, let alone costing them annually for each area. Moreover, grantees would have a clear incentive not to reveal their preferences and to declare no interest in the services concerned. The surest way of using grants to ensure that minimum standards were attained would be to have lump-sum specific grants which met the whole cost of providing each service at the standard required. Of course, this would be very costly to the grantor and, as Oates (1972, pp. 92–3) has noted, it might be able to raise standards to the level desired in most areas, if not all, by an array of matching specific grants that involved a much lower aggregate outlay.

However, if minimum standards were successfully enforced, then it could well be argued that subcentral authorities were acting as agents of the grantor, and not as independent bodies, except on those occasions when they provided services at levels beyond the minimum ones. On this view, it could also be argued that any grantor wishing to secure minimum standards should finance the whole cost incurred in meeting those standards.

Grants to enforce grantor preferences for macroeconomic reasons

Another interpretation of grants used to enforce grantor preferences concerns macroeconomics. There are essentially two separate issues here. First, if the grantor were the central government, then it might properly have some interest in the level (and perhaps the composition) of subcentral spending. This interest could arise for many reasons, including concern with the overall level of demand in the economy and the balance of demand

between the private and public sectors. The relevant issues are discussed in Chapter 9. For the present, it should be noted that, if total subcentral spending were to be influenced by grants, then the appropriate grants might be effort-related general grants, which, it has been seen, generate greater changes in grantee outlay for each dollar change in grants. Moreover, the larger these grants were in relation to total grantee revenue, the larger the effect the grantor could hope to secure.

Secondly, subcentral authorities could be allocated some role in regional development. As noted in Chapter 1, this role could vary from being, in effect, agents of the central government to having the primary responsibility for formulating, implementing and financing development policies. If the authorities acted essentially as agents, then presumably their development activities would be financed by specific grants and would be subject to close control. If they were essentially autonomous, then they might expect no specific grants, although their development needs might affect their revenue-sharing and equalisation grant receipts; however, as explained in Chapter 1, the central government might well wish to exercise some overall control, and it might attempt to do so by paying specific grants with conditions attached.

Specific grants in practice

The previous sections of this chapter indicate that specific grants are usually the appropriate instruments for grantors to use if they wish to encourage subcentral authorities to take account of their externalities or if they wish to enforce their own preferences. In contrast, later sections will suggest that the other purposes of grants, namely revenue-sharing and equalisation, can perhaps be tackled best with general grants. Accordingly, it seems appropriate to pause for a moment to see how far the specific grant arrangements in selected countries can be related to the two purposes so far discussed.

Grant aid in the United States relies particularly heavily on specific grants: such grants made up 88 per cent of federal aid in 1978 and 90 per cent of state aid in 1977.[11] The term specific grants is used here to cover three groups of grants that are frequently distinguished from one another, namely project categorical grants, allotted formula categorical grants and block grants. In this context, categorical grants are narrow specific grants, while block grants are broad specific grants. All three types are used for aid by the federal government, although the present federal administration is seeking to increase the importance of block grants at the expense of categorical grants. Almost all specific state grants to local authorities, on the other hand, are in the form of formula categorical grants. Allotted formula categorical grants and block grants are distributed on the basis of formulae to all eligible grantees, both the formulae and the eligibility criteria being established by law or regulation. With project categorical grants, the grantor invites applications for assistance in the funding of some specified activities; if a total grant payment has been set and the combined value of the applications exceeds this total, then the grantor

may have considerable discretion in deciding which applicants will benefit.

Most of the federal grant outlays on these specific grants are related to five main functional areas. In 1978, for example, some 12 per cent of these grants were concerned with transportation, 18 per cent with health, 20 per cent with income security, 19 per cent with education, and 14 per cent with training, employment and other social services. Specific grants paid by states to local authorities focused on similar items, with education forming 67 per cent of such grants in 1977, public welfare 16 per cent and highways 7 per cent. These services also dominated the grants a few years earlier when Herber (1975, p. 81) argued plausibly that the transportation grants, which related chiefly to roads, could be seen as concerned with the correction of externalities. He supported this contention by pointing out that federal legislation dating from 1956 had used specific grants to states for highways as part of its efforts to secure a large network of high-speed and limited access inter-state highways. Herber also argued that the educational grants could be explained with reference to externalities (p. 81) but that welfare grants could not (p. 85).

However, the existence of specific grants concerned with education, health and income security is perhaps capable of various interpretations. For example, it could be argued (as discussed in Chapter 1) that people are concerned about the welfare of citizens in all jurisdictions, not just their own, in which case there could be substantial external benefits for all those services and not from education alone. Alternatively, the grants might seek to encourage minimum standards, which (it has argued earlier) may be needed for services with a large redistributive impact. Finally, education and health services may be supported for merit good reasons.

It should be appreciated that specific grants in the United States are not characterised by stringent matching requirements. For example, Break (1980, p. 131) has noted that, in 1975, 39 per cent of the federal government's categorical grant programmes had no statutory matching requirements, and only 13 per cent required grantee matching of 50 per cent or more. This suggests that the aim of encouraging grantees to take externalities into account may frequently be unimportant, for it has been seen that the most suitable grants for this purpose are matching ones. Moreover, Oates (1972, p. 89) has observed that most of the matching grants are closed matching specific grants, and they are closed, typically, at modest spending levels; while it is possible that correcting for externalities would justify matching grants whose proportions diminished as grantee spending rose, he felt it unlikely that it would justify closed grants since that would imply that, at a certain level of public provision, spillover benefits suddenly became zero. It could be that some specific grants are concerned to promote minimum standards or merit goods, but the fact that many programmes, particularly block grant programmes, pay great heed to grantees' resources and needs suggests that many may primarily have equalisation and revenue-sharing objectives. They do not prove this, however, for it could be that higher grants to low-resource and high-need areas are needed merely to persuade them to achieve satisfactory service standards. Nevertheless, it seems fairly clear that those objectives do form part of the *raison d'être* of many programmes.

Specific grants play a dominant role in Canada too. In 1978, for instance, they made up about two-thirds of federal aid to the provinces and six-sevenths of provincial aid to local authorities.[12] As in the United States, health, education and general welfare programmes accounted for most of the federal aid: these three services received respectively 46 per cent, 21 per cent and 23 per cent of total specific grant payments, leaving only 10 per cent for other items. Much of the aid for the three services concerned is paid as part of the Established Programs Financing (EPF) arrangements introduced in 1977 and is paid in the form of lump-sum grants with the intention of securing a basic level of service in all parts of the country, although the grants are paid unconditionally so that there is no guarantee that the federal government's desired standards will always be met. Indeed, the federal government's dissatisfaction with the results of EPF is likely to strengthen its resolve to cut the sums involved, though as Macdonald has observed (1982, p. 150) it would seem more appropriate for rules or conditions to be attached to the grants. Other specific federal grants are generally in the form of matching grants, typically involving the federal government in meeting half a province's expenditure on particular programmes. These matching provisions could mean that the other grants are concerned with externalities or merit goods, but in practice the main concern seems to be with minimum standards (ACIR, 1981d, p. 61). It is, however, arguable that the grants are used also as substitutes for general revenue-sharing grants, and Burns (1977, p. 61) has noted that the health grants contain explicit equalisation elements.

The pattern of specific-purpose grants in Australia has been dominated by the same items as those in North America. In 1978–9, for example, 13 per cent of specific grants paid by the commonwealth government to states and local authorities were accounted for by transport, 45 per cent by education and 36 per cent by health, welfare and social security.[13] Mathews (in ACIR, 1981c, pp. 10–11) has pointed out that these specific grants have incorporated elements of horizontal fiscal adjustment, thereby reducing to some extent the need for general equalisation grants, and that in the early 1970s particularly they were used as a substitute for general revenue-sharing grants for which there was a growing need. However, recent developments in general grants (outlined in the next chapter) have reduced this last case for specific grants (though they have not affected the main programmes concerned), and these fell from 49 per cent of commonwealth payments to the states in 1975–6 to around 43 per cent in 1979–80.

Many of the specific grants in Australia refrain from making burdensome matching requirements. By way of contrast, specific grants in the United Kingdom are usually matching grants. Further, specific grants to local authorities in the United Kingdom accounted for a mere 21 per cent of central government grant payments in 1981.[14] These facts suggest that the grants there may be sharply designed for the purposes of handling externalities and enforcing strictly limited grantor preferences; certainly they have little part to play in revenue-sharing or equalisation, for which purposes the general grant programme is specially designed (as explained in the next chapter). However, specific grants are not really necessary to

enforce grantor preferences in the United Kingdom, for these are addressed in a wide array of detailed government controls. The inference is that specific grants are used principally for externality correction, and indeed externalities may well be important in the case of some of the services attracting this form of grant, particularly transport and roads, which attracted 15 per cent of specific grant payments in 1981. However, the importance of externalities is less clear in the main functions eligible for specific grants, namely police services and the administration of justice, which attracted 42 per cent. Education attracted 17 per cent, but the funds concerned were substantially used by grantees to finance their maintenance grants to students, a service they effectively perform for the central government on an agency basis. It is possible that the central government also regards local authority provision of police and administration of justice services as agency activities, and hence feels obliged to make specific grants for them, but these grants cover only about half local spending on those items.

Grants to correct for fiscal imbalance

It is argued in Chapter 8 that subcentral authorities should generally use loans alone to finance capital spending and taxes and charges alone to finance current spending. A third broad purpose of grants is to compensate grantees for any mismatch there may be between their aggregate current expenditure needs on the one hand and their aggregate tax-raising capacity (plus charge-raising capacity) on the other hand; such a mismatch is sometimes termed a fiscal imbalance, or, more precisely, a vertical fiscal imbalance. It is important to notice that this purpose would apply only if the authorities in a tier had an imbalance when considered as a group. On some occasions there may be no imbalance in aggregate, yet some authorities may have inadequate revenue capacities while others have more than adequate revenue capacities. This situation of horizontal fiscal imbalance calls for an equalisation scheme, and is discussed at length in the following sections. It should be stressed that aggregate vertical fiscal imbalances have increased in recent years in many countries, particularly where subcentral authorities have been dependent on taxes whose yields (at constant rates) do not rise as fast as real or money incomes, for their expenditure needs have often risen faster.[15]

It might be thought that there need never be a mismatch between expenditure needs and tax-raising (and charge-raising) capacity if subcentral authorities had access to at least one tax, for then they could always raise adequate funds simply by raising their tax rates. However, there might be freak situations, such as an authority with access to a tax on motor fuel which found that the tax rate that produced the highest possible yield still produced a revenue too low to finance a level of expenditure considered adequate by the grantor. Even so, it would be more precise to say that a revenue deficiency exists when the aggregate level of expenditure that the grantor would like to see falls short of the tax (and charge) revenue that could be raised at the subcentral tax (and charge) rates that it would like to see.

A grantor faced with this situation has broadly four options. First, it could encourage or empower (depending on the legal framework) the subcentral authorities concerned to raise their revenue from charges. Secondly, it could encourage or empower them to levy new taxes. Thirdly, it could institute a formal arrangement for one (or more) of its own taxes whereby each authority automatically received a uniform percentage, or possibly the whole, of the yield of the tax (or taxes) generated in its own area; this arrangement will be termed 'tax-sharing' since each authority would acquire a specified share of some of the tax revenue generated in its own jurisdiction. Fourthly, it could give out grants to the authorities, in which case the receipts of any particular grantee might depend directly on factors such as its own revenue or spending effort or its spending needs, or inversely on its own tax resources; such a system of grants will be termed 'revenue-sharing' grants since the grantor eases the revenue problems of the grantees by sharing out with them some of its own total revenue. Of course, the grantor need not rely wholly on any one of these four possibilities, but could use any combination of two or more. However, it is perhaps fair to say that there are reasons why the grantor might wish to make only limited use of the first three options. It is necessary, therefore, to note those reasons, and then discuss what form revenue-sharing grants might appropriately take.

As far as charges are concerned, there are certain principles that can be laid down to suggest when they should be used and how they should be determined in order to help secure an efficient allocation of resources; the issues involved here are discussed in Chapter 7. Thus, if it were thought that the correct use of charges was already being made, then clearly it would be inappropriate to make further use of them to ease the problem of fiscal imbalance.

The principal problem with attempts to extend subcentral taxation powers is the difficulty of finding suitable taxes for subcentral authorities to collect. Chapter 6 discusses this issue in detail, and it notes a number of criteria that should be met before a tax could be considered wholly satisfactory when raised at a subcentral level. It is not difficult to show, nor, perhaps, is it surprising to find, that no tax is perfect from this point of view, and this conclusion may lead to a reluctance to encourage further subcentral taxation powers. However, it should be noted that a significant level of subcentral taxation is important for two crucial reasons.

First, subcentral taxes give the authorities some discretion over their total revenue, and therefore ease their attempts to tailor the levels and types of their services to the tastes of their citizens; the analysis in Chapter 1 implies that without these attempts, much of the case for subcentral provision would disappear. It may be noted that an authority's ability to alter its total revenue will depend not only on how much revenue it raises in taxes over whose rates it has discretion, but also on whether it is in receipt of grants whose values depend in any way on its own tax effort. Indeed, if all an authority's grant receipts were proportional to its tax revenue, then a doubling of its tax effort would double its total revenue (from taxes and grants) irrespective of the proportion of the total that was attributable to taxes. There are, however, problems with grants that

include a tax-effort factor, as is shown in the next chapter, and these problems are more serious the larger such grants are in relation to tax revenue. Accordingly, grants of this type seldom form a large part of grantees' incomes, and so autonomy over revenue levels does generally require substantial independent taxation power.

Secondly, subcentral taxes reduce the extent to which subcentral authority activities will be financed by grants and thus reduce the importance of one reason for which they could legitimately be subjected to control by the grantor. The grantor is necessarily responsible to its taxpayers for the way in which their tax payments are spent, and therefore it is quite understandable that some control will emerge over the way in which grantees spend grant receipts. These controls may not take the form of making all grants specific grants, and may not even manifest themselves as conditions attached to grant payments. Rather, they may appear to be unrelated and take the form of legislation, circulars or ministerial controls.

The grantor must weigh up the advantages of subcentral tax power against the disadvantages of possible individual subcentral taxes. This judgement should bear in mind the fact that grantees may already be substantially dependent on grants paid for other purposes. However, the judgement may well come down against extending subcentral taxes to the point where no revenue-sharing or tax-sharing arrangements are needed.

Tax-sharing arrangements, as here defined, do not give individual subcentral authorities the first advantage of subcentral taxes, namely the ability to set their own tax rates, whereas revenue-sharing grants can be made to relate to grantees' own tax efforts. Tax-sharing arrangements may or may not give them the second advantage, namely avoiding grantor control. This latter advantage depends very much on the constitutional arrangements lying behind the tax-sharing scheme. If these arrangements guaranteed some specified share of grantor taxes to grantees as of right, then the grantees' revenues would be seen to be unquestionably theirs, and the grantor would not be responsible to its taxpayers for the way in which they were spent. If the arrangements were made at the unilateral discretion of the grantor, on the other hand, then that body might feel it had every right to oversee the way in which the grantees spent their revenue, just as it would feel in the case of grants. In general, therefore, tax-sharing arrangements have no special attraction except where subcentral receipts are constitutionally guaranteed. The most important situation where this sort of guarantee operates is perhaps in West Germany, and Chapter 6 contains a brief outline of the arrangements used there. As it happens, there are a few other tax-sharing schemes where such guarantees do not exist. For example, some states in the United States share the motor vehicle tax with local authorities and distribute the proceeds among the authorities according to where the vehicles are principally garaged.

Grantors may reduce the need for revenue-sharing grants by relying on the alternatives, but virtually all use such grants to some extent. In any event, as Mathews (1980a, p. 19) has pointed out, even if subcentral authorities were at some stage given tax (or other revenue) powers commensurate with their spending needs, and hence had a fiscal balance,

then the yields of different taxes and the expenditure needs of different levels of government could fluctuate over time and destroy this balance. From time to time, it might be possible to restore the balance by changing subcentral taxation and spending powers; such changes cannot easily be made frequently, however, nor are they necessarily susceptible to fine-tuning, and there is thus always likely to be a residual need for revenue-sharing grants.

Where revenue-sharing grants occur, there is clearly a case for general grants, since the necessity for these grants does not relate to any particular services. Even so, it has been noted earlier in this chapter that the problems of fiscal imbalance may be partially, if not wholly, addressed with specific grants. Where specific grants are used, then the grantor may well opt for matching grants, since it was suggested in Chapter 3 that lump-sum specific grants generally have the same effects as lump-sum general grants. Where general grants are used, the grant payable to any grantee could be made dependent on various parameters: it could, for instance, be a per capita grant, or it could instead be related more explicitly to the area's spending needs, its taxable capacity and, perhaps, its revenue effort. Similarly, specific grant payments could be made to depend on these factors as well as on spending on the services concerned.

Later sections put forward arguments for general equalisation grants that put all grantees on an equal footing irrespective of their spending needs and taxable capacities; such grants might also be made to depend on grantees' revenue efforts. Where these arguments are accepted, then it is likely that grantees' revenue-sharing grant receipts, and particularly any receipts from general revenue-sharing grants, will be taken into account when calculating their equalisation grant receipts. This can be most easily accomplished by having a single scheme of general grants to handle both equalisation and revenue-sharing functions. A pure equalisation scheme can be readily modified in this way simply by allowing it to have a positive net cost; that is to say, by ceasing to require the positive grants to poor or needy areas to be financed fully by negative grants or precepts on the others, but permitting more, if not all, areas to be beneficiaries and meeting the net cost of the scheme out of the proceeds of the grantor's taxes. The design of grant schemes intended to tackle both equalisation and revenue-sharing is a complex subject, which forms the material for the next chapter. It should be noted, though, that if it were not accepted that all grantees should be out on an equal footing in this way, then there would be no point in having an equalisation scheme and, perhaps, no point in trying to allocate revenue-sharing grants in relation to taxable capacity or needs. In these circumstances, allocations might be made on some other basis such as a specified sum per capita.

The equity case for equalisation grants

The remaining sections of this chapter consider equalisation grants, which are used widely to neutralise, either partially or wholly, any differences there may be in tax resources and spending requirements between different

areas within a tier of subcentral authorities. The principal case for these grants stems from horizontal equity considerations, and it is examined in this section. A secondary argument for equalisation grants relates to concern on efficiency grounds with the migration that might occur if horizontal inequity were permitted to persist, and this concern is discussed in the next section. Although the case for equalisation grants is widely supported, it is possible to put forward arguments to suggest that they are not necessary, and later sections of this chapter consider three of these: first, there is the possibility that equity would be achieved without grants as a result of adjustments in land prices; secondly, there is the possibility that subcentral authorities could finance their activities by benefit taxes, so that, without grants, any two citizens in a country who paid equal taxes would receive equal benefits and be equitably treated; and, thirdly, there is the possibility that equalisation grants could be replaced by a system of central government taxes whose rates varied from area to area in order to help citizens in poor or needy areas. The final two sections of this chapter consider whether the introduction of equalisation grants would itself have adverse side-effects on the productive efficiency of a country. The widespread existence of equalisation schemes suggests that not all grantors accept that they are unnecessary or undesirable. Many issues arise when designing equalisation schemes and these are discussed in the next chapter.

Table 4.3

		Symbol	*Area A*	*Area B*
1	Unit costs (index)	C	1.00	1.00
2	Unit needs (index)	N	1.00	1.00
3	Income per head ($'000s)	T	20.00	16.00
4	Tax rate (proportion)	t	0.05	0.05
5	Tax yield per head ($'000s)	tT	1.00	0.80
6	Service levels (index)	tT/CN	1.00	0.80

The equity case for equalisation grants can be illustrated with the help of an example such as that given in Table 4.3, which relates to two subcentral authorities (A and B) in a country where the currency unit is called a dollar. The first line, the unit cost index (C), reflects any difference there may be in the costs of producing a unit of output of public services. For example, it might cost more to put out each fire in a remote area if fuel prices were higher and hence the fire engine's running costs were higher, and it might cost more to educate each child in a sparsely populated area if school buses had to run more miles per pupil. In the example, it is assumed for the moment that the weighted average of all these costs is the same in each area, and, for convenience, that it equals some standard level determined by the grantor and defined as unity. This standard

might be an average figure, or the figure for some area taken to be a standard area, or simply an arbitrary level. The second line refers to the unit need index (N), and reflects any factors that will affect the number of units of output per capita needed to secure given levels of services. Thus unit needs will tend to be high where there is, for example, a high ratio of schoolchildren to be taught, in relation to population, or a high proportion of criminals to be caught, in relation to population. In this example, it is again assumed initially that the weighted average of all needs factors for each area equals some standard level defined as unity. The third line relates to the areas' per capita tax bases (T). It is supposed that each area levies a proportional income tax and that the levels of per capita income can be taken to indicate per capita tax bases. It can be seen that A's per capita tax base is higher that B's.

Suppose, now, that each area levies a tax rate (t) of 5 per cent, as shown in the fourth line. In this case, the per capita tax yield will be higher in A than in B, as shown in the fifth line. In the absence of any grants, A's expenditure per head will be higher, and so, assuming comparable efficiency in service provision, its service levels will also be higher. This is reflected by the service level index given in the sixth line; for present purposes this is defined as expenditure per head (in \$'000s) divided by both the unit cost index and the unit need index. Now, if Mr α in A and Mr ß in B were equals in that each had an income of \$30,000, then each would pay \$1,500 in tax, but they would receive different levels of benefit in return. A similar situation would arise with differences in unit needs or unit costs. Table 4.4 shows what would happen if the two areas had the same income per capita and the same unit cost index, but if B had a higher unit need index of 1.25. Here, if each area set a tax rate of 5 per cent, as shown, then B would have a service level index of 0.80 compared with 1.00 for A. The same result would occur, if, for example, B had a unit need index of 1.00 but a unit cost index of 1.25.

Table 4.4

		Symbol	*Area A*	*Area B*
1	Unit costs (index)	C	1.00	1.00
2	Unit needs (index)	N	1.00	1.25
3	Income per head (\$'000s)	T	20.00	20.00
4	Tax rate (proportion)	t	0.05	0.05
5	Tax yield per head (\$'000s)	tT	1.00	1.00
6	Service levels (index)	tT/CN	1.00	0.80

It follows that if Mr α and Mr ß live in areas with different unit costs, different unit needs or different tax bases, then they must expect to find

equal tax payments securing different service levels. Thus, to use a term suggested by Buchanan (1950, p. 588), they would be likely to have unequal fiscal residuums – that is to say, unequal differences between the value of benefits received and taxes paid. It is not certain that this will be the case for any two individuals, however; A might devote most of its effort to services of little value to Mr α while B might concentrate on services of great value to Mr ß. Even so, B cannot give the same residuums to all its citizens that their equals in A obtain.

A suitable equalisation grant to B would enable it to provide equal fiscal residuums to those in A. In the example considered, B suffered in turn from a tax base 20 per cent below A's, unit costs 25 per cent above A's, and unit needs 25 per cent above A's. In each case, a grant to B equal to 25 per cent of its tax yield – that is, a grant of $200 per head – would raise its service level index (suitably redefined to equal $(tT+G)/CN$ where G equals per capita grant receipts) by 25 per cent, that is from 0.80 to 1.00, so that it equalled A's. Accordingly, B would be in a position where it could give all its citizens the same fiscal residuums that their counterparts in A obtain, although it would probably have to provide the same pattern of services to do so.

A decision that horizontal equity is desirable is a value judgement, but clearly, where this view is held, equalisation grants will have a strong appeal. It should be noted, however, that the grant does not necessarily secure such equity. Even with the grant, Mr α and Mr ß might face different residuums if A and B had different service patterns, or, in a case where authorities could raise more than one tax, if A and B made equal overall tax efforts but relied most on different taxes. Moreover, the analysis has ignored any problems that might arise if the areas made unequal tax efforts; these are considered in Chapter 5. All that can be said for equalisation grants is that they make horizontal equity attainable for equals in different areas.

It is possible, though, not to make a value judgement in favour of horizontal equity. Grewal, Brennan and Mathews (1980, pp. 176–7) have pointed out that it is seldom argued that a Frenchman and an Englishman on equal real incomes should necessarily face the same residuums, and they suggest that it might be no less proper to ignore horizontal inequity between states in a federal country than between different countries. In their view, the Buchanan equity argument reflects an essentially unitary (that is, a non-federal) standpoint. This view is taken also by Scott (1964, p. 251), who has argued that complete horizontal equity is not a prime goal in a federation. However, it is not clear that federal countries will necessarily be less concerned with horizontal equity than unitary ones. The fact is, as Oates (1972, p. 85) has observed, that if poorer jurisdictions were to have significantly higher tax rates than wealthy areas, and yet have inferior public services, then this would hardly seem fair to citizens living in the poor areas. In practice, therefore, in both unitary and federal systems, value judgements in favour of horizontal equity are often made and may lead to substantial equalisation payments.

The efficiency case for equalisation grants

Horizontal equity may not in fact be the only consideration that encourages the introduction of an equalisation scheme. In the example considered above, Mr α in A was able to secure a higher level of subcentral services than Mr ß in B, even though both had the same level of income and paid the same amount of tax to their subcentral authorities. In these circumstances, there is a clear incentive for Mr ß, and indeed all other citizens of B, to move into A. It is explained in the next section that such migration might be substantially, and conceivably wholly, deterred by capitalisation effects on land values; the efficiency case for equalisation grants arises where it is felt that the capitalisation process will not work so perfectly that no migration is induced. A major argument in this case is that migration is liable to require a substantial use of resources. In the example under consideration, there would be a need for new facilities in A such as houses, hospitals and roads. Of course, some of these facilities may have to be provided by A's authority itself, and so tax rates in A might rise as a result and thereby reduce the amount of any migration and resource usage involved. It nevertheless seems possible that there would be some resource usage, and all resources have alternative uses. Another objection to the migration might be that citizens would be tempted to settle in low tax rate areas even if their wages were lower there; this could mean that they would settle where their contributions to gross domestic product were lower, and hence total output might fall. Thus, unless migration of this sort were to bring clear advantages to a country, an equalisation scheme might perhaps be introduced for efficiency reasons to enable authorities to eliminate this incentive to move. To consider whether the migration would bring clear advantages, it is necessary to look in turn at the three possible reasons why A was able to offer a better deal to its citizens than B.

The first possibility was that A had a higher tax base per head. In the case of a subcentral income tax, this meant that, on average, A's citizens had higher incomes than B's. One explanation, and perhaps the most likely explanation, why this situation might occur between two areas of one country could be that one area (A) has a higher proportion of, say, stockbrokers and coalminers whose equilibrium wages might be relatively high, while the other area (B) has a higher proportion of, say, textile workers and farm hands whose equilibrium wages might be relatively low; in the absence of any clear gains to society from moves of textile workers further away from mills or farm hands further away from farms, an equalisation scheme to neutralise resource differentials and counter the incentive to make such moves has obvious advantages. Another explanation could be that A is endowed with more land, capital and entrepreneurship in relation to labour, and hence has a higher marginal product of labour and higher wages; the efficiency case for an equalisation scheme is a little less clearcut here, and is considered later in this chapter. Of course, subcentral authorities may not levy an income tax, and where other taxes are used it will be necessary to consider afresh whether there would be gains to society arising out of moves from low tax base areas

to high tax base ones. If the tax was on expenditure on goods and services, then it would be hard to envisage substantial gains arising from encouraging people to move nearer to shopping centres. If the tax was on property, in the form of buildings, then it would be hard to see a case for encouraging migration from, say, areas with little business property to areas with much.

The second possibility was that A had a lower unit need index than B, perhaps because it had fewer schoolchildren or criminals in relation to population than B. However, migration from B to A would not alter the total number of children to be taught, or criminals to be caught, and it would thus appear to bring no clear gains to society; once again, therefore, an equalisation scheme to counter the incentive to move would be an attractive proposition.

The third possibility was that A had a lower unit cost index than B. It might, for example, cost more to extinguish fires in B if petrol prices there were relatively high, and it might cost more to educate children in B if they were widely scattered and school buses made long journeys. Now, it could be argued that society would gain if people were to move from areas where the costs of units of output were high to areas where they were low, and thus there might be less support for having an equalisation scheme that neutralised these differentials. Schemes that ignore cost differences may do so for precisely this reason, though they may do so also because such differences may be relatively unimportant compared with differences in resources and needs. It should be noted, however, that it is not always easy to decide whether a particular difference between authorities should be seen as one of costs or needs. For example, a sparsely populated area (B) may have more miles of roads to maintain in relation to population than a densely populated one (A). This could be seen as a difference in needs, where B needs more units of output (miles of maintained road) per head than A, and it could be argued that migration from B to A would bring no gains as the total amount of road maintenance required would not be altered. Alternatively, it could be argued that a unit of output should be defined as a system of two-lane roads between each town in an area, in which case each area might desire one unit, but the cost would be higher in B; there would be a case for encouraging all B's citizens to move into A to reduce the size of the total road network needed. The latter approach might be more appropriate before the network was constructed, and the former more appropriate afterwards.

It is worth concluding this discussion of the efficiency case for equalisation grants by observing that the case will doubtless be felt to be stronger within countries (whether federal or unitary) where migration between areas is typically free from regulations and therefore relatively easy than between countries where migration is typically controlled and therefore relatively hard.

Once a decision is made to have a system of equalisation grants, the structure of the grant scheme needs to be determined. Since the grants are designed to combat a general lack of resources or generally high needs or costs in certain areas, there seems to be a clear case for general grants, although it will be recalled from an earlier section that the problems of

equalisation may be addressed to some degree with specific grants. However, comprehensive attempts at equalisation ordinarily rely on general grants, and where these are used there is the possibility of making them lump-sum grants or effort-related grants. This issue is discussed fully in the next chapter.

Capitalisation as an alternative to equalisation

Attention now turns to three considerations that suggest that equalisation grants may not be essential to secure horizontal equity and so, in turn, efficiency. The first, raised by Barnett and Topham (1980, pp. 244–7),[16] concerns the possible capitalisation effects on house prices that might occur if different subcentral authorities were to face different resources or needs in the absence of equalisation grants. Suppose that a nation forms a single housing market, that all housing is rented, and that when subcentral authorities are created for the first time there are just two (A and B), which provide a single service (X). Suppose, too, that average incomes are higher in A than B, and that the authorities levy an income tax to finance X. In these circumstances, A can offer higher service levels than B at a similar tax rate. Consequently, a citizen in A is liable to get a more attractive fiscal residuum than a citizen with similar tastes and incomes in B; citizens in B are therefore likely to consider migrating to A. Now the sudden rise in demand for housing in A, and the sudden excess supply in B, could lead to changes in land prices and property rents, so that migration ceased to be tempting after all. In other words, taking X and housing together, A might be as expensive as B. In these circumstances, horizontal equity might be held to obtain, and the need for equalisation grants might be felt to be invalid.

Of course, horizontal equity could also hold if an equalisation grant were introduced to B's benefit at the same time; the next chapter outlines one scheme (for the equalisation of fiscal capacity) where horizontal equity would be secured at only one common tax rate, and two other schemes where it would be secured at any common tax rate. With such grants, there would have been no general incentive to move from B to A, and so no general change in land prices or house rents. As it happens, there are at least four reasons for believing that it might be better to secure equity with equalisation grants rather than with capitalisation adjustments.

First, if some of B's citizens were, in fact, owner-occupiers, then they would prefer the equalisation alternative, for with capitalisation the value of their houses would fall, and they would probably not feel that a capital loss on their house was much compensation for the higher tax cost of local services that they would have compared with the cost for citizens in A. Secondly, the presence of equalisation grants would mean that each citizen's investment in land was protected from windfall changes in value arising from migration stimulated by changes in the tax base of his area or the tax bases of other areas. Thirdly, it must be emphasised that land is not used for housing alone; in many areas much of it is used for agriculture. Accordingly, differences in domestic property rents in A and

B may not be sustained; for, in the long run anyway, the supply of land for these properties could rise in A (and perhaps fall in B) thereby reducing the rent differences and so allowing inequity to reappear, and indeed encouraging migration to take place.

Finally, the fear that a reduction in the tax base per head would result in capital losses could make citizens in all areas reluctant to allow any housing development that would attract new residents who might cause such a reduction. In other words, if A and B used an income tax, there would be likely to be widespread opposition to the development of housing that appealed to citizens of below-average incomes, while if A and B levied a property tax or a sales tax, there would be opposition to the development of housing that would have below-average values, or that appealed to citizens liable to have below-average expenditures. In short, the absence of an equalisation scheme raises the very real and ironic possibility (discussed by Hamilton, 1976, p. 752) that poor people would be better off with a poll tax, because this would not result in areas seeking to prevent the development of suitable housing for them. This possibility would be avoided with equalisation grants.

Benefit taxes (type I) as an alternative to equalisation

The second possibility to be considered is whether equalisation grants would be needed if subcentral authorities were to finance their expenditures with taxes where payments by individuals corresponded with their benefits from subcentral services. There are various ways in which an individual's benefits might be measured for (so-called) benefit tax purposes. This section considers the possibility of using the individual's own valuations of subcentral services. The next section considers the possibility of measuring benefits in relation to the costs of providing the services.

It will be recalled that the equity case for equalisation grants arises because, in the absence of such grants, an authority with low resources (or high needs or costs) would be unable to provide services to the same standards as an area with high resources (or low needs or costs) if it set the same tax rate; consequently, citizens with equal incomes who paid the same amount in taxes might find themselves provided with much lower levels of services in one area than another and thus have very different fiscal residuums. This situation could, in turn, induce migration that produced no benefit to society as a whole, and so could create an efficiency case for equalisation grants. Of course, it should be stressed that equalisation grants will not necessarily secure horizontal equity in the sense of securing equal fiscal residuums for individuals with equal incomes. The residuums could still be different if the authorities chose to have different budget levels or different bundles of services; all that equalisation grants do is to make it possible for horizontal equity to be secured.

Now it is clear that the equity and efficiency cases for equalisation grants arise because, without them, individuals would inevitably find some areas fiscally more advantageous than others. Commenting on a paper by Buchanan (1950), Musgrave (1961a, pp. 118–19) once suggested that, if subcentral authority tax payments were to correspond with benefits, then

each citizen would have a zero fiscal residuum wherever he lived, and thus there would be no horizontal equity problem for equalisation grants to tackle. By implication, there should be no incentive for migration, and so no efficiency problem either. Musgrave noted that a system of benefit taxation would arise if, for example, areas were to provide a service to individuals with benefits proportional to their incomes, and if they were to finance it with a proportional income tax. However, as Buchanan (1961, pp. 123–4) subsequently showed, it is important to be specific about whether tax payments would relate to marginal benefits or total benefits. He argued that a system of marginal benefit taxation would not ensure that all citizens were subject to equal fiscal treatment (and by implication would not ensure horizontal equity) and would not remove the incentive to move to areas with greater resources; conversely, a system of total benefit taxation would ensure equal fiscal treatment and remove the incentive to migrate.

The situation can be discussed in more detail with the help of Figures 4.2 and 4.3, which relate to two areas (A and B), each of which provides a single service (X) that brings no benefits to non-residents; the figures relate to a period in time. X may be assumed to be a public good subject to congestion problems where the cost of providing a given level of service is proportional to population. For an area with a given population, the level of service is directly related to the number of units of X provided. A and B each have a population of two, the citizens in A being $\alpha.1$ and $\alpha.2$, and the citizens in B being ß.1 and ß.2. $\alpha.1$ and ß.1 are assumed to have identical tastes and incomes. $\alpha.2$ is the richest of the four and ß.2 is the poorest. The curves α_1 and $ß_1$ show the identical marginal valuations for X of $\alpha.1$ and ß.1; the curves α_{1+2} and $ß_{1+2}$ show respectively the vertical summations of $\alpha.1$'s and $\alpha.2$'s valuations and of ß.1's and ß.2's valuations. It can be deduced from the position of these curves that the rich $\alpha.2$ is assumed to have a higher valuation schedule than the poor ß.2.

Now assume that each area faces the same marginal cost schedule for X, and assume, for simplicity, that these costs are constant and are shown by the curves labelled *MC*. In these circumstances, OQ_A and OQ_B represent the optimal outputs of X in each area (assuming prices of other items equal marginal costs) since those outputs ensure that the sum of the marginal valuations of X in each area equals the marginal cost of X. On the usual interpretation of benefit taxation, citizens are confronted with a tax-price equal to their marginal valuations. So if, for example, the outputs OQ_A and OQ_B were provided, then $\alpha.1$, $\alpha.2$, ß.1 and ß.2 would face tax prices of *OC*, *CD*, *OE* and *EF* respectively, and the authorities would collect $OCC'Q_A$, $CDD'C'$, $OEE'Q_B$ and $EFF'E'$ from them respectively, and they would just meet their total costs of $ODD'Q_A$ and $OFF'Q_B$. It will be noted that the benefit received by each individual from the last unit of X provided precisely equals the tax paid by him for that unit (and indeed each unit). Buchanan (1961, p. 124) has suggested that this condition is necessary but not sufficient for horizontal equity; in so doing, he appears to depart from his original approach (pp. 122–3) of viewing fiscal residuums as the differences between taxes

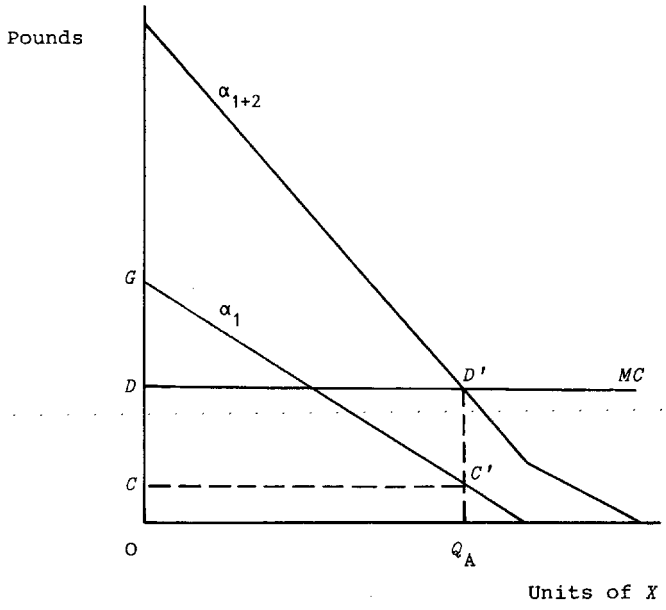

Figure 4.2

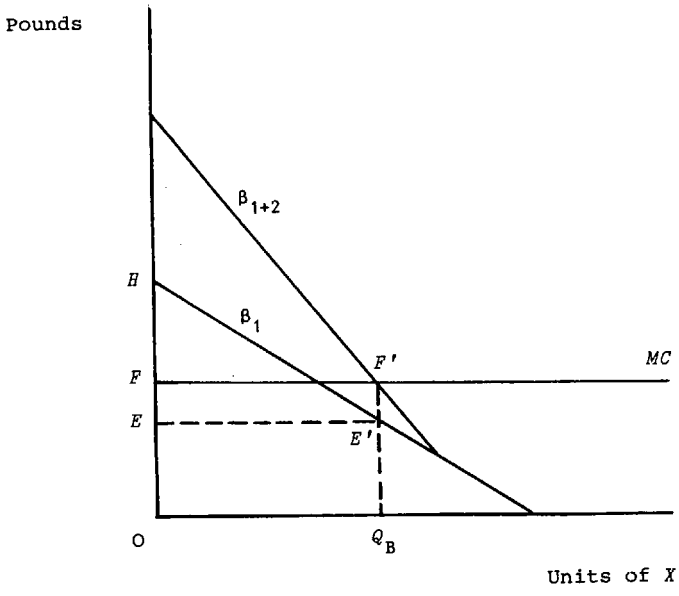

Figure 4.3

and benefits measured in cost values, and seems to view them instead as the differences between taxes and marginal benefits (as valued by taxpayers) multiplied by outputs. Musgrave (1961b) has implied that this condition is sufficient for horizontal equity. In fact, it is very difficult to see how this condition relates to horizontal equity at all. It is quite possible for it to hold even in circumstances like those shown in the figures where ß.1 pays more in tax than $\alpha.1$ ($OEE'Q_B$ being larger than $OCC'Q_A$) even though his authority provides fewer units of X (OQ_B being smaller than OQ_A). Thus it is far from clear that a marginal benefit taxation scheme would produce horizontal equity, and it would certainly not discourage citizens from wishing to migrate since they would generally be better off in areas with higher resources (or lower needs or costs).

Buchanan (1961, p. 123) has also implied, however, that the problems might in principle disappear with a system of benefit taxation in which tax payments by individuals were related to the total benefits that they received rather than the marginal benefits, though he noted that such a system would be nonsensical in practice because governments could use it to exact from each individual almost everything above the bare subsistence minimum. The relevant issues can be related to the example depicted in Figures 4.2 and 4.3. It could be argued that adoption of the marginal benefit taxation system would leave $\alpha.1$ with a much larger fiscal surplus from the provision of OQ_A units of X (CGC') than it would leave ß.1 from the provision of OQ_B units of X (EHE'). This difference in fiscal surplus could be taken both as an indication of horizontal inequity (though only if horizontal equity views fiscal residuums as the difference between taxes paid and total benefits as valued by taxpayers) and as the reason that ß.1, for example, would envy the position of an equal such as $\alpha.1$ and wish to move from B to A.[17] Now if each citizen had to pay a tax equal to his total valuation of benefits received, which would be $OGC'Q_A$ for $\alpha.1$ and $OHE'Q_B$ for ß.1, then (on this approach) each would have a zero fiscal surplus so that there would be horizontal equity and there would be no incentive to migrate. However, the authorities would acquire more money from each citizen than on the marginal benefit taxation approach. Thus they would acquire more than was needed to cover the costs of their services and they might leave citizens with little to spend on other items.

It seems clear that Buchanan is correct in supposing that a system of benefit taxation could hope to avoid the need for a system of equalisation grants only if it related tax payments to total benefits rather than marginal benefits. However, such a system of taxation would probably be inoperable as well as nonsensical. The usual interpretation of benefit taxation, which relates to marginal benefits, is itself ordinarily regarded as inoperable because each citizen would be unwilling to reveal the true extent of his marginal benefits; by pretending that these marginal benefits were small, he could attempt to reduce significantly his tax payments while expecting an insignificant fall in the quantity of public services as a result. This same problem would arise if citizens were asked to reveal their total benefits, though it might be argued that citizens would have an even greater incentive to provide the authorities with underestimates.

It does not appear, therefore, that such a system of benefit taxation is a feasible alternative to a system of equalisation grants. It may be added that the concept of horizontal equity referred to in this discussion is rather curious. On this approach, any system of taxation that exacted equal sums from citizens equally able to pay, and provided them with equal services, would be horizontally inequitable unless their tastes were such that they valued the benefits equally. Although it would be legitimate to form the value judgement that those who most appreciated public services should pay the largest taxes, the impossibility of acquiring honest valuations of benefits makes such an arrangement unworkable. In general, therefore, differences in tastes are ignored for equity purposes, and systems that provide equals with equal services (often loosely described as equal benefits) for equal tax payments are held to be equitable, irrespective of the valuations citizens actually place on those services. It is horizontal equity in this operational sense that equalisation grants make possible.

It may be observed that such equity could be clearly infringed by a system of marginal benefit taxation, since in the example described earlier $\alpha.1$ paid less in taxes than ß.1 for a higher service level. Such equity could also be infringed by a system of total benefit taxation. It was not infringed in the above example where $\alpha.1$ received a higher level of service than ß.1 but paid more in taxes because his total valuation of the benefits was higher. However, he might have paid less for OQ_A units of X than ß.1 would have paid for fewer OQ_B units of X; this could have happened if the curve α_1 had become much flatter to the left of C' and if the curve $ß_1$ had become much steeper to the left of E', for then $\alpha.1$'s total valuation of the benefits from OQ_A units of X might have been lower than ß.1's total valuation of the benefits from OQ_B units, so that $\alpha.1$ would have paid less tax than ß.1.

Benefit taxes (type II) as an alternative to equalisation

In the light of the discussion in the last section, it is worth considering a different approach to benefit taxation from that used by Buchanan and Musgrave. The key difference in this approach lies in the way in which an individual's benefits are valued, for this makes no reference to his own total or marginal valuations. Instead, the aggregate benefit of a public service is related to the total cost of providing it, and this aggregate is divided among beneficiaries in proportion to their consumption levels. For example, in the case of a service provided by a central government and held to be consumed equally by all the nation's citizens, all citizens would be held to derive equal benefits valued at the total service cost divided by the population. On this basis, such a government could be considered to finance the service with a system of benefit taxation if it levied a poll tax at a rate that resulted in a balanced budget, because in this way the citizens regarded as having equal benefits would pay equal amounts in tax, and the total tax receipts would just cover the total costs of the service, which are regarded as an indication of its aggregate benefit. It will be noted that each citizen would have a zero fiscal residuum if this

were defined as the difference between his tax payment and his benefit as measured in this way. Now, if subcentral authorities used similar benefit taxes, then they too could provide zero fiscal residuums, as here defined, for all citizens. Accordingly, it might be thought that there would be no horizontal equity or migration problems and no need for equalisation grants. However, this is not the case.

Suppose, for example, that subcentral authorities were to provide police patrols designed to protect citizens from crime; in this case, the cost per head and the poll tax rate needed to secure a given level of protection would be higher in areas were criminals were thick on the ground. Alternatively, suppose the authorities provided a fire (extinction) service; in this case, the cost per head and the poll tax rate needed to secure a given level of service might be higher in sparsely populated areas. Accordingly, it would be hard to claim that a system of benefit taxes that measured benefits as indicated in the last paragraph would lead to horizontal equity if different areas faced differences in unit needs and costs. Certainly, such a system would not prevent some areas being generally more attractive fiscally and hence would not prevent systematic inducements to migration. This migration might be felt to be beneficial if it were the result of differences in costs, but it has no virtue when caused by differences in needs.

It is clear that it is not satisfactory to measure benefits with reference to the actual costs incurred in each area, for these could vary even when service levels do not. A more satisfactory arrangement is to value the benefits of an authority's services with reference to the costs that would be incurred in providing services to a similar level in an area of standard costs and needs. On this basis, two authorities providing similar standards of services would be regarded, reasonably enough, as providing equal benefits irrespective of whether differences in needs or costs required one to spend more per head than the other. In order for each area to provide zero fiscal residuums, however, it would be necessary for each to levy poll taxes at the same rate – specifically at the rate that would be needed in an area of standard costs and needs. This means that equalisation grants would have to be paid to areas with high needs and costs, while negative grants would be exacted from areas with low needs and costs.

It appears that for such a system of benefit taxation to secure horizontal equity, equalisation grants would be required to offset differences in costs and needs; the sole advantage of benefit taxation over non-benefit taxation seems to be that it would obviate the necessity of equalising differences in resources. This conclusion would hold even if differences in costs were ignored; in that case, a system of benefit taxation appears to require needs differences alone to be equalised, while non-benefit taxation would require differences in needs and resources to be equalised. Although this conclusion suggests that benefit taxes of this sort would not be a complete substitute for equalisation grants, it does not in fact indicate fully the extent to which they would be a substitute, for with such taxes it would not be necessary to equalise all differences in needs. For example, if the subcentral authorities provided schooling and levied a benefit tax on parents (who might be required to pay on behalf of their children), then the cost to a parent would not depend on whether his area had a relatively large or small

proportion of its population attending school, and there would be no need to help parents in areas with large proportions of schoolchildren by means of extra equalisation grants to these areas.

This discussion suggests that the equity and efficiency objectives of equalisation grants can be handled by much more modest schemes if these grants accompany a system of benefit taxes than if they accompany non-benefit taxes. However, it is possible that this discussion understates the case for benefit taxes. It is arguable that some equity and efficiency problems will remain whenever other taxes are used, even if an equalisation scheme were in operation; these problems would certainly not arise if the benefit taxes under consideration here were coupled with such a scheme.

To understand these problems, it is useful to consider an example similar to those used earlier in this chapter. Suppose that Mr α in A and Mr ß in B have equal incomes and circumstances, and suppose that A and B levy income taxes at equal rates. Suppose, too, that they receive equalisation grants that enable them to provide similar service patterns and levels. This seems satisfactory from a horizontal equity standpoint. Moreover, Mr α and Mr ß would have equal fiscal residuums if these were defined as the differences between their taxes paid, which would be equal, and their shares of the costs that would be incurred in an area of standard needs and costs if it sought to provide services at the same levels as A and B. Of course, A and B might choose to provide different bundles of services, in which case Mr α and Mr ß might have different fiscal residuums. As a result, either (or both) might be tempted to move to the area of the other, or they might be tempted to move to areas other than A or B. Such moves would be in reaction to different fiscal packages that reflected geographical variations in preferences, and it was suggested in Chapter 1 that such moves, which might take place in and out of all areas, should be seen as an advantage of subcentral government rather than as a disadvantage. These moves are very different from the moves caused by the inequity that would occur in the absence of equalisation grants, when there would be a systematic incentive for all citizens to move into the areas with the highest resources or the lowest needs; moves of this sort were criticised on efficiency grounds earlier in this chapter.

While there seems to be no problem in this case with non-benefit taxes if A and B set (any) similar tax rate, the situation is different if they do not. Suppose, for instance, that A and B initially set the same tax rate, and that B then doubled its tax rate. This would double its tax receipts and should in turn lead to a rise in its service levels. As a result, Mr ß would pay twice as much tax as Mr α, and receive more benefits, and his fiscal residuum would almost certainly cease to equal Mr α's. The reason for this is that all citizens in B would enjoy the same rise in benefits, but their rise in tax payments would depend on their incomes; thus there would be only one income level for which the rise in benefits equalled the rise in tax payments and hence result in the same fiscal residuum as before. Only if Mr α and Mr ß were on this income level would they continue to have equal residuums. It follows that horizontal equity, in the sense of equal fiscal residuums for Mr α and Mr ß, would usually not occur if A and B set different tas rates. In general, if α and Mr ß had relatively

high incomes, then their tax payments would exceed their benefits, and Mr α's small residuum in low tax rate A would appear more attractive than Mr ß's large residuum in high tax rate B; if they had relatively low incomes, then their benefits would exceed their tax payments, and Mr α's small residuum in low tax rate A would appear less attractive than Mr ß's large residuum in high tax rate B. Overall, there appears to be much horizontal inequity in terms of unequal residuums, and a systematic inducement for the rich to locate in the area with the lowest tax rate and for the poor to locate in the area with the highest tax rate.

By way of contrast, it should be noticed that such problems might not arise if all areas levied benefit taxes where benefits were valued, as suggested earlier, with reference to the costs that would be incurred in an area of standard costs and needs. In this situation, each area would produce zero fiscal residuums as here defined for all citizens; and, of course, a doubling of benefit tax rates in any area coupled with a doubling of service levels would mean that each citizen would pay twice as much in tax for twice the benefit and continue to have a zero residuum.

However, it is not clear how serious the problems with non-benefit taxes are. The point is that citizens do not, in fact, look only at the fiscal residuums as here defined. Rich citizens are likely to place high valuations on public services; accordingly, they will not necessarily feel that their counterparts in the lowest tax rate area get the best treatment, and they will not necessarily wish to move there. Conversely, poor citizens are likely to place low valuations on public services; accordingly, they will not necessarily feel that their counterparts in the highest tax rate area get the best treatment, and they will not necessarily wish to move there. These considerations make it hard to show convincingly that a system of non-benefit taxation, accompanied by a suitable equalisation scheme, would lead to general systematic inequity, even though equals living in areas with different tax rates would have different fiscal residuums. Certainly it would not create systematic incentives to move to extreme areas. Nevertheless, it can be shown that non-benefit taxes cause citizens to locate non-optimally, and this important issue is explored in Chapter 6.

For the moment, it is sufficient to note that a system of benefit taxes where benefits were valued in relation to shares of service costs (or, rather, shares of the service costs that would be incurred in an area of standard costs and needs) would largely but not wholly obviate the necessity for equalisation grants. The difficulties of operating such a scheme should not be underestimated. In particular, there could be considerable controversy about the relative consumption levels, and hence assumed benefits, of different citizens for services such as roads if consumption levels were clearly unequal; also, the incidence of the tax payments might be significantly regressive. On the other hand, at least the system would not require citizens to reveal their own valuations in a context where they had an incentive to under-record them. This is an important difference between this type of benefit taxation and those inoperable sorts that measure benefits in relation either to citizens' own total valuations, hence in principle making equalisation grants quite unnecessary, or to marginal valuations, when extensive equalisation grants would still be needed.

Differential central tax rates as an alternative to equalisation

The third way in which an attempt could be made to secure horizontal equity without equalisation grants would be for the central government to levy one or more of its own taxes at different rates in different areas. The most frequently discussed possibility is that put forward by Buchanan[18] where central government income taxes would be levied at low (or, possibly, negative) average rates in low-resource areas and high average rates in high-resource areas; the rates could be adjusted to take account also of variations in needs and costs. The method can be explained with reference to the example used earlier in this chapter (in relation to Table 4.3) where Mr α in A and Mr ß in B, who each had an income of $30,000 a year, each paid taxes of 5 per cent of their incomes (that is, $1,500) to their subcentral authorities. Now, A was a richer area, and it obtained $200 per head more than B, so that if each area's services benefited its citizens equally, then A's citizens might be held to have fiscal residuums $200 smaller than those of B's citizens. In principle, this horizontal inequity could be removed by altering the central government's income tax bands and rates so that Mr α paid $200 more than Mr ß, and, likewise, all other citizens of A paid $200 more than their equals in B.

In countries where equalisation schemes operate, a move to replace them with differential central tax rates would presumably seek to result in citizens of poor areas paying lower rates of income tax to the central government in return for their authorities losing grant receipts. Buchanan (1950, p. 595) has suggested that one advantage of such a switch in the method of support for poor areas would be that less central tax revenue would be levied, 'an important feature in this era of big central government'; of course, this benefit would be secured only by an increase in the total yield of subcentral taxes, as poor areas would set higher tax rates to recoup the loss of grant income. Perhaps a more important advantage would be the fact that low-resource and high-needs subcentral authorities would have much greater fiscal independence. In principle, then, differential taxation by area could act as a substitute for lump-sum equalisation grants. Oates (1972, p. 84) has suggested that grants are retained in practice because there typically exist constitutional or other political obstacles to geographically discriminatory tax rates; these obstacles might increase if attempts were made to use them as a substitute for a system of effort-related grants, for in this case the central income tax rates in any poor area should be cut each time the area's authority raised its own tax rates.

Equalisation grants and locational efficiency

The final problem to be discussed is whether equalisation grants hinder productive efficiency, or, to put it another way, whether they result in a lower total output than would be produced in their absence. There are two separate reasons why a system of equalisation grants might have adverse effects on output levels. First, by changing the relative advantages and disadvantages of living in different areas, they might influence location decisions by workers and, perhaps, cause them to live in areas where their

marginal products were lower than they would be in the area that they would choose in the absence of grants. Secondly, equalisation grants might discourage recipient authorities from encouraging development in their areas. These two issues are discussed in separate sections.

The national output of private goods and services will be maximised when labour and other inputs are combined in such a way that the marginal product of any particular type of labour is the same in each area of the nation. So long as the marginal product of, say, van drivers is higher in area A than in area B, then output would rise if drivers were to migrate from B to A. In a perfectly competitive economy, the value of the marginal product of any type of labour would equal its wage rate, and, assuming migration entailed no private costs, workers of any sort would migrate until wage rates, and so marginal products, were equal in each area, and hence output would be maximised. This process of migration would be slowed down by any friction such as private migration costs, but in terms of output maximisation there is clearly a case for not adding to any friction that might exist.

Now suppose that the areas A and B are in a disequilibrium position, and that citizens are moving from, say, low-wage B to high-wage A. Next, suppose that each area has a subcentral authority that provides a local service (X) financed by income taxes, and suppose that the total resources needed to produce constant levels of X are not increased as a result of the migration; if they were, this would make the migration less desirable. Since wages are higher in A than in B, the authority in A can offer its citizens a better deal than the authority in B: a given service level can be offered at a lower tax rate, or a higher level can be offered at the same tax rate. In the circumstances, this factor forms part of the case for migrating from B to A. However, if an equalisation scheme were introduced, then (depending on the scheme chosen) this factor would diminish or disappear and the migration process could be slowed down. This argument against grants was put forward by Scott (1950, p. 419), who concluded that the best way of stimulating migration would be to have subcentral provision without equalisation grants.

It should be noticed, though, that this conclusion might be less clearcut if A and B relied on any tax other than an income tax. It would be possible for low-wage B to have a higher tax base than A if a property tax or expenditure tax were in use, and in this case the absence of equalisation grants could encourage migration from A to B, the opposite direction to that required. In such a situation, the introduction of an equalisation grant scheme could reduce or even eliminate any encouragement for citizens to move to areas where they were less productive.

An interesting corollary of Scott's argument is that the incentive to move from B to A would also be reduced if the central government, instead of considering a system of equalisation grants, had itself taken over responsibility for the provision of the previously local service X. This is because the central government would probably seek to supply X at uniform levels in all areas and would probably levy its taxes at uniform rates. Accordingly, citizens in B would no longer fare badly in comparison with the citizens of A as far as X was concerned. Now, if centralisation

of local services tends to discourage desirable migration, it follows, as Scott implicitly concluded, that desirable migration in any nation with central government services would rise if the country were Balkanised; at least, it would do so unless this process turned out to enhance barriers to migration such as psychological factors.

However, it is Scott's objection to equalisation grants that is of primary concern in the present context. Buchanan (1952a) countered it with two arguments. First, he pointed out that there would be no case for abandoning equalisation grants if the income differentials between areas were initially in equilibrium. Suppose that the wage rate for each type of labour is uniform over the country as a whole, but that the proportions of citizens employed in different types of occupation vary from area to area, so that average incomes are higher in A than B. In this case, an equalisation scheme would be necessary to discourage migration from B to A since this would reduce total output.

Secondly, he claimed that, even if the initial situation was one of disequilibrium, the maintenance of grants might not hinder output growth. He agreed that desirable migration of labour would increase if the equalisation grants were abolished; but he argued that wage-rate equality between areas, and the associated high output levels, would be promoted by other forces too, and suggested that these forces might be weakened with the abolition of grants. For example, he claimed that B's lower wage rate might be the result of high ratios of unskilled labour to other factors such as capital and entrepreneurship. The migration of unskilled labour from B to A would tend to reduce wage differences, but so would the migration of the other factors from A to B, and, in the disequilibrium situation, this migration should take place since these factors would be relatively scarce in B and hence have higher rates of return. If equalisation grants were abolished, however, then B might, for example, raise property taxes and cut spending on roads, and such actions might reduce the immigration of capital and entrepreneurship. In cannot be said, *a priori*, whether the abolition of the scheme would overall hinder or hasten the equilibrating process. A net hastening effect would be more likely if B's low wages stemmed in large part from poor natural resource endowments, as suggested by Scott (1952); in this case, increases in output would be greatly helped if scarce labour was given an extra incentive to move to A, and if B did become less tempting to capital and entrepreneurship, because A can be assumed to have the best soil, ports, mines, sites, and so on.[19]

Of course, this discussion of Buchanan's second point should not cause his first point to be overlooked: there is no case for abolishing grants and stimulating migration where the initial situation is in equilibrium. The same might apply even if it were only near to equilibrium, for then the abolition of grants might stimulate much more movement from B to A than was required.

More recently, Buchanan has somewhat altered his stance. With Wagner, he has developed a model to show that labour would not migrate in a way that led to output maximisation either in the absence of equalisation grants or if full equalisation were attempted (Buchanan and

Wagner, 1970). Indeed, it is claimed that a de-equalising scheme might be required, though a scheme of partial equalisation is considered the sort most likely to be needed. The model used to demonstrate these results values public products as the total fiscal surplus of all beneficiaries – that is, the sum of each individual's total valuation of the benefits less his taxes paid; further, it assumes that each subcentral authority operates a fixed-size public facility. As an area's population rises, so its citizens gain, since per capita tax payments fall, but they may also lose since it is assumed that (ultimately, at least) congestion will set in and therefore the benefits from the utility to existing citizens will fall. Buchanan and Wagner have argued that an individual should migrate from one area to another when any resulting increase (or fall) in the output of private goods and services more than offsets (or is more than offset by) any net loss (or gain) to the citizens of the two areas concerned – including the citizen whose move is under consideration – arising from changes in tax payments and benefits from the public service. Ideally, therefore, individuals should consider all the gains and losses accruing to society when considering a move; as, generally, they will consider only the private gains and losses, so, generally, the actual amount of migration is unlikely to be the optimal amount. A system of grants is needed that makes underpopulated areas more attractive, and overpopulated ones less attractive.

This model rests critically on the assumption that authorities operate fixed-size public utilities subject to congestion. Few subcentral services are of this nature. With many services, it may not be unrealistic to suppose that a constant level of per capita expenditure could secure a constant level of service irrespective of population size; certainly, authorities of vastly different size do often seem to provide similar benefits from similar per capita expenditures, and it was noted in Chapter 1 that evidence of economies of scale in practice is, to say the least, very limited. If this assumption is widely applicable, then the only relevant effect on output from migration would be any change there might be in the production of private goods and services. Most of the discussion in this section rests on this assumption.

It seems, then, that in the absence of economies of scale in the provision of subcentral services, the ideal distribution of population would be secured if labour migration occurred only in response to wage differentials between areas for the same job; labour supply should always fall in low-wage areas, thus raising marginal products and wages, while the converse should occur in high-wage areas. It would be necessary to discourage potential migrants from being affected by other factors. In particular, it would be necessary to discourage moves into any area where the only attraction was that the per capita tax base was high, thereby resulting in lower subcentral tax rates or higher service levels than elsewhere. This discouragement could be effected by a full equalisation scheme, though, for reasons noted by Scott, such a scheme might prevent undesirable migration at the cost of reducing the rate at which desirable migration took place. Further work is needed to see how far, in practice, subcentral services are provided by fixed-size facilities for partial equalisation to be appropriate for the reasons explained by Buchanan and Wagner.

Equalisation grants and development

This chapter is concluded by considering a quite different way in which a system of equalisation grants might result in output levels being lower than they would be without the grants. The argument is that subcentral authorities might well encourage development in their areas if there were no equalisation grants, and yet be indifferent or hostile to development if there were some. Consider, first, an authority in a situation where no equalisation grants existed. In this case, development in its area might well increase its tax base in relation to population and thereby permit existing service levels to be provided at lower tax rates, or higher service levels to be provided at existing tax rates. The increase in the tax base might take place directly if the authority were to levy a property tax, and indirectly if it were to levy an income tax (since, for example, development might raise incomes by lowering unemployment) or taxes on expenditure (since expenditure would rise as a result of any rise in incomes). Thus development might well appear attractive.

Next, suppose that there was an equalisation scheme that resulted in the relationship between an area's tax effort and its per capita tax and grant income being independent of its tax base.[20] In this case, the authority would have no fiscal incentive to encourage or permit (as the case might be) any form of development that would increase its per capita tax base, for any such increase would be offset by a corresponding fall in grant receipts. Such indifference to development might lead to less development than would occur without grants.

The situation would be rather different if the equalisation scheme were to result in the relationship between an area's tax effort and its per capita tax and grant income being dependent on its tax base.[21] This situation occurs with equalisation schemes that involve lump-sum general grants; as explained in the next chapter, such schemes are used in, for example, Australia and Canada. In this case, an authority might find that, in the absence of an equalisation grant, the relationship between its tax rate and its income per head would be that shown by the line OT in Figure 4.4 (which relates to a period in time). Now, lump-sum grant schemes often seek to ensure that grantees could achieve a specified total tax and grant income per head, say Oy, if they set a specified tax rate, say Ot. If the area concerned in Figure 4.4 received such a grant, then the relation between its tax rate and per capita income (from tax and grant) would be shown by the line FG. FG is parallel to OT and represents the effect of a lump-sum grant with a per capita value of OF; this is sufficient to give the authority a total per capita income of Oy if it sets the tax rate Ot.

Now, if development were to take place, then the line OT might be replaced by OT' – showing, in effect, that a higher tax revenue would be produced than before at any specific tax rate. With a higher base, the authority needs a lower lump-sum grant to enable it to have the income Oy if it sets the tax rate Ot. In fact, it needs a grant of OF' per head, and, with a grant of this amount, the relationship between its tax rate and per capita income would be that shown in Figure 4.4 by the line $F'G'$. With such a grant scheme, the authority's attitude to development would

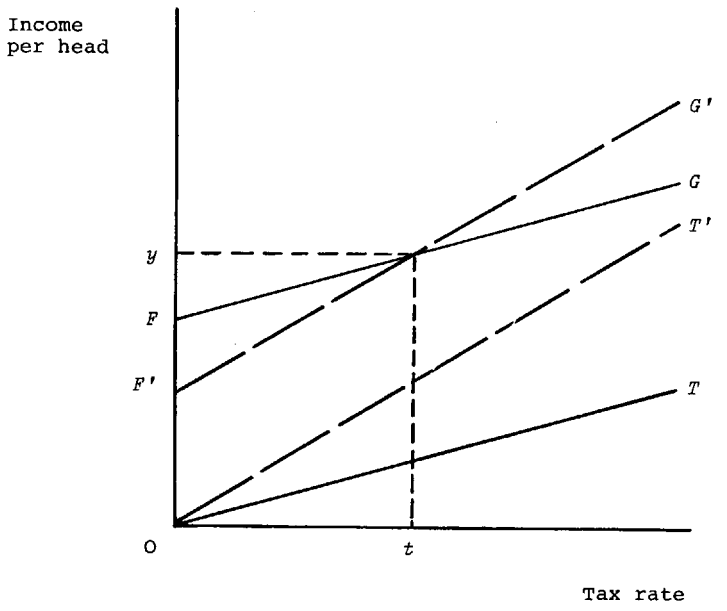

Figure 4.4

depend on the per capita income level it sought. If it were to seek a level below O*y*, then it would be hostile to development, for any such level could be obtained at a lower tax rate without it; *FG* is to the left of *F'G'* at such income levels. Conversely, if it were to seek a level in excess of O*y*, then it would encourage development, for any such level could be obtained at a lower tax rate with it; *F'G'* is to the left of *FG* at such income levels.

The possibility that equalisation grants might encourage indifference, or even hostility, to development is clearly of some concern to those anxious for development, but it might not be unwelcome to all. In Canada, for example, where the often modest equalisation arrangements for local authorities do not seem to have these effects, Hayes (1977, p. 171) has noted that 'it is often argued that decisions by local authorities about municipal planning and development are unduly influenced by a desire to improve the real property tax base'.

It has been suggested by some of those who presumably are anxious for development that the possibility that equalisation schemes would make authorities indifferent or hostile to development could be avoided by arranging that estimates of their tax bases should include, say, only 90 per cent of any increases due to development (Foster *et al.*, 1980, pp. 549–50). In time, though, this might lead to large discrepancies between an area's estimated tax base and its actual tax base. An alternative might be to add the increases in full, but to add them only after a specified period

of, say, three years had elapsed. However, any device that created a difference between estimated and actual tax bases would mean that an equalisation scheme would not fully secure the equality it was intended to bring about.

It may be noted that there is a quite different reason why subcentral authorities might be hostile to development. This is that they might find that development resulted in a need for items such as extra roads or sewers, which they would have to provide. There would be no need for such hostility, however, if the authorities were able to recoup the costs incurred by levying appropriate charges on the developers. Where such charges are not levied, the expenditure that the subcentral authorities would have to meet from their own resources might more than offset any advantages of development that a higher tax base would bring in the absence of equalisation grants. If an equalisation scheme were in operation, then the need for expenditure on items such as roads and services could be incorporated into the authorities' needs indices and this cause of hostility to development could be overcome. Indeed, this possibility might mean that the introduction of an equalisation scheme could improve the climate for development by replacing a state of hostility with one of indifference.

Summary

It is contended in this chapter that the reasons why grants may be given fall into four groups. First, there are grants for use in situations where authorities provide some services that benefit non-residents. The problem in these situations is that authorities are likely to ignore external benefits when determining their output levels, and hence they are unlikely to produce optimal levels. In principle, a system of matching specific grants could be introduced for each service concerned to encourage provision at optimal levels, although it would be very difficult to estimate with precision the grants needed. An interesting conclusion is that the introduction of such a system could result in a higher aggregate consumption of the service concerned and yet a lower aggregate output by encouraging provision to be concentrated in those areas that generate the highest spill-outs.

Secondly, there are grants intended to enforce grantor preferences. These may be introduced for certain services regarded as merit goods by the grantor if it feels grantees will produce too little if allowed to choose service levels themselves. For example, it might feel grantee voters were unaware of the benefits of a new service or ignored the benefits of certain services to non-voters or minority groups; or it might wish to tie revenue-sharing or equalisation grant payments to specific services if it feels its own taxpayers would prefer this arrangement to general grants. Alternatively, the grantor may wish to enforce minimum standards for some services, perhaps to protect citizens unable to migrate from areas inclined to low service levels, perhaps to reduce the need for complementary services for which it is responsible, perhaps to ease the minds of citizens who have to migrate and want a choice of areas with certain basic service levels,

and perhaps to ensure that subcentral levels of merit good provision do not vary so much that migration problems arise. Finally, central government grantors may use grants to affect the level and, perhaps, the composition of subcentral spending for macroeconomic reasons.

Thirdly, grants may be used for revenue-sharing purposes to help a tier of authorities that cannot meet the expenditure deemed desirable by the grantor at tax rates and charge levels it regards as acceptable. An alternative to such grants would be to increase the tax powers of the tier concerned, provided satisfactory taxes could be found. The tier concerned is likely to have more autonomy with tax receipts than with revenue-sharing grants.

Finally, grants may be paid for equalisation purposes, to neutralise authorities in part or in whole against differences in their tax resources and variations in their costs of and needs for services. Without equalisation grants, some areas will be able to offer more attractive fiscal packages than others, thereby causing horizontal inequity between equals in different areas and stimulating migration that could reduce output and would consume resources.

It has been suggested that it is not necessary to have equalisation grants to solve these problems, but the three alternatives examined do not seem very promising. For example, relying on capitalisation effects on land prices means that owner-occupiers of houses might get windfall gains or losses whenever a factory opened or closed in their area, and they might oppose housing developments that would attract poor people and lower their area's per capita tax base. The most promising system of benefit taxes is one where tax payments reflect benefits measured in relation to individual consumption levels rather than individual valuations; however, such a system would be hard to operate for services such as roads, it would probably produce a steeply regressive incidence of taxation, and it would not wholly remove the need for equalisation grants. Differential central government tax rates between areas could be a more straightforward alternative, but might be unacceptable for constitutional or political reasons.

The effects of equalisation grants on location decisions by citizens depend very much on the circumstances. It is certainly possible that the grants would reduce the rate at which people moved to areas where they could obtain higher real wages and be more productive, but they might also increase the rate at which capital moved into areas most in need of development. Equalisation grants could also make authorities indifferent, and in some circumstances hostile, to development instead of enthusiastic; some people might see this as an argument against the grants, while others might not.

Notes

1 The analysis here is derived from Oates (1972, pp. 99–104).

2 $MRT_{xy} = MC_y/MC_x$ by definition. Suppose $MC_x = 1$, then $MRT_{xy} = MC_y$, and a grant meeting half the cost of Y (so that $g = 0.5$) would be worth $MC_y/2$ or $MRT_{xy}/2$ per unit, that is, $MRT_{xy}.g$.

3 Equation (6) is found by multiplying equation (3) by ß and subtracting it from equation

(2). Equation (7) is found by multiplying equation (2) by α and subtracting it from equation (3).

4 For example, equation (4) shows that $G_A = MRT_{xy} - \sum_{i=1}^{m} MRS_{xy(A)}$. Using the value of $\sum_{i=1}^{m} MRS_{xy(A)}$ implied in equation (6), it follows that $G_A = MRT_{xy} - MRT_{xy}.(1-ß)/(1-\alpha ß)$. Thus $G_A = MRT_{xy}.[(1-\alpha ß)-(1-ß)]/(1-\alpha ß) = MRT_{xy}.ß(1-\alpha)/(1-\alpha ß)$ as in equation (8).

5 This analysis closely follows Boadway (1979, pp. 423–4).

6 For a fuller discussion of the problem of determining an optimal set of subsidies when there are reciprocal externalities, see Davis and Winston (1962).

7 The possibility that output might be higher without grants was first discussed by Williams (1966), although its dependence on there being a redistribution of real income was first demonstrated by Brainard and Trenery Dolbear (1967); both these papers assumed, in effect, that any units of *Y* produced by A and consumed in B could not simultaneously be consumed in A. The possibility that output might be higher without grants even where, in effect, units of *Y* produced in A could be consumed simultaneously in A and B was shown by Pauly (1970, pp. 574–7); Pauly's model was essentially a special case of that considered in the text with $\alpha = ß = 1$, but his conclusion would apply with any positive values for α and ß.

8 It is possible that there might eventually be less total consumption as well as less total production if all parties were to move from independent action to collective action. See Buchanan and Kafoglis (1963).

9 Their argument at this point discusses merit goods in relation to individuals rather than subcentral authorities.

10 Their argument at this point discusses merit goods in relation to individuals.

11 The figures for the United States are derived from Break (1980, pp. 126 and 180) and the United States Department of Commerce (1981, p. 284).

12 The figures for Canada are derived from the Ministry of Supply and Services (1981, pp. 829–31).

13 The figures for Australia are derived from Mathews (1980b, pp. 178 and 185).

14 The figures for the United Kingdom are derived from HMSO (1982b, pp. 55–6).

15 See Mathews (1980a, pp. 10–11) for a discussion of this point.

16 See also the subsequent discussion of King (1982) and Barnett and Topham (1982).

17 Musgrave (1961b, p. 133) was not initially convinced that the migration induced by different surpluses should be seen as distorting, but later (see Musgrave and Musgrave, 1976, p. 629n) he seemed to accept that it was.

18 See Buchanan (1950, pp. 595–6) and Oates (1972, pp. 79–81).

19 In a reply, Buchanan (1952b) noted that land quality differences might be reduced with the advent of scientific rain-making and the possible peacetime utilisation of atomic energy.

20 Figures 5.1, 5.2 and 5.3 and Figures 5.7, 5.8 and 5.9 in Chapter 5 illustrate two such schemes.

21 Figures 5.4, 5.5 and 5.6 in Chapter 5 illustrate such a scheme; for simplicity, Figure 4.4 shows per capita (tax and equalisation grant) income on the vertical axis rather than an index of service levels. For a given area, the two should be directly related.

5

Equalisation and Revenue-Sharing Grant Schemes

Introduction

It was explained in the last chapter that two of the purposes for which grants are frequently given – namely, equalisation and revenue-sharing – might ordinarily be felt to warrant schemes of general grants. As it happens, it is possible to devise general grant schemes that can handle both of these objectives simultaneously. This chapter considers the issues that arise when such schemes are designed.[1]

One particularly important issue is whether the grants paid to individual grantees should be effort-related grants or lump-sum grants. The next two sections consider two widely discussed schemes: in one, grantees' receipts are effort-related in such a way that they vary proportionately with grantees' tax efforts, while, in the other, grantees are allocated lump-sum grants. The relative advantages and disadvantages of these two schemes are noted in a further section. This is followed by an outline of another scheme in which grantees' receipts are effort-related but do not vary proportionately with tax effort; this scheme seeks, as far as is possible, to embody the advantages of both the earlier schemes.

Since all these schemes are generally concerned with neutralising differences in expenditure requirements and resources, it is important to consider how those two items might be measured. Two further issues arise in schemes that take account of grantees' revenue efforts – namely, how revenue effort might be measured and whether the grantor can predetermine its total grant outlay before grantee tax rates are known. A separate section is devoted to each of these four problems.

After this theoretical material, attention turns to the equalisation and revenue-sharing arrangements in Australia, Canada, the United Kingdom, the United States and West Germany. There are separate sections for each country, but these are preceded by a section concerned to show the relative necessity in each of those countries for equalisation and revenue-sharing grants.

A scheme to equalise fiscal potential

To help compare different possible grant schemes, it will be supposed that there are two areas (A and B), which levy a proportional income tax, and where the unit need indices (N), the unit cost indices (C) and the per

Table 5.1

		Symbol	*Area A*	*Area B*
1	Unit costs (index)	C	1.00	1.09
2	Unit needs (index)	N	1.00	1.22
3	Income per head ($'000s)	T	20.00	16.00
4	Tax rate (proportion)	t	0.05	0.05
5	Tax yield per head ($'000s)	tT	1.00	0.80
6	Service levels (index)	tT/CN	1.00	0.60

capita tax bases (T) are as shown in the first three lines of Table 5.1. These items were all explained in Chapter 4.

Suppose that the grantor wishes to introduce a general equalisation grant scheme intended to neutralise fully any differences there were between costs, needs and resources in different areas. Such an aim suggests that areas whose indices of costs or needs were above some standard figures, or whose resources were below one, would expect to receive positive grants. Conversely, areas whose costs or needs were below some standard level, or whose resources were above one, would expect negative grants, that is to say they would expect to make payments to the grantor. If the standards used were average figures, then negative grants might be as common as positive ones, a situation that must be expected in any grant scheme devoted solely towards equalisation.

In practice, many schemes avoid negative grants by setting standards that are not averages and that result in most, if not all, areas receiving positive grants. The few areas that might expect to receive negative grants may be ignored, so that no negative grants arise, though at a cost of having a few areas with citizens not treated equitably with those elsewhere; such an arrangement has long been a feature of equalisation grants in the United Kingdom. Broadly speaking, schemes that ensure that most areas are recipients are intended to accomplish a revenue-sharing function as well as an equalisation function. With composite schemes of this sort, the standards used may be arbitrary figures, or they may be the figures applying to one particular relatively well-off area. For the sake of simplicity, suppose in the present example that the standard figures for C, N and T used by the scheme are those applying to area A, which may be termed a standard area and which thus appears to merit no grant.

The question arises of how much grant area B should receive. The authority in B might argue that if it were to set a tax rate (t) of say 5 per cent, then (as shown in lines 5 and 6 of table 5.1) it would achieve a service level index of only 0.6, whereas the same rate in the standard area A would result in a service level index of 1.0. Accordingly, B could say it needed a grant to boost its income and expenditure (of $800 per head) by two-thirds so that it too could secure a service level index of 1.0 with a tax rate of 5 per cent. This means that it would want a per capita grant (G) of $533. When grants are paid, the service level index must be redefined as $(tT+G)/CN$, where G is the grant receipt; for area B, with a t of 0.05,

this would equal unity (as in A) if G were to equal \$533, so that total income would be this plus the \$800 from taxes, or \$1,333 per head.

It is important, of course, to be able to estimate B's grant if it set another tax rate. Suppose it set a rate of 10 per cent. Its authority could point out that a 10 per cent rate in the standard area, A, would raise twice as much income as a 5 per cent rate and thus result in a service level index of 2.0; and it could argue that it should receive sufficient grant to enable it to secure an index of 2.0 when it set a rate of 10 per cent. Such a rate would yield \$1,600 per head in taxes, and, if coupled with a grant of \$1,066, would give a total income of \$2,666; this is double the income needed for a service level index of 1.0 and results in a service level index of 2.0 as required. More generally, B could claim that at any tax rate t, the standard area A could acquire an income of $t\bar{T}$, where $\bar{T}$ is the standard tax base applicable to area A. For B to provide the same service level index at t, B would need a total income of $t\bar{T}(CN/\bar{C}\bar{N})$ where C and N are B's indices of costs and needs, and $\bar{C}$ and $\bar{N}$ are the standard indices applicable to A; since $\bar{C}$ and $\bar{N}$ are each defined as unity here, the income needed by B simplifies to $t\bar{T}CN$. Thus B needs a grant that, combined with its tax income of tT, gives it this total income. Accordingly, an area's grant is determined by $G = t(\bar{T}CN - T)$. This is proportional to its tax rate t and is clearly effort-related.

The example in the last paragraph assumes that unit cost differences will be allowed for. For reasons noted in Chapter 4, this is not always the case. Cost differences can readily be ignored by assuming that each area has the standard unit cost index of 1.0. Then, for grant purposes, service levels will be defined by $(tT+G)/N$, and actual differences resulting from different Cs will be overlooked. In turn, the formula for G simplifies to $G = t(\bar{T}N - T)$. In practice, some schemes also ignore differences in unit needs. This can be done by assuming that each area has the standard unit need index of 1.0; then, for grant purposes, service levels are defined as $tT+G$ and the grant formula becomes $G = t(\bar{T} - T)$.

The effects of this scheme for A and B can be seen in Figures 5.1 and 5.2. Without the grant scheme, the service level index for any tax rate t will be given by the lines OT_A and OT_B for A and B respectively. Thus, with rates of 5 per cent they could achieve service level indices of 1.0 and 0.6 respectively, as shown in Table 5.1. If the two areas were to set equal tax rates (other than zero), then A would always be able to provide higher service levels. With the scheme, the areas' service level indices for any tax rate would be given by the lines OG_A (which coincides with OT_A) and OG_B respectively; these are identical, so if A and B were to set any common tax rate then they could secure similar service level indices. It will be noticed that the gap between OT_B and OG_B widens as t rises, indicating that B's grant receipts would increase if it raised its tax rate.

Figure 5.3 shows the position for a third area C, where a favourable combination of costs, needs and resources produces the pre-grant relationship between its tax rate and service level index shown by OT_C. At a rate of 5 per cent, the service level index is 1.2, thereby exceeding the level of 1.0 in the standard area A. This area qualifies for negative grants, so that it has to pay sums to the grantor; these leave it finally on

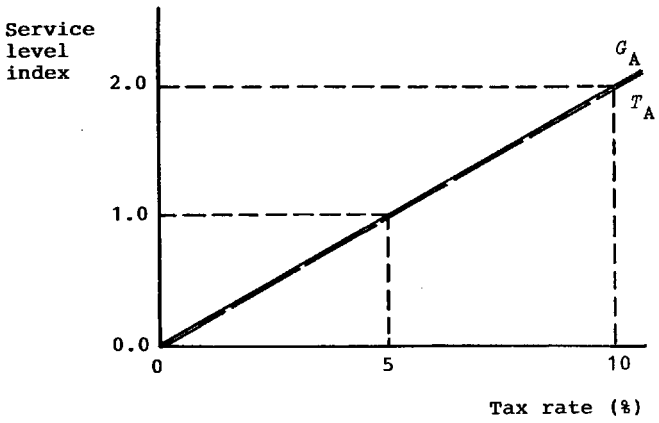

Figure 5.1

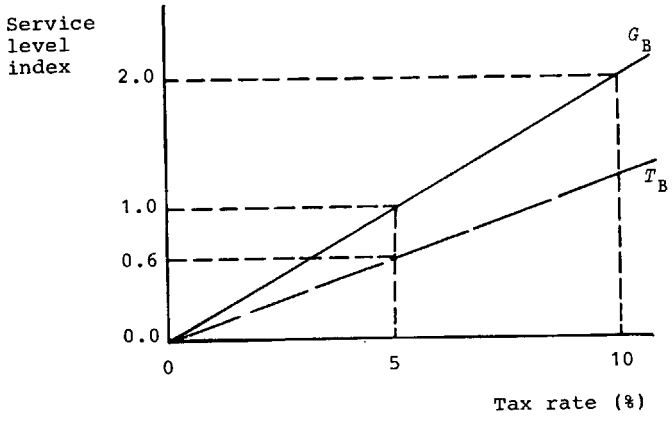

Figure 5.2

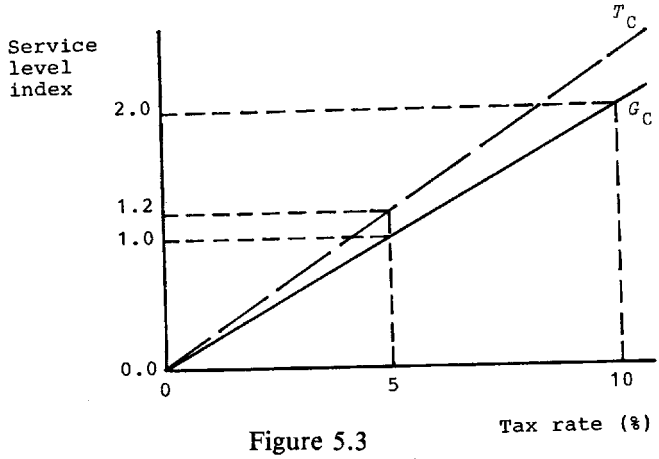

Figure 5.3

OG_C, which is identical to OG_A and OG_B, and ensure that C, like A and B, would have a service level index of 1.0 if its tax rate were 5 per cent and of 2.0 if the rate were 10 per cent. The gap between OG_C and OT_C widens as t rises, indicating that C's payments to the grantor would rise if it increased its tax rate. This situation might be common in a scheme concerned with equalisation alone where the final standard service level index schedule (depicted by OG_A, OG_B and OG_C) would be defined by the tax schedule (OT) of an average authority. In a scheme concerned also with revenue-sharing, where the final standard service level index schedule might lie above most (and perhaps all) authorities' OT lines, this situation could be rare. In practice, these few authorities might be left out of the scheme to avoid the need for unpopular negative grants, though there would then be inequity between the citizens in each authority omitted from the scheme, and between the citizens in those authorities and those in the areas covered by the scheme.

The scheme just described is similar to one discussed in a seminal paper by Musgrave,[2] and was termed by him a scheme for equalising fiscal potential, though it is also sometimes known as one for equalising district power. It is worth noting that there are elements of fiscal potential that the scheme ignores. For instance, differences in development between areas stemming from different policies on mineral exploitation might affect per capita incomes (and perhaps other tax bases) and be compensated for by differences in grants, but these would not accurately reflect differences in fiscal potential.[3]

A scheme to equalise fiscal capacity

Musgrave contrasted this scheme with another, often encountered in practice, in which grants are not effort-related. In this scheme, the grantor defines a standard tax rate ($\bar{t}$) and then works out the grant each area would receive if it set that rate in precisely the same way that would be used with the previous scheme. Thus, if $\bar{t}$ were set at 5 per cent, then, as before, A would receive no grant if it set that rate, while B would receive \$533, and, as before, each area could achieve a service level index of 1.0. The difference with this scheme is that the areas would receive these amounts of grant whatever tax rates they actually set. Thus the grant (G) is fixed as on the previous scheme with the assumption that the tax rate chosen is $\bar{t}$, and so is defined by $G = \bar{t}(\bar{T}CN - T)$. The right-hand side of this equation would become $\bar{t}(\bar{T}N - T)$ if unit cost differences were ignored, and $\bar{t}(\bar{T} - T)$ if unit need differences were also ignored.

To see the effects of this scheme, suppose that A and B set tax rates double $\bar{t}$, that is 10 per cent. A, being the standard area in receipt of no grant, would be able to achieve a service level index of 2.0, as before, but B would not. B would raise \$1,600 in taxes and still receive \$533 in grant. Thus $tT + G$ would equal \$2,133, and its service level index, given by $(tT + G)/CN$ would be only 1.6. The effect of this scheme for areas A and B is shown in Figures 5.4 and 5.5. As before, OT_A and OT_B show the service level indices attainable for each tax rate in the absence of grants.

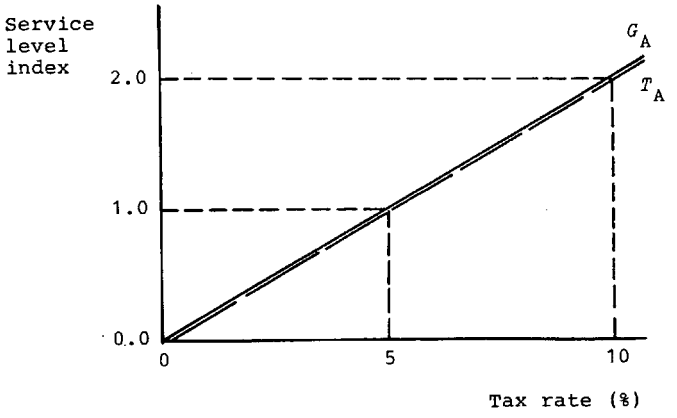

Figure 5.4

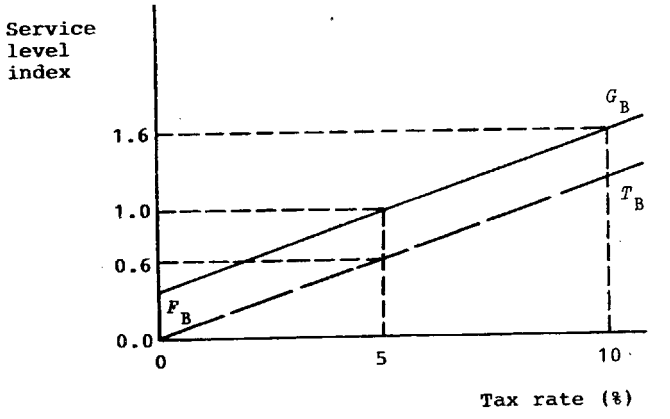

Figure 5.5

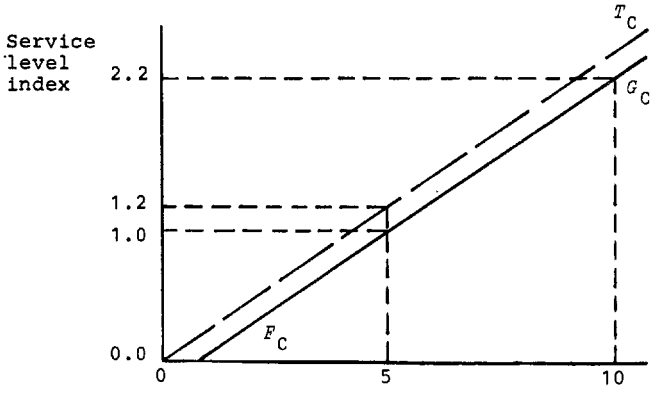

Figure 5.6

The new grants put the areas on OG_A and F_BG_B respectively. Figure 5.6 shows the situation for area C. Unless it were omitted from the scheme, it would be obliged to pay a lump sum to the grantor that would take it from OT_C to F_CG_C; in this case, it too would have a service level index of 1.0 if it set a rate of 5 per cent, but it would have an index of 2.2 if it set a rate of 10 per cent. Musgrave described this scheme as one that equalised fiscal capacity.[4]

Equalisation of fiscal potential and equalisation of fiscal capacity compared

In terms of horizontal equity, there is little doubt that the first scheme is preferable to the second. With the second scheme, for example, it is possible for one area, B, to set a tax rate of 10 per cent and secure a service level index of only 1.60, while another area, A, could set a rate of just 8.5 per cent and achieve an index of 1.70, and yet another area, C, could set a rate of only 8 per cent and achieve an index of 1.72. As noted in Chapter 4, Oates (1972, p. 85) has argued that this situation would not seem fair to citizens in B; it is thus a little surprising to find that when he discussed the possibility of grant formulae taking tax rates into account (1977, p. 315), he said 'I see no legitimate role for such provisions'. Another feature of the second scheme that might be felt inequitable is that equal absolute or percentage changes in tax rates benefit A more than B but less than C. For instance, a rise in t from 5 per cent to 6 per cent raises the service level index in B from 1.00 to 1.12, while it raises the index in A from 1.00 to 1.20 and the index in C from 1.00 to 1.24.

Musgrave (1961a, p. 103) has argued that the second scheme has another disadvantage resulting from the fact that it makes no allowance for the grantee's own tax effort. With this scheme, the grantor might be levying taxes on citizens throughout the country to contribute to public services in areas such as B even though the residents of B might fail to make even a modest effort to meet their own problems. To this, citizens elsewhere 'will rightly object'.

However, the first scheme has one feature that has often been criticised, for example by Grewal (1981b, pp. 40–1), and this is that it encourages non-optimal service levels. Authorities such as B, which are in receipt of grants, will find the grants meeting fixed proportions of their expenditures.[5] In effect, the grants reduce the perceived marginal cost of the authority's services, and citizens will be encouraged to consume them beyond the optimal point where the sum of the marginal rates of substitution equals the marginal rate of transformation. By contrast, authorities such as C, which make payments to the grantor, will find that their payments amount to a fixed proportion of their expenditure. In effect, the grant increases the perceived marginal cost of the authority's services, and citizens are encouraged to consume sub-optimal amounts. In practical terms, the grantor may feel that such a scheme would encourage areas such as B to set tax rates that were too high, and acquire large grants in the process, and encourage areas such as C (if there are any) to set tax

rates that were too low. These incentives for non-optimal service levels would be avoided with the second scheme, for its lump-sum grants do not alter the perceived marginal cost of grantee services.

It is not altogether clear, however, how much importance should be attached to this criticism. Le Grand (1975, p. 535) has argued that giving authorities in receipt of grants an incentive to increase public expenditure is 'a feature of the scheme which may not appeal to the budgetary agents of the central government'. However, if the central government were the grantor, and if it felt the scheme encouraged too much subcentral spending, then it could induce a reduction by making the scheme less generous, for instance by raising $\bar{C}$ or $\bar{N}$ or by cutting $\bar{T}$. Moreover, it is arguable that most grants schemes in practice have a revenue-sharing function as well as an equalisation function. Typically, this means that the central government wishes subcentral authorities to have higher service levels than they would without the grant scheme; in such circumstances its budgetary agents might well prefer effort-related grants since they secure a given increase in grantee spending with a lower amount of grant than would be needed with lump-sum grants, and hence make smaller financial demands on the central government. Their main grievance in this situation would probably be the high grants paid to the areas with the highest subcentral spending. Ultimately, of course, the chief attraction of the second scheme is that each authority is induced to provide an optimal quantity of services with the funds available, but this efficiency characteristic is obtained at the cost of inequity between citizens in any pair of authorities with the same tax rate (unless it is the standard rate).

An alternative equalisation scheme

The choice appears, therefore, to depend on a value judgement between the two schemes. The first makes horizontal equity possible and gives little help (but some incentive) to areas that are unwilling to help themselves, but results in some allocative inefficiency between the private and (subcentral) public sectors and in high grants to extravagant areas in receipt of grants; the second scheme produces converse results. The choice of schemes may, however, be wider. Of course, if no allocative inefficiency were to be permitted, then a scheme such as the second, with its lump-sum grants, would have to be chosen with its attendant strengths and weaknesses; but work has been done recently to develop schemes that retain the attractions of the first scheme while trying to reduce its weaknesses.

One such scheme, developed by King,[6] is illustrated in Figures 5.7, 5.8 and 5.9. As before, the relationships between tax rates and service level indices in areas A, B and C in the absence of grants are shown by the lines OT_A, OT_B and OT_C. With grants, the relationships are shown by the curves, F_AG_A, F_BG_B and F_CG_C; these are identical, but their precise curvature can be determined by the grantor. As F_AG_A, F_BG_B and F_CG_C are all the same, horizontal equity between A, B and C will be possible at any common tax rate, and, as a corollary, any subsequent equal changes in the rate would secure equal changes in service levels. Moreover, the

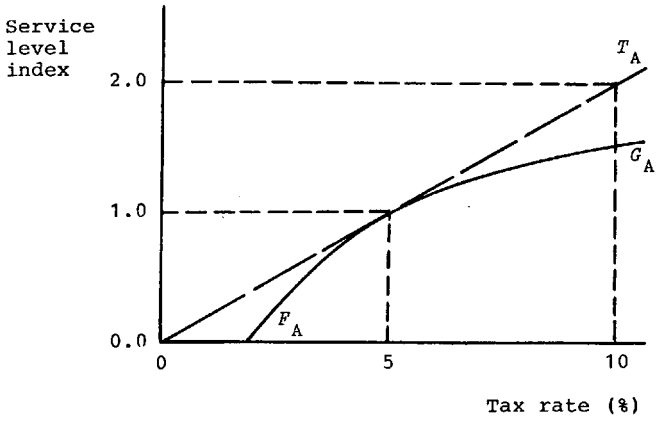

Figure 5.7

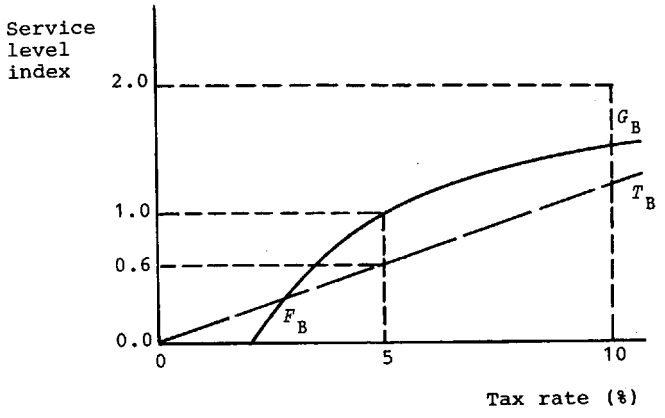

Figure 5.8

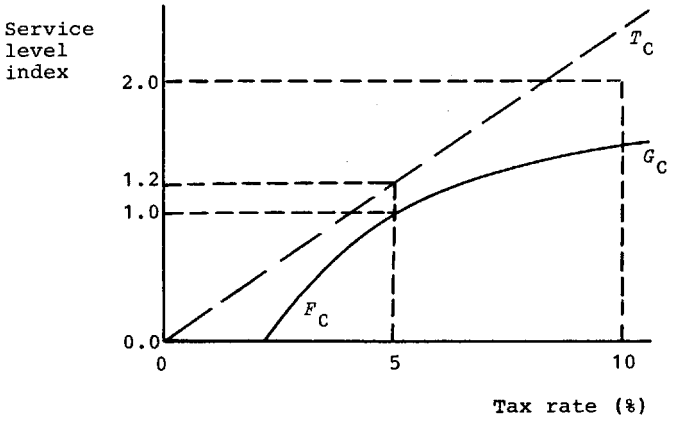

Figure 5.9

schemes would not help areas if they set low tax rates, for negative grants operate at low levels even for B; indeed, areas with low rates are given substantial incentives to increase these rates, for the slopes of F_AG_A, F_BG_B and F_CG_C are high at low rates. Thus this scheme embodies the principal advantages of the first scheme considered.

However, unlike the first scheme, this scheme discourages areas in receipt of grants, such as B, from raising their tax rates very far. F_BG_B in Figure 5.8 has the same slope as OG_B in Figure 5.2 at a 5 per cent tax rate, but flattens out progressively at higher rates. Thus such areas find increases in their tax rates progressively less rewarding. Indeed, the slope of F_BG_B eventually equals, and then gets less than, that of OT_B, indicating that grant receipts eventually cease rising and then start falling when the tax rate reaches a certain high level. It follows that the most extravagant areas will not necessarily receive the highest grants. Moreover, this scheme, again unlike the first, discourages areas such as C, liable to make payments to the grantee, from cutting their tax rates very far. F_CG_C in Figure 5.9 has the same slope as OG_C in Figure 5.3 at a 5 per cent rate, but gets progressively steeper at lower rates. Thus such areas find cuts in their tax rates progressively more painful. Indeed, as the tax rate falls, the slope of F_CG_C eventually equals that of OT_C and then exceeds it, indicating that payments to the grantor eventually cease falling and then start rising when the tax rate reaches a certain low level. It should be stressed that the scheme achieves these results with the curved shape of F_AG_A, F_BG_B and F_CG_C. This shape in fact discourages all areas from setting very high tax rates, whether they receive grants or not, for increases in t above 5 per cent become progressively less rewarding for all areas; it also discourages all areas from setting very low tax rates, whether they pay sums to the grantor or not, for reductions in t below 5 per cent have progressively larger adverse effects on service levels. One advantage of this latter property is that the scheme might enable the grantor to dispense with many of its controls or grants designed to secure minimum standards.

It will be seen in Figure 5.7 that the standard area A receives negative grants at any tax rate except 5 per cent, a fate it escapes in the other two schemes. If the scheme were intended to handle revenue-sharing as well as equalisation, then few areas, if any, would be as well off as A, and most or all would be in a position akin to B's where positive grants would be obtained for a wide range of tax rates. If the scheme were intended to handle equalisation alone, then, as with all such schemes, many areas must expect to find themselves eligible for negative grants.

However, if equalisation were the sole objective, then it would be necessary with any scheme to fix the position and slope of the FG lines to ensure that the aggregate value of the positive grants equalled the aggregate value of the negative grants. As this third scheme is typically less generous to both high tax rate and low tax rate areas than the other schemes, it will need to be more generous to those in between. Accordingly, it might achieve equalisation while permitting a higher service level index than the others for areas with a tax rate of 5 per cent, it used the same standard tax rate. In other words, the fate of any area, such as A, under the different schemes might be more truly brought out by the

appropriate figures if curve F_AG_A in Figure 5.7 were higher so that a tax rate of 5 per cent produced a service level index of, say, 1.1; F_BG_B and F_CG_C in Figures 5.8 and 5.9 would, of course, be higher also.

It may be concluded that, where the grantor has an overwhelming preference for avoiding any incentives for allocative inefficiency, then the second scheme will still be chosen; but where value judgements lead to a more finely balanced decision, then some effort-related scheme such as the third may be chosen in preference to the first.

Measuring expenditure requirements

Whichever scheme is chosen, it will be necessary to decide how to measure resources. If an effort-related grant scheme is used, then it will also be necessary to decide how to define revenue effort. These two issues are discussed in the next two sections. The present section is concerned with the measurement of needs and costs and so of expenditure requirements. It should be stressed that, in practice, not all equalisation schemes pay heed to variations in needs, despite the arguments in favour of so doing outlined in the last chapter; few schemes pay heed to variations in costs.

There are three steps in estimating the relative per capita expenditure requirements for any area. First, it must be decided whether or not differences in unit output costs are to be taken into account. If they are, then the aim will be to estimate a composite index (CN) for the area to use in determining its grant receipts; if not, then the aim will be merely to estimate the needs index (N).

Secondly, each service must be taken in turn, and an attempt made to see how much expenditure per head would be required to provide a given level of service in the area. This figure should then be divided by the expenditure required in some standard area to get an expenditure requirement index for that service. If unit cost differences were to be allowed for, then the estimate of expenditure per head required in the area would be made with reference to actual output costs there; if they were to be ignored, then the estimates would assume that the area could produce output units at some standard prices. This step is conceptually straightforward, but it is empirically hard. In estimating a needs index for road maintenance, for example, it would be necessary to measure factors such as the miles of road per head, the usage of the roads, the frequency of heavy frosts and snowstorms, and the age and construction of the roads, and then to define some standard level of maintenance and see how each of these factors affected the expenditure needed to sustain that level.

Thirdly, a weighted average of the indices for each service must then be calculated to arrive at a single overall index for the area itself. This raises the issue of how much weight to attach to the indices for each service, or, to put it another way, how much weight to attach to each service. Typically, the same set of weights is applied to each area. This process occurs explicitly in cases where the grantor defines some standard level for each service. It occurs implicitly in cases where needs indices are

estimated by regression (a process used in the late 1970s in the United Kingdom) to see which factors of need are most prevalent in high-spending areas; in such cases, the relationship between, say, miles of roads per head and total expenditure per head that applied to all authorities taken as a group would also be held to apply to each individual authority irrespective of how much importance it attached to roads.

King (1980b, pp. 51–8) has shown, however, that some horizontal inequity could appear if this approach were used. To illustrate the point at issue, a somewhat extreme example will be used. Suppose that differences in unit output costs are being ignored, and that two areas A and B have identical resources per head, and are entitled to provide just two services – X, where needs are exceptionally high in A and low in B, and Y, where the converse applies. Suppose, though, that X and Y are weighted equally by the grantor and that each area is credited with the same overall needs index. Suppose also that A and B levy taxes at the same rates, so that they raise equal tax revenues per head and, given their equal overall needs indices, subsequently merit equal equalisation grant receipts per head. Suppose finally that tastes are the same in A and B and that their citizens want all subcentral expenditure devoted to X. In this case, A and B would spend equal sums on X, with the result that standards would be higher in B where the needs index for X is lower. Thus, if Mr α in A and Mr ß in B were to have identical incomes and pay identical taxes, they would not enjoy equal service levels.

King has shown that this problem would disappear if each area's needs index was calculated by weighting the services in accordance with the weights given to them by that area.[7] Thus, in estimating the needs indices for A and B, the grantor would note that each attached no importance to Y and would therefore look exclusively at their needs for X, thereby giving a higher needs index to A. As a consequence, A would get more grant, enough to enable it to provide as high a level of X as B if equal tax rates were set.

Measuring resources

The major issue in measuring resources is whether reference should be made to an area's tax base or to some measure of its income; either way, there is, as Akin (1973, p. 281) has implied, a case for adding in an allowance to cover the extent to which the area can tax non-residents. This section focuses on this issue, which has been evaded so far in the text by assuming that areas levy an income tax and that the two approaches yield similar results. If another tax were used, say a general sales tax, then it would be possible to use either incomes or the tax base simply by defining the resources index (T) in the formulae as per capita income or the per capita sales tax base as appropriate. Clearly some areas would do better on one approach and others would do better on the other. Manvel (1971, p. 199) showed how states in the United States (in 1966–7) would fare if resources were related in turn to state tax bases and state personal incomes; he found that some states were very sensitive to the choice made.

For example, with the first measure, Connecticut and Wyoming had indices, respectively, of 109 and 148 (each national average being 100), while with the second they had indices of 125 and 93.

The case for using incomes is that under a normal equalisation scheme there are, *ceteris paribus*, bigger grants to low local tax resource areas than to high local tax resource areas. It is possible that a poor area endowed with high local tax resources could be seen as doing badly compared with a rich area endowed with low local tax resources. For instance, if regional governments were established in the United Kingdom, and if they were endowed with a tax on sales of tobacco, then Scotland would almost certainly get lower per capita grants than South East England. Since average incomes are higher in the South East, this might be felt to lead to poor vertical equity between areas of different incomes. In fact, though, Scotland would probably not do badly compared with the South East, since the grant would be financed from the proceeds of other central taxes, which would presumably be much higher in the richer area.

As it happens, there is a powerful horizontal equity argument against using incomes. Suppose an equalisation scheme were introduced in which tax effort was measured as the ratio of tax yields to incomes. A given tax effort in poor but heavy-smoking Scotland would, on this basis, mean a much lower tobacco tax rate than in the rich but light-smoking South East. If each region were to make the same tax effort, then there would be widely ranging tax rates. This would mean that people with equal incomes and similar smoking habits would be taxed differently in different areas, even though the grant scheme would seek to ensure they had the same (or similar) service levels, so the system would unquestionably lead to horizontal inequity. It is worth noting that the general revenue-sharing scheme in the United States started in 1972 by using a definition of tax effort that related to incomes. By 1975, it was being argued by Nathan, Manvel and Calkins (1975, p. 145) that 'the revenue-sharing law would be improved materially if it did not rely upon per capita income as a proxy indicator of relative fiscal capacity' but used indices relating to tax-raising potential. It is probable that this conclusion would apply to any other schemes that use income as an index of resources.

In general, then, it seems better to relate resources in equalisation and revenue-sharing grant scheme formulae to tax bases than to incomes where the two differ. An analysis by Foster *et al.* (1980, pp. 243–4), however, implies that, in the case of a property tax, there is a case for using modified tax bases rather than actual ones. The strength of the case depends on the reasons why the tax base varies from area to area. Suppose, first, that property values for similar properties are the same in all areas. Citizens in one area, A, may have a higher tax base per head than those in B simply because more expensive domestic properties are more common or because there is a higher proportion of non-domestic property. So, if each area sought similar service levels, then citizens in B would have to pay higher tax rates and thus more in taxes than people living in similar houses in A unless an equalisation scheme were introduced that fully offset differences in the tax bases. Without such a scheme, there would be an incentive to migrate from B to A, even though such migration offered no clear gains to society.

Suppose next that the tax bases differed only because similar domestic properties had higher capital and annual values in A, perhaps because A offered advantages such as a better climate or prettier countryside. In this case, citizens in B would have to pay higher tax rates than those in A, but because they had lower tax bases than people living in similar houses in A, whether those be the annual or capital values of their properties, they would not pay any more in tax. Thus there would be no differences in fiscal residuums and no necessary incentive to move from B to A as a result of subcentral authority taxes and expenditures; accordingly, there would be no need at all for equalisation grants.

In reality, differences in property tax bases occur for all these reasons. There is perhaps a case for defining the property tax base of an area for equalisation grant purposes as the value that its domestic property would be if property prices for any given type equalled average prices in the country as a whole for that sort of property, plus the actual value of all non-domestic property. In this way, the effects on domestic property of special locational advantages and disadvantages would be ignored, while differences in the mix of domestic property and in the value of non-domestic property would be taken into account.

In fact, it may be appropriate to equalise with respect to modified tax bases in the case of other taxes too, although the modifications will typically be less significant. Consider, first, a subcentral general sales tax or value added tax, and suppose to begin with that pre-tax retail prices are uniform through the country. Per capita sales values will be high, and tax rates might be low, in areas where retail outlets are concentrated; there will be inequity between citizens in these areas and those elsewhere unless an equalisation scheme compensates areas with few retail outlets. Suppose, instead, that the only reason why per capita sales values vary from area to area is that pre-tax retail prices are high in certain regions, such as off-shore islands, where substantial delivery costs are entailed. The island authorities could indeed levy lower tax rates than those on the mainland to provide similar service levels, but their citizens would be paying lower rates on higher prices, and so they would pay no less in tax than citizens elsewhere. Thus there would be no horizontal equity problem to tackle. In practice, sales tax bases may vary for both reasons. To ensure that equalisation grants relate to the first alone, it would be necessary to calculate each area's tax base on the assumption that it faced average pre-tax prices. In this way, variations stemming from varying pre-tax prices would be ignored.

Consider, next, a subcentral tax on earned incomes or a payroll tax, and suppose to begin with that the wage rates for any given job are uniform throughout the country. Per capita income tax bases will be high, and tax rates might be low, in areas where high-income jobs are concentrated and where employment levels are high; there will be inequity between citizens in these areas and those elsewhere unless an equalisation scheme compensates areas with low tax bases. Suppose, instead, that the only reason per capita tax bases vary from area to area is that the wage rates for certain jobs are higher in some areas than others; for example, they may be higher in areas where the climate is particularly unattractive.

The authorities in these areas could indeed levy lower tax rates than those in more temperate areas to provide similar service levels, but their citizens would be paying lower rates on higher wages, and so would pay no less in tax than citizens elsewhere. Thus it might be felt that there would be no horizontal equity problem to tackle. In practice, income tax bases may vary for both types of reason. To ensure that equalisation grants related to the first alone, it would be necessary to calculate each area's tax base on the assumption that its citizens earned average wage rates for their type of job. In this way, variations stemming from the second sort of difference would be ignored.

Measuring revenue effort

It was noted in Chapter 3 that effort-related grants could, in principle, take into account grantee revenues from taxes, charges and loans, though generally they consider tax revenue alone. To simplify the discussion, it will be assumed initially that grantees can raise revenue from no source other than a single tax. Then the possibility of levying more than one tax will be explored. Finally, the effects of their being allowed also to levy charges and to borrow will be considered.

Where a single tax can be levied, it seems clear that an area's tax effort should be related to its tax yield. The appropriate measure of effort will then depend on whether the area's resources are being defined in relation to some measure of its income or its tax base. Where resources are defined in relation to incomes, then it would be expected that an area's tax effort would be defined as the ratio of its tax income to its total relevant income. Where resources are being defined in relation to tax bases, then it would be expected that an area's tax effort would be defined as the ratio of its tax income to its tax base; that ratio, of course, is directly related to the area's tax rate (t). The tax rate is the index of revenue effort used, for example, in the United Kingdom.

If more than one tax can be levied, then there is no real problem in connection with the income measure of resources, for an area's tax effort can still be defined as the ratio of its total tax income to its total relevant income, as is the case with the United States' revenue-sharing grants. It is not so easy, however, to get a satisfactory measure of effort if resources are related to tax bases. One possibility would be to operate several separate schemes, one for each tax, but this arrangement would encourage areas to rely wholly on the tax where their resources were lowest, since in that way a given total tax yield would attract the highest amount of grant.[8] As a result, tax rates for any one tax could vary greatly from area to area, thereby perhaps stimulating much unproductive migration of residents, labour, capital and trade. To avoid this problem, it is necessary to have a single scheme, and to derive a single measure of effort from the tax rates applied. King (1980b, pp. 22–6) has developed suitably modified formulae for such schemes and has suggested that an area's effort be defined as its total tax yield divided by the total yield that would be produced if its taxes were levied at specified standard tax rates, so that it would be given

by $\Sigma (tT)/\Sigma (\bar{t}T)$. In this way, its effort would be proportional to its total tax yield but would not depend on which taxes it used most. There would therefore be no incentive to rely more on one tax and less on another.

The effect of allowing authorities to levy charges depends to some extent on how much autonomy they are given. For example, if an authority is permitted to levy charges for its public transport services but forbidden to run them at a profit for subsidising other services or at a loss to be met from other revenue, then these services will arguably be separate from all its others and need not be considered in general grant programmes.

Suppose, though, that the authority can run the services at whatever profit or loss it chooses. In this case, its charges can be seen as, in effect, a tax on transport users and should really be treated along with all its other taxes. If such charges were not so treated, then, as Fisher (1979, p. 182) has explained, authorities in receipt of effort-related grants would be tempted to cut their charge revenue and raise their tax revenue by a corresponding amount. In this way, their total own revenue would be constant, but the rise in taxes would attract extra grants whereas the fall in charges would not produce any fall in grants. In short, excluding such charges gives grantees an incentive to have zero charges and make up the lost revenue through high taxes.

Analogously, it could be shown that if loan revenue was excluded from revenue-effort measures, then grantees would have an incentive to cut their loan receipts and raise their tax (and any eligible charge) receipts. However, this incentive would be tempered by the fact that raising taxes imposes a burden on present citizens, while cutting borrowing eases the burden (of repayment) on future citizens. In any event, it is suggested in Chapter 8 that, on the whole, charges and taxes should be used to finance current spending, while loans should be used to finance capital spending. If this arrangement were enforced, and if, as is the usual case, effort-related grants sought to help finance current spending alone and were not to be spent on capital projects, then loans would cease to be substitutes for other revenue sources and there would be no problem in ignoring them in revenue-effort measures.

The determination of aggregate grants in effort-related schemes

A further problem with schemes where tax efforts affect grantees' receipts is that they necessarily affect grantor payments. Grantees would prefer to fix their tax rates when the *FG* schedules for the scheme had been fixed; yet a grantor anxious to secure a specific aggregate payment might not wish to fix the relevant parameters until all grantees' tax rates were known. If grantees were required to fix their tax rates before the grantor fixed its parameters, and thus before they knew how much grant they would receive, then this would increase any uncertainty they might have about their total revenue. Conversely, if the parameters were fixed before the tax rates were known, then the grantor would not be able precisely to determine its own aggregate outlay.

One way of eliminating uncertainty is to introduce a lag whereby grant payments in one year depend on the tax rates set in the previous year. One problem with this is that areas that change their tax rates have to wait a year before getting the appropriate changes in grant, and this might be felt to be serious, particularly in countries where the political complexion of authorities can change rapidly. Another possibility would be to use an iterative approach, which would require subcentral authorities to indicate their forecasted tax rates on the basis of some arbitrary values for the grant parameters. The grantor could then estimate its forecasted grant total, which, of course, would be unlikely to equal the total desired. It could then adjust the parameters and ask the authorities for revised tax rates. The process could continue until the grantor arrived at some parameter values which led to the authorities forecasting tax rates that would warrant grants whose sum equalled the required total. Unfortunately, there might not be time to find a perfect set of parameters. Moreover, if the scheme were to work, then the authorities would have to be compelled to set their final forecast tax rates; hence they would have to make each forecast with considerable care and, so, with considerable effort.

King (1980b, pp. 65–73 and 76) has suggested that a better arrangement might be for the grantor to set its parameters in advance and accept that its total grant payments might not be exactly as planned. Next year, it could adjust the parameters and try again. It should be possible for total grant payments, aggregated over a number of years, to be kept near to a desired figure. An alternative approach would be to try and reduce the importance of revenue-sharing grants and thereby the magnitude of any uncertainty that might be felt by the grantor or the grantee, as the case might be. In many countries, the importance of the grants could be reduced by increasing the tax powers of grantees, thereby diminishing the need for equalisation grant schemes also to satisfy a revenue-sharing purpose and allowing them to concentrate more on pure equalisation, and, in turn, to have a lower aggregate value.

In practice, however, grantors operating effort-related schemes seem to prefer to fix final grant payments after all grantee tax rates are known. Certainly this applies to the two major relevant schemes outlined later in this chapter, namely the revenue-sharing scheme used in the United States since 1972 and the rate support grant scheme used in the United Kingdom. One result of this arrangement is that grantees cannot be certain of their total revenues when they set their tax rates. This is potentially quite serious in the United Kingdom where the relevant grant receipts can form a large proportion of a grantee's total income. Another result is that if an area changes its tax effort, then its income may not vary as much as the grant scheme suggests.

This last point can be explained with the help of a simple example. Suppose a grantor wishes a grantee's total tax and grant receipts to be proportional to its tax effort. Suppose also, initially, that there are just three areas (A, B and C), which have equal populations and which make equal tax efforts and are identical in all other respects so that each qualifies for a similar grant that turns out to be £40. Now, if A considers doubling its tax effort, then it would presumably hope its grant would rise to £80,

double that of B and C. However, if the grantor wishes to maintain total grant payments at the original total of £120, then it could pay A twice as much as B and C only if it gave A £60 and the others £30 each. It is true that, *ex post*, receipts would be proportional to effort, but the point for A is that it knows a doubling of its tax effort would, *ceteris paribus*, raise its grant by a mere 50 per cent, despite the alleged proportionality of the scheme.

The problem for A is less serious, though, the more grantees there are or the smaller it is in relation to the grantees as a whole. Suppose, for instance, there were four identical areas (A, B, C and D) making identical efforts and receiving £40 each. If A doubled its effort, then it should get twice as much as each of the others. If the total outlay remained at £160, then A would get £64 and the others £32 each. This time, a doubling in A's tax effort has raised its grant by 60 per cent.

It follows that the intended relationship between effort and grant cannot be fully achieved if the total outlay is predetermined, but the relationship for a given area will be affected less when the number of grantees is high and the area is relatively small. Ironically, perhaps, Fisher (1979, p. 182) has suggested that the United States revenue-sharing scheme should allocate fixed totals to small areas, such as counties, so that each local authority would compete with few others. This is because he feels the effort-related nature of the scheme encourages overspending, and that this encouragement would be mitigated by his proposal, which would reduce the benefits to a given area from raising its tax effort. However, his proposal means that such benefits would vary from area to area depending on how many other areas each was competing with. It would arguably be better to put all areas on a similar footing by putting them all in the same pool, and reducing any incentive for overspending by having a scheme where grant receipts were less sensitive to tax effort.

Equalisation and revenue-sharing grants in practice

The preceding sections of this chapter have given a theoretical analysis of some of the major issues that arise when grant programmes are designed with equalisation and revenue-sharing objectives in mind. The following sections are concerned with outlining the programmes that are (or have been) in use in a number of selected countries – specifically, Australia, Canada, the United Kingdom, the United States and West Germany. Before looking at these arrangements, however, it is useful to make three general observations, and then to consider the relative importance of equalisation and revenue-sharing in each country.

First, and perhaps not surprisingly, the various pros and cons of schemes to equalise fiscal potential and fiscal capacity, and of intermediate types of scheme, have led different countries to adopt different arrangements. In practice, the arrangements tend not to be as simple as those outlined in the previous sections; they are generally modified and constrained by specific national problems or electoral wishes. Nevertheless, it is possible to relate the basic philosophy of some countries' schemes to those

considered above. In Australia, Canada and West Germany, for example, the tax effort of a state, province or *Land* is not considered when its grant receipts are calculated. Accordingly, as Mathews (1977, p. 426) has observed, it is broadly the equalisation of fiscal capacity approach that is in force in these three countries. It should be appreciated, however, that it is inevitable that the scheme in West Germany ignores tax effort, for the *Länder* (as opposed to local authorities) are not given individual discretion over the tax rates that they set and so these rates are uniform over the whole country. In contrast to the schemes in these three countries, the schemes in the United States and United Kingdom do take tax effort into account; they bear some relationship, respectively, to the scheme for equalising fiscal potential and the alternative scheme illustrated in Figures 5.7, 5.8 and 5.9.

Secondly, it should be stressed that only the Australian and United Kingdom schemes attempt to neutralise variations in needs as enthusiastically as they attempt to neutralise variations in resources. Variations in costs are ignored everywhere except in Australia.

Thirdly, all the grant schemes described below have a revenue-sharing objective as well as an equalisation objective, except the scheme in West Germany (where the tax-sharing arrangements described in the next chapter minimise the need for revenue-sharing grants). Accordingly, it is only the West German scheme which involves negative grants for some areas, and these are used to finance positive grants to the others.

It was explained in the last chapter that wherever there are differences in per capita resources and spending requirements there is a case for equalisation grants. However, it is not unreasonable to say that the strength of this case depends chiefly on the extent to which resources and needs vary from area to area. Unfortunately, it is not possible to compare variations in needs in all the five countries under discussion because no comprehensive figures on needs are available in the three countries that ignore variations in them for equalisation purposes. As it happens, the evidence from the other two countries suggests that needs may vary less than resources. Of course, the same would not necessarily apply elsewhere; in particular, it should be stressed that the large resource variations in the United Kingdom are a result of local authorities being permitted to levy just one tax, the property tax, which (as noted in the next chapter) is very unevenly distributed.

The variations in per capita spending needs by the upper tier of authorities in the United Kingdom (or, more strictly, England and Wales, since comparable figures are not available for authorities in Scotland and Northern Ireland) have been estimated by Bennett (1982, pp. 315–17) and show, for 1980–1, a ratio of 2.00:1 between the authority with the highest needs index, Powys, and the authority with the lowest index, Surrey. This compares with a ratio in per capita tax resources (that is in the per capita values of the property tax base) of 4.01:1 between the richest authority, Greater London, and the poorest, Mid-Glamorgan (CIPFA, 1982, pp. A1 and B1).

In Australia, variations in needs and resources are given in figures produced by the Commonwealth Grants Commission (1982, pp. 228–9).

Ignoring Northern Territory, which is treated as a special case, the per capita spending requirements for a standard package of services at standard levels ranged in 1980–1 from $954 in Tasmania to $763 in Victoria, that is in a ratio of 1.25:1. On the other hand, the per capita revenues that each state would have acquired if it had levied a standard package of taxes at standard rates ranged from $437 in Western Australia to $282 in Tasmania which is in a ratio of 1.55:1. It is interesting to note that, in Australia, a single authority has both the highest needs and the lowest resources, while in the United Kingdom, the authority with the highest needs, Powys, has the second lowest resources.

It will be noted that the proportionate variations in the per capita resources for the highest tier of subcentral authorities are much higher in the United Kingdom (at 4.01:1) than in Australia (at 1.55:1). The ratios in Canada, the United States and West Germany all seem to fall between these two, and these remaining countries will be taken in turn in descending order.

In Canada, estimates have been prepared to show the per capita receipts each province would have if it set a representative array of taxes at representative tax rates. In 1980–1 (as shown in ACIR, 1981d, p. 55) these varied from around $3,025 in Alberta to $935 in Prince Edward Island, a ratio of some 3.24:1. Alberta's resources were boosted by its high per capita oil and gas revenues, but even the second richest province, British Columbia, with its more modest per capita natural resource revenues, had an overall per capita resource level more than double that of Prince Edward Island.

In the United States, one straightforward way of comparing states' relative resources is to look at their per capita personal incomes.[9] In 1980, these ranged from $12,790 in Alaska to $6,580 in Mississippi (United States Department of Commerce, 1981, p. 429), a ratio of 1.94:1. If Alaska is regarded as a special case, and also the District of Columbia (which had the second highest level), then the comparison would be between Connecticut, at $11,720, and Mississippi, giving a ratio of 1.78:1.

The range of per capita resources between *Länder* in West Germany can be derived from per capita tax revenue figures since all their taxes (whether involved in the tax-sharing scheme or not) are effectively levied at the same tax rates in each *Land*. In 1976 (see ACIR, 1981b, p. 24) these revenues ranged from 2,259 DM per head in Hamburg to 1,303 DM per head in Saarland, that is a ratio of 1.73:1.

It is perhaps even harder to make comparisons of the relative need for revenue-sharing grants. The ideal arrangement would be to have an aggregate account for each country's subcentral authorities, with each of these accounts drawn up on an identical basis, and for these accounts to show what proportions of aggregate subcentral revenue arose from revenue-sharing grants on the one hand and from taxes, tax-sharing arrangements, charges, loans and remaining grants on the other hand. It should perhaps be emphasised that the only other grants yielding a positive aggregate revenue would be those paid in connection with externalities or enforcing grantor preferences; a scheme of equalisation grants concerned with equalisation alone, and not concerned with

revenue-sharing, would entail negative grants for some areas to finance positive grants to the others, and would not affect the aggregate revenue of subcentral authorities.

Unfortunately, the available aggregate accounts are not drawn up on an identical basis. In particular, they tend to differ in how far they show receipts from charges and expenditure financed from such receipts, and, perhaps more importantly, they do not specify precisely which grants are paid for revenue-sharing purposes and which for other purposes. The arguments presented in the last chapter suggested that grants for correcting externalities and enforcing grantor preferences should be specific grants while revenue-sharing grants should be general. Clearly, if these prescriptions were always followed, then the aggregate payments of general grants in any country would indicate the size of the revenue-sharing programme.

However, as also noted in the last chapter, it seems highly probable that many central governments handle much of their revenue-sharing (and equalisation) problems by means of specific grants. The use of specific grants for revenue-sharing reasons is not in fact as curious as it might at first seem. Central governments have a legitimate case for wishing to influence the way in which any of their grant payments are spent. In the case of revenue-sharing grants, this desire for control could be manifested by a general grant scheme coupled with detailed regulations of one sort and another. Arguably, this approach is used in the United Kingdom, where the substantial general grant payments are paid with few detailed strings attached, but where, in 1971, it was reported that the central government department paying them out had more than 1,000 statutory controls over the expenditure behaviour of local authorities;[10] admittedly, some 300 less frequently used ones were abolished in 1980. The desire for control could be manifested instead by using an array of specific grants each with its own conditions. This approach may often be more acceptable in political terms for federal-state grants in federal countries, and it would be consistent with the view that a large slice of the specific grants in the United States should be seen as primarily concerned with revenue-sharing.

Given these data limitations, the only way in which even crude international comparisons can be made is to compare the relative importance of tax revenue (including genuine tax-sharing receipts), net borrowing and all grants. The fact that specific grants account for only about one-fifth of grant payments in the United Kingdom possibly suggests that most of the aggregate payment of grants there is concerned with revenue-sharing, and the same may be the case elsewhere; the fact does not prove this of course, because, for example, the United Kingdom's central government may use general grants plus regulations to enforce grantor preferences.

The relative importance of these three sources of revenue is shown for selected countries in Table 5.2, where grants (or intergovernmental revenue as they are known in some official publications) exclude advances or loans from other levels of government, and where the figures for net loans are in many cases derived by subtracting net indebtedness at the beginning of the year from net indebtedness at the end. It should be stressed that

Table 5.2 *The relative sizes of tax, grant and (net) loan revenue in selected countries*

Country	*Year*	*Level*	*Taxes* %	*Grants* %	*Net loans* %	*Tax: grant ratio*
Australia	1979–80	State	30	63	7	0.48
		Local	57	24	19	2.36
Canada	1977–8		66	24	10	2.76
	1977	Provincial	40	50	11	0.80
		Local				
United Kingdom	1981		39	60	0*	0.65
		Local				
United States	1980	State	65	30	5	2.13
		Local	39	46	14	0.84
West Germany	1976	*Land*	70	19	11	3.67
		Local	49	43	8	1.13

* The figure was under 0.5 per cent; net borrowing was particularly small in 1981.

Note: The percentage figures for a given set of authorities may not add up to 100 as a result of rounding errors.

Sources: For Australia, Mathews (1982a) pp. 164–5; for Canada, Ministry of Supply and Services (1981) pp. 822–3, 836–9 & 844; for the United Kingdom, HMSO (1982b) pp. 55–6; for the United States, United States Department of Commerce (1981) pp. 279 & 289; and for West Germany, ACIR (1981b) p. 63.

the figures relate to a variety of years and that different sources were used for different countries. In particular, it is not clear just how comparable the net loan figures are and, although loans are always relatively small, it is probable that the tax:grant ratios shown in the final column are the least unsatisfactory ones to use for comparisons. It should also be stressed that the figures mask wide variations between individual grantees in the different countries.

The figures in Table 5.2 suggest that, at the state level, the Australian authorities are by far the most dependent on grant revenue and the least well endowed with tax revenues. The Canadian provinces, where tax revenues include such items as natural resource revenues and the liquor profits made by the provincial agencies that exercise monopolistic control over the distribution of alcoholic beverages, seem slightly less in need of grants than the states in the United States, but rather more in need of them than West Germany's *Länder*. It should be recalled, though, that the high tax revenue of those *Länder* gives a misleading impression of their autonomy since an individual *Land* cannot determine its own tax rates.

At the local authority level, the need for grants is least in Australia. It is smaller in West Germany than in Canada or the United States, and

it is highest in the United Kingdom. It should be pointed out that the 39 per cent share shown for taxes in the United Kingdom in 1981 is much higher than in most recent years. The figure was forced up by the unusually low borrowing level (explained by unusually low capital expenditure) and a tight grant policy; in 1976, for example, taxes produced less than a quarter of the total for taxes, grants and net loans.

Now that some idea has been given about the need for equalisation and revenue-sharing grants in the selected countries, it is possible to look at their actual grant schemes. The countries will be taken in an order that involves describing the schemes with the simplest basic principles first and the ones with most complex basic principles last. As it happens, this means taking to begin with the three schemes where there is no revenue-effort factor.

The West German arrangements

Table 5.2 indicates that the revenue-sharing requirements in West Germany are, at most, very modest. This, of course, is a result of the tax-sharing scheme in use there. There is still a clear need for an equalisation scheme, however, and the principal element of equalisation in West Germany is the financial settlement between the *Länder*.[11] In essence, a figure is computed to show the total per capita yield of each *Land*'s own taxes together with its share of shared taxes and (as it happens) half the yield that the local authorities in it would raise if they set standard tax rates. This per capita yield is then compared with the average figure for all the *Länder*, and transfers are made by the richer areas to the poorer areas. As noted earlier, little attention is paid to variations in need, although a little extra is allowed to areas with high population densities. The transfers are not sufficient to bring the rich areas down to the average or the poor ones up to the average; in other words, there is only a partial equalisation of fiscal capacity.

This equalisation scheme, which involves no federal grants, is augmented by supplementary federal grants made on a small scale to *Länder* with below average resources or above average needs. Even so, in 1976, the total *Land* incomes from taxes, inter-state equalisation payments and supplementary federal grants varied quite considerably in per capita terms, from 128.0 DM in Hamburg to 96.5 DM in Bavaria (Traber, 1980, p. 277).

As far as municipalities in West Germany are concerned, some equalisation is effected by grants from the *Länder*. Although the municipalities have some discretion over their tax rates, their grant receipts do not vary with their tax rates.

The Canadian arrangements

In Canada,[12] general grants make up only about one-third of the payments by the central government to the provinces, and it is arguable that the large specific grants, especially those relating to health services and higher education, are in some measure a substitute for general revenue-

sharing grants. Nevertheless, the equalisation grants, which form the bulk of the general grants, add a significant net sum to the aggregate provincial revenue. In effect, there have been some twenty-nine separate equalisation of fiscal capacity schemes since 1977 covering virtually all the sources of provincial revenue (other than loans), although the number rose to thirty-three in 1982.

Between 1977 and 1982, a province's grant entitlement for a given tax was in general calculated as the difference between (i) the yield it would have had if it had set the national average tax rate and (ii) the yield it would have had if it had set that rate and if its tax base per head had equalled the national average tax base per head. In 1982 the scheme was modified slightly; one change is that, instead of referring to the national average tax base per head, it now refers to a standard tax base per head computed with reference to the bases in the five median provinces. An interesting feature of this scheme is that each province's tax base must be calculated for each of the thirty-three sources of revenue, even though few provinces, if any, actually levy all thirty-three. A given province might have a positive entitlement for some taxes and a negative one for others, but simple addition of the thirty-three entitlements gives a final net entitlement for that province; this final figure might also be positive or negative. No allowance at all is made for varying needs. If, overall, a province has a positive entitlement, then it receives a grant equal to its aggregate entitlement, but if, overall, it has a negative entitlement – suggesting that it is, overall, favoured with above-average tax bases – then it is ignored by the scheme and enjoys a superior fiscal position to the others; typically, Alberta, British Columbia and Ontario have been in this position. The use of fiscal capacity equalisation at the provincial level means that an individual province's grant receipts are, in principle, independent of its own tax rates. In practice, this is not quite so, for its own tax rates enter the calculations for the national average tax rates; as these are weighted averages, changes in the rates set by one of the larger provinces, at least, can have a noticeable effect in the same direction on its own grant receipts.

An important, and somewhat controversial, feature of the 1982 changes concerns the treatment of revenues from non-renewable resources, in particular oil and natural gas revenues. Until 1982, the equalisation treatment for these revenues differed from the general arrangements outlined in the previous paragraph in that provinces with low bases were given entitlements sufficient to take them only half way up to the national average base. Thus such provinces were treated less generously than those suffering from low bases for the other revenues. Since 1982, though, the anomalous treatment of resource revenues has been terminated.

The provinces run a variety of revenue-sharing and so-called equalisation grant schemes for the local authorities in their areas. However, all the so-called equalisation schemes entail gross outlays for the provinces concerned, thereby raising aggregate local authority revenues. The revenue-sharing schemes generally distribute funds on a per capita basis. The equalisation schemes are principally concerned to reduce the disparities in resources between local authorities, although grant payments in some provinces make some allowance for needs, particularly in respect of varying

road mileages. Most of the schemes involve lump-sum grants, but the equalisation schemes in Quebec and Saskatchewan involve tax-effort factors. In all these cases, each grant is broadly proportional to the grantee's tax effort, so that in this respect these two schemes are related to equalisation of fiscal potential schemes. However, neither of these schemes secures complete equalisation.

The Australian arrangements

The arrangements in Australia[13] have changed markedly in recent years. Until 1975, most of the revenue-sharing aid given to the states by the federal government took the form of general financial assistance grants and substantial specific grants. Between them, these grants went some way towards neutralising the differences in resources, needs and costs of the states. They were supplemented by carefully thought out equalisation grants that sought to bring recipient states up to the average position of the two richest states, New South Wales and Victoria, which were known as the standard states. Naturally, these equalisation grants always entailed a positive gross outlay for the federal government. Unlike the West German and Canadian schemes, the Australian scheme attempted to make full allowance for needs and costs. In practice, a recipient state could be seen as receiving two grant elements. The first element, concerned with resources, was sufficient to ensure that its combined resources element plus tax revenue would be the same in per capita terms as the average tax revenue of the standard states if it chose to set the average tax rates set by them. The second element, concerned with expenditure requirements, was sufficient to ensure that it would have to find from other sources no more money per head than would be needed, on average, by the standard states if it chose to provide the average range and level of services provided in those states; in other words, the recipient was compensated if its needs and unit costs exceeded the levels in the standard states. The calculations made no reference to the tax rates or service levels that a recipient state actually chose.

This system attempted to secure a much greater degree of equalisation than the West German scheme or the Canadian one, but three factors prevented complete equalisation of fiscal capacity. First, the richer of the two standard states was not brought down to the average of the two. Secondly, the poorer of the two was not brought up to the average of the two. Thirdly, three of the four other states did not actually claim the grants because they had benefited from special financial deals with the Commonwealth government and, in Mathews's view (1982b, pp. 14–15), were effectively over-equalised anyway. In fact, the positions of the standard states are very close to each other, so the first two of these factors were of modest importance.

It should be noted that the equalisation grant entitlements were calculated by the Commonwealth Grants Commission, whose recommendations were always accepted by the central government. In calculating the resources element, this Commission took into account the

possibility that some claimant states might have some revenue sources, such as mineral royalties, not available to the standard states; it also bore in mind the fact that the base of some taxes, such as gambling taxes, might be affected by the recipients' tax rates. In calculating the expenditure requirements element, the Commission allowed for the fact that a recipient might face relatively high costs for a unit of output, as well as the fact that it might need a relatively high number of units of output per head.

In 1975, however, a start was made on replacing the general financial assistance grants and most of the specific grants with a system of revenue-sharing under which the federal government would pay to the states a share (around 33–40 per cent) of its income tax receipts. From 1981–2, the states have instead been paid a share of total federal government tax receipts. The Commonwealth Grants Commission was asked to advise on the division between the states of the sum concerned in each year from 1981 onwards. In essence, their proposed new scheme was a development of the old, which closely approached a pure equalisation of fiscal capacity scheme. In effect, each state's receipts could be seen as comprising two elements. Variations in resources would be equalised by making payments from the revenue-sharing funds in such a way that if each state levied taxes and, interestingly, charges at the average rates then the total per capita income from the first element and the taxes and charges would be the same in all states. Variations in expenditure requirements would be equalised by making payments of the second element from the revenue-sharing funds in such a way that if each state sought to provide services of the average range and level then each would need the same per capita revenue from sources other than this element. So far the Commonwealth Grants Commission's proposals have not been fully implemented, however, and until 1984–5, at least, the relative amount of the total payments received by each state seems likely to be influenced more by its pre-1981 share of federal payments than by the Grants Commission's proposals (see Mathews, 1982b).

This discussion of Australian arrangements may be concluded by noting that the federal government also pays modest general grants to local authorities on the advice of the Commonwealth Grants Commission. These grants have an equalisation basis, but they do not by any means secure complete equalisation. Like the grants paid by the states to local authorities, they ignore grantees' tax efforts, except insofar as these are taken into account in determining the standard tax rates used by the schemes.

The United States' arrangements

In contrast to the grant schemes considered so far, the general revenue-sharing scheme in the United States and the rate support grant scheme in the United Kingdom do take tax effort into account (so, incidentally, do revenue-sharing grants in Switzerland[14]).

Compared with Australia, Canada and West Germany, federal governments in the United States have been relatively unconcerned with inter-state equalisation. The most systematic attempt at a scheme

embodying equalisation principles was the general revenue-sharing scheme, introduced in 1972 for the benefit of both states and local authorities but continued since 1980 for the benefit of local authorities alone. This section concentrates on the arrangements as they were from 1972 to 1980.[15] It should be stressed that the grants paid under this scheme have always been a relatively small part of total federal grants, although they form the bulk of general grants. The remaining grants vary from wide specific grants (known as 'block' grants) to narrow specific grants (known as 'categorical' grants) and are, to some extent at least, a substitute for general revenue-sharing grants. Many of these other grant programmes take recipients' fiscal positions into account but, in 1980 at least, there appeared to be a small positive correlation (of 0.154) between states' grant receipts and their revenue capacities (Davis and Lucke, 1982, p. 350).

The revenue-sharing arrangements for states were somewhat complicated by the fact that they used two different formulae. In one, the Senate formula, the per capita grants paid to individual states varied in inverse proportion to their resources (defined in relation to personal income) and in direct proportion to their tax efforts (defined as total state and local tax yields as a percentage of personal income); no reference at all was made to needs. In the other, the House formula, 17 per cent of the grant was paid to states in proportion to their tax efforts, and a further 17 per cent was paid in proportion to the states' own income tax collections (subject to statutory upper and lower limits); the remaining 66 per cent was paid on a lump-sum basis that favoured states with low resources and urbanised areas, this last factor being the only reference made by the formula to needs.

It should be stressed that, while poorer states might expect to receive more than richer ones, the sums involved were always far too small to make much impact on reducing the gap between them. In 1972, there appeared to be a small negative correlation (of -0.19) between the states' per capita revenue-sharing grant receipts and their per capita general revenue capacities, and there was a slightly larger negative correlation (of -0.29) between their per capita receipts and their per capita tax capacities (Nathan *et al.*, 1975, p. 87). In 1979–81, however, there appeared to be a very small positive correlation (of 0.076) between per capita receipts and per capita revenue capacity (Davis and Lucke, 1982, p. 351). This may have been a result of a tendency for those states having lower capacities to make less effort, or of some of the other factors affecting receipts in the House formula being stronger in the richer states.

The grant distribution process involved two stages. First, each state (and the District of Columbia) had its entitlement calculated on the basis of each formula, and was then given the higher of these two amounts, subject to any uniform percentage reduction that might be necessary to reduce the total payment of grants to the sum appropriated for the year concerned. It follows that if a state actually received a grant under the Senate formula then, *ceteris paribus*, an increase in tax effort there would be expected to raise its grant receipts almost proportionally. If it received a grant under the House formula then increases in tax effort would have only a modest effect until they were sufficient to raise its entitlement under the Senate formula above that under the House formula. State governments actually

collected only one-third of the state entitlements. The remaining two-thirds were allocated to general purpose local authorities in the state. These proportions applied to all states, irrespective of the division of functions between the states and local authorities.

The second stage concerned the allocation of the local authority shares to the 39,000 or so relevant local authorities. In principle, states could either determine their own allocation procedures or they could use a procedure suggested by the federal government. In practice, all states adopted the latter approach. To begin with, the local authority share was allocated between county areas on the principles embodied in the Senate formula. Then, within each county, the funds were divided between the county government on the one hand and the municipalities and townships on the other in proportion to their (non-school) tax revenues. Finally, the distribution between municipalities and townships was done on the basis of the principles of the Senate formula. In general, therefore, municipalities and townships had receipts that were in principle proportional to their tax efforts, a result indicating some similarity between this scheme and schemes that seek to equalise fiscal potential.

Three points should be made about the relationships between tax effort and grant receipts. First, there were statutory floors and ceilings on the per capita amounts any local authority might receive: no local authority could receive less than 20 per cent or more than 45 per cent of the average per capita payment to local authorities in the state. Secondly, any change in an area's tax rates had an uncertain effect on its grant receipts in the following year because the sums available for distribution and the effort of many other authorities affected the final outcome. Thirdly, each authority was competing for a share of a cake of given size; thus, as explained in an earlier section, the fewer competing authorities there were in any one case, the less proportional any one authority's receipts were likely to be to its tax effort.

The arrangements for state grants to local authorities vary, of course, from state to state. Overall, specific grants typically form some 90 per cent of the total (Break, 1980, p. 180), many of them relating to education, although specific grants (especially ones relating to education) often have equalising elements built into them. Virtually all states provide some general grants, however. Around half of this general grant aid comes in the form of the tax-sharing arrangements noted in the last chapter, and around half comes in the form of revenue-sharing grants allocated on a per capita basis. Thus general grant assistance typically ignores variations in need and tax effort, and it frequently also ignores variations in resources.

The United Kingdom's arrangements

The present system of revenue-sharing and equalisation grants used in the United Kingdom is embodied in payments of the rate support grant (RSG). The RSG system has evolved over many years, and the description in this section is concerned chiefly with the arrangements as they have been in England since 1981–2, following the introduction of significant changes.[16] Since the United Kingdom is not a federal country, it might

be supposed that the RSG system would apply uniformly throughout, but this is not the case. The arrangements in Wales were very similar to those in England, and those in Scotland broadly similar, but the central government had the power to withhold part (or possibly all) of a Scottish grantee's RSG entitlement if it wanted that authority to spend less. Special arrangements were in force in Northern Ireland where the need for revenue-sharing, at least, was much less than elsewhere since most local government functions there were taken over by the central government in 1973.

The RSG has two components, one known as the 'domestic' element and the other as the 'unitary' (or block) grant because it is a union of the previously separate resources and needs elements. To understand the system, it should be borne in mind that the country has a broadly two-tier system of local government, that the sole local tax is the property tax known as 'rates', and that each authority is free to set its own tax rate known as the 'poundage'. The lower-tier authorities collect their own tax and also, on an agency basis, the tax for the upper tier. In principle, the same rate of tax is levied on the occupiers of domestic and non-domestic properties, but the domestic element of the RSG is paid to lower-tier authorities in amounts that enable them all to give a uniform reduction in the poundage actually collected from domiciles. In 1981–2, this element accounted for around 7 per cent of the RSG (and 9 per cent of rate collections), whereas it was just over 10 per cent between 1974 and 1978.

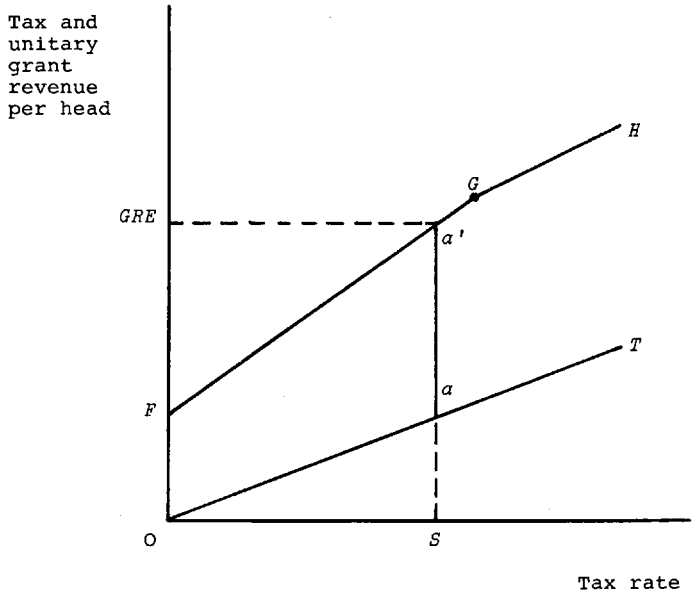

Figure 5.10

It follows that the entire equalisation exercise and the bulk of the revenue-sharing exercise are shouldered by the unitary element. Like the Australian scheme, this attempts to take both resources and needs into

account; unlike that scheme, it includes a tax-effort factor. A figure known as the grant-related expenditure (GRE) is calculated for each authority, which represents an estimate of how much it would have to spend to provide its services at standard levels, after allowing for receipts of charges at centrally acceptable rates and all other grants. This is shown in per capita terms by the level *GRE* in Figure 5.10, which relates to a hypothetical authority. The line O*T* shows how much tax this authority could raise per head from various tax rates.

The first stage in calculating entitlement to the unitary grant is to suppose that the authority set a specified tax rate, O*S*. In this case, its per capita grant receipts would be set at *aa'*, which would be sufficient to bring its total income up to the *GRE* level. This stage is broadly similar to the system used in Australia, but, whereas there each state is actually given that entitlement irrespective of its tax effort, in the United Kingdom it may be given more or less depending on its effort. The 1981–2 situation is illustrated in Figure 5.10 where the line *FGH* shows total per capita tax and unitary grant income, so that the gap between this line and O*T* represents unitary grant receipts. As the tax rate fell below O*S*, so the grant fell and the gap narrows. As the tax rate rose above O*S*, so the grant rose and the gap widens. However, the kink in *FGH* at *G*, a point reflecting tax and unitary grant income 10 per cent higher than *GRE*, shows that, beyond *G*, tax rate rises became rather less rewarding for this area.

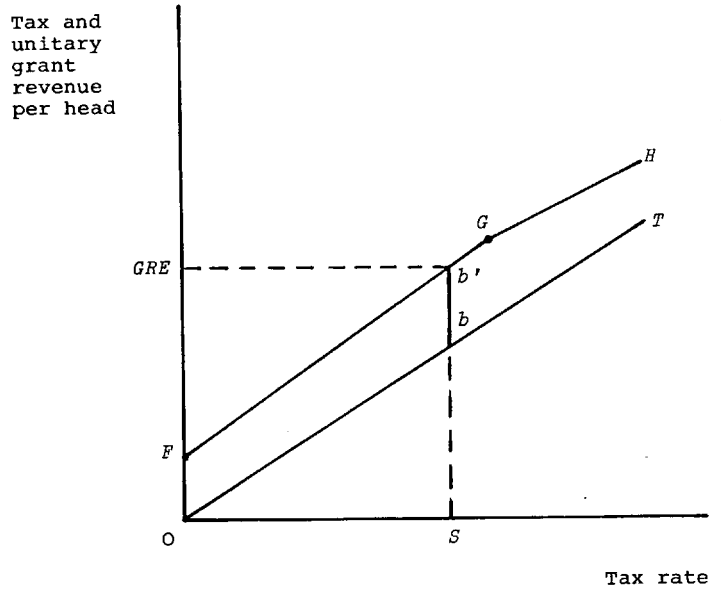

Figure 5.11

Figure 5.11 shows the situation for an area that had similar responsibilities but that had higher resources and lower needs. Higher resources are indicated by a steeper line *OT* for per capita tax receipts, while lower needs are indicated by a lower per capita *GRE*. If this area had set the standard tax rate *OS*, then it would have received a grant of *bb'*. The grant at other tax rates is given by the gap between *FGH* and *OT*, where the two segments of *FGH* have the same slopes as those in Figure 5.10 and kink at a level 10 per cent above this second authority's *GRE*. For this authority, *OT* was steeper than *GH*, so tax rate rises beyond the kink would have resulted in progressively smaller grants since the gap between *OT* and *FGH* narrows; eventually the grant entitlement could in theory have become negative, but such authorities would have been 'let off' with zero grants. As it happens, few authorities had sufficient tax resources to be in the position shown in Figure 5.11 with *OT* steeper than *GH*. The average authority found that each extra pound of tax revenue would bring in an extra 33 pence of grant along *FG* and an extra 7.5 pence along *GH*.

This scheme is related to the one shown in Figures 5.7, 5.8 and 5.9 in that tax increases eventually become less rewarding. The relationship is a little more marked in Wales, since the section of the line equivalent to *GH* is curved for authorities there (see Bennett, 1982, pp. 114–15), and it became stronger elsewhere in 1982 when the government introduced a further kink to try to curb high-spending authorities. In general, these further kinks fall along *GH*, but the reforms are complex and were done in a rushed and untidy way (see Bennett, 1982, pp. 117–25) so that the effect was different for different authorities; it is probable that the new system will be tidied up before long. With the RSG scheme, tax cuts leading to low spending are deterred by lower grants, but the scheme presented in Figures 5.7, 5.8 and 5.9 provides much more of a deterrent against low tax rates. A significant difference between these two schemes is that, although areas in the United Kingdom with similar responsibilities face *FGH* lines with similar slopes, so that equal rises in tax rates (on the same side of the kinks) lead to equal changes in per capita revenues, such rises will not generally lead to similar changes in service levels; areas with high needs will be unable to raise service levels by as much as areas with low needs if they receive equal increases in per capita incomes. This problem arises because the RSG scheme takes needs indices into account when assessing an area's GRE (and hence the grant that it would merit if it set the standard tax rate *OS*), but it ignores needs indices when deciding how much more or less grant is deserved at tax rates different from *OS*. Nevertheless, two areas with similar non-standard tax rates could have service levels differing only because of differences in their needs; in Australia and Canada, for example, they could differ also as a result of variations in their resources.

Three further points deserve mention. First, if an authority has a very high tax base, then its *OT* line could in principle pass above the *GRE* level at the standard tax rate *OS*. This is rare, but does happen in, for example, the Orkney and Shetland Islands in Scotland. Full equalisation would demand negative grants for such areas, but these are not made.

Secondly, each authority's grant as calculated by the formula should be seen as an initial estimate. If the aggregate payout calculated in this way exceeds the central government's predetermined total grant figure, then grantees will receive rather less than these initial estimates.

Thirdly, although the assessment of needs attempts to cover a wide variety of relevant factors, these are generally concerned solely with variations in the number of units of output per head required. For example, the police component of an area's GRE will depend on its crime rate, and the education component on the number of children. Unlike the Australian scheme, variations in the costs of output units are not generally included. For instance, regional variations in school heating costs are ignored. The education component does, however, permit a higher level of spending on, for example, the children of immigrants, manual workers and one-parent families, and the children in families with four or more children.

This service-by-service approach to expenditure needs has been criticised by Jackman (1981, p. 33) as presenting a serious, if latent, threat to local autonomy on the grounds that the central government may see the figures as desirable guidelines and attempt to make grantees conform with them. This approach to needs nevertheless removes a major problem in the aggregate approach used from 1974 to 1981. On that scheme, an annual multiple regression analysis was made of total local spending by each authority against various factors of need, such as miles of road and numbers of children. The results were used to show, for example, the relationship between the number of children and total spending, and this relationship was used when assessing the impact of children on the then equivalent of GRE. A problem was that if a group of authorities with a similar characteristic, such as a high population density, set out to raise their service levels, then they would simultaneously raise their spending. The regression analysis for the next year would then show that spending and apparently therefore needs were higher in high-density areas than previously imagined, and those areas would subsequently have higher needs indices and attract higher grants. It was understandably felt to be unsatisfactory for a group of authorities to be able to raise their needs indices solely by raising their spending levels.

Summary

Equalisation schemes seek to reduce the extent to which some subcentral authorities might be able to offer their inhabitants more attractive fiscal packages than others. There are three reasons why these packages might vary. First, some areas might have higher per capita tax bases and so be able to raise equal per capita sums at lower tax rates. Secondly, some areas might need more units of public services per head than others to provide similar standards of services. Thirdly, the costs of units of services might be higher in some areas than others. All equalisation schemes seem to address the first of these problems and some address the second, but few address the third.

A pure equalisation scheme would seek to put all authorities in a similar

position. This position would be that of an authority with average resources (and needs and costs if these are considered). Grants to below-average areas would be financed by negative grants or precepts on the others, so that the net cost to the grantor would be zero. In practice, few schemes are as pure as this. Some schemes do not attempt full equalisation, but merely seek to reduce the differences between the areas. Many schemes bring disadvantaged areas up to the average, or beyond, but either leave the other areas alone or give them small grants. These arrangements mean that the grantor incurs a substantial net cost, and suggest that it is undertaking revenue-sharing as well as equalisation.

In the past, most attention has focused on two very different grant schemes. In one, effort-related grants are used, and payments are proportional to grantees' tax efforts. This scheme ensures that any two areas with a common tax rate can provide similar service levels, but it alters the marginal cost of services to grantee citizens and can lead to overprovision. This drawback may not be felt to be very serious if the grants have in part a revenue-sharing purpose, for such grants are generally intended to raise aggregate grantee spending above the level that would otherwise emerge. In the other scheme, lump-sum grants are used. With this scheme, there is no incentive for overspending, but two areas with a common tax rate will not be able to secure a common level of services unless they happen to choose one particular tax rate specified by the grantor. A new scheme of effort-related grants is put forward here in which a grantee's grant receipts fall in relation to its revenue as its tax rate rises. It is suggested that this scheme retains the advantage of the earlier effort-related scheme while mitigating its disadvantages.

All equalisation schemes need to measure grantees' resources. In practice, resources may be measured in relation to the grantees' per capita tax bases or their per capita income levels. It is argued that it is necessary to use the former approach in order to make possible equal treatment for equals in different areas.

Schemes that take varying needs (and costs) into account have to estimate the relative per capita spending requirements in each area. Implicitly, the schemes concerned always seem to estimate the relative requirements of a given grantee by comparing the expenditure that would be needed to provide a standard package of services there with the expenditure that would be needed in an area with standard needs (and costs). On this approach, the relative importance of each service in the package is determined by the grantor. It is shown that this can lead to unequal treatment of equals in different areas. It is suggested that the grantor should take note of the varying preferences of each grantee and estimate the relative spending requirements of a given grantee by comparing its own total spending with the spending that would be needed to provide that grantee's chosen package of services in an area of standard needs (and costs).

Effort-related grant schemes need measures of revenue effort. Where a single tax is in use, the tax rate is clearly a satisfactory measure. Where more than one tax is in use, it would be possible to have a separate scheme for each tax, but it is shown that this could lead to each area choosing

to rely wholly or largely on the tax with which it is least well endowed, and wide inter-area variations in tax rates could emerge. Instead, it is suggested that there should be a single scheme in which a grantee's tax effort is defined as its total revenue divided by the amount it would raise if it set standard tax rates determined by the grantor.

One feature of effort-related schemes in practice is that the final formulae are generally settled after the grantee tax rates have been fixed in order that the total payment equals a sum predetermined by the grantor. This secures certainty for the grantor at the cost of uncertainty for the grantees, who must fix their tax rates before they know exactly how much grant these rates will attract. It is suggested that it might be better to fix the formulae first to secure certainty for the grantees. The grantor might end up in a given year with a total payment higher or lower than it would prefer, but it should be able to adjust the formulae from year to year to keep the overall descrepancy over a period of years small.

Notes

1 For discussions of equalisation grant design see King (1980b), Mathews (1977), and Musgrave (1961a).

2 See Musgrave (1961a, pp. 109–11). Musgrave's plan 6, as summarised by his equation 6.2, is equivalent to the scheme in the text if (i) cost indices are ignored (they appear to be included in the needs index as defined by Musgrave) and (ii) the last term of Musgrave's formula (designed to secure a specific aggregate grant for all areas taken together) is ignored. For a discussion of this equivalence, see King (1980b, p. 61, n1).

3 I am grateful to Professor R. L. Mathews for raising this point.

4 See Musgrave (1961a, pp. 102–3). His equation 3.4 is equivalent to that given in the text for this scheme if the unit cost index is ignored. Note that an area's grant receipt will not be wholly independent of its tax rate if the standard rate is an average rate.

5 The grant is given by $t(\bar{T}CN - T)$ and total expenditure by this plus tax revenue (tT), that is by $t(\bar{T}CN)$. Hence the grant meets the fraction $(\bar{T}CN - T)/\bar{T}CN$ of a recipient's expenditure, provided expenditure financed from sources other than this grant and the tax is ignored.

6 See King (1980b, pp. 32–7, 39–42, 47 and 63–4). Some elements of this particular scheme are to be found in a scheme put forward independently by Reischauer and Hartman (1973, p. 85).

7 On the normal approach, N for any area, say B, is found from $N = \Sigma \bar{E}/\Sigma \bar{S}$, where $\Sigma \bar{E}$ is the expenditure needed in B (assuming it faces standard unit costs) to provide subcentral services at *standard levels*, and $\Sigma \bar{S}$ is the expenditure needed in an area of standard needs to provide the same standard levels. King suggests using $N = \Sigma E/\Sigma S$, where ΣE is the expenditure needed in B to provide services at *B's own levels*, and ΣS is the expenditure needed in an area of standard needs to provide the same service levels as those provided in B. If variation in output unit costs were to be allowed for, then the index CN could be defined in the same ways, though $\Sigma \bar{E}$ and ΣE should reflect the actual output unit costs in B.

8 Some writers have concluded that this problem could be solved only if areas had to maintain fixed relationships between their various tax rates. See, for example, Cripps and Godley (1976, pp. 33–4). The present text shows that there is another less restrictive solution.

9 Nathan *et al.* (1975, pp. 45–6n) have explained that the measure of resources used in the United States revenue-sharing scheme is related to personal income per head.

10 This was reported in *The Times* for 23 March 1971.

11 For a discussion of the equalisation arrangements in West Germany, see Hunter (1973, especially Chapters III and IV) and Traber (1980).

12 For a discussion of the relevant federal–provincial grant arrangements in Canada, see Boadway and Flatters (1982, pp. 1–5 and 61) and Burns (1977, pp. 54–66); for a description of the provincial–municipal arrangements, see Auld and Eden (1983).

13 For a discussion of the federal–state arrangements in Australia, see Mathews (in ACIR, 1981c, Chapter 3).

14 For an outline of the revenue-sharing arrangements in Switzerland, see Bieri (1979, pp. 53–67).

15 For a summary of the revenue-sharing arrangements in the United States, see Nathan *et al.* (1975, especially pp. 45–6) and Break (1980, pp. 142–68, especially pp. 147–8). For a description of state–local authority grants, see Maxwell and Aronson (1977, Chapter 4) and ACIR (1980).

16 For a discussion of the new arrangements in the United Kingdom, see Bennett (1982, pp. 103–30), Jackman (1981), and Crispin (1980, pp. 27–31).

6

Taxes

Introduction

In this chapter, and the two that follow, it is supposed that a set of functions is to be entrusted to subcentral authorities (perhaps for the reasons considered in Chapter 1) and that a system of subcentral authorities has been created (their sizes and boundaries being determined, perhaps, by the factors considered in Chapter 2). Further, it is supposed that the authorities are in receipt of grants to encourage them to take externalities into account when deciding on their service levels, and, perhaps, some grants to enforce grantor preferences.

The question then arises of how they should finance the rest of their expenditure requirements. There are essentially five possible sources – namely, revenue-sharing grants, some form of tax-sharing, taxes (probably accompanied by a system of equalisation grants), charges and loans. It was argued in Chapter 4 that one major problem with revenue-sharing grants, and with tax-sharing arrangements unless they are constitutionally guaranteed, is that it is unlikely that the grantor, presumed to be a higher tier of government, would make the payments concerned without laying down conditions about how the money should be spent; indeed, it is questionable whether the grantor should pay them without conditions, since it would be responsible to its own taxpayers for the way in which their taxes were spent. Such conditions might cover the proportions of payments to be spent on different services by grantees, and the type or quality of each service to be provided. How much importance will be attached to these conditions in any particular case will naturally depend on their stringency and how far the preferences about services patterns in the several grantee areas differ from those of the grantor, but it is likely that sufficient importance will be attached to them for there to be a desire that these two revenue sources are not used to the total exclusion of the others. Another reason for not relying wholly on these two sources is that with them alone authorities would have no means of altering their total spending levels; although, of course, it should be stressed that the degree of subcentral control over spending levels depends not only on the extent to which authorities can alter their tax, charge or borrowing levels, but also on the extent to which revenue-sharing and equalisation grant receipts vary with tax effort.

It seems reasonable to assume, therefore, that the question of giving subcentral authorities the powers to raise taxes, to levy charges and to borrow will be considered. These three sources are considered, in turn,

in three separate chapters, and it will be seen that different issues arise in each case.

This chapter is concerned with taxes, and has three main purposes. First, it discusses the case for having similar taxes in each authority of any tier of government. Secondly, it examines a number of criteria by which taxes may be judged to see how suitable they are for use by subcentral authorities; on each criterion the performance of each of a number of possible taxes is discussed. Not surprisingly, perhaps, the discussion of these criteria suggests that no tax appears to be perfect for subcentral use. Thirdly, the chapter considers the various taxes in use by subcentral authorities in a number of developed economies and looks particularly closely at the tax-sharing scheme in use in West Germany. This scheme is of special interest because (as shown in Chapter 5) it enables the *Länder* to finance very high proportions of their expenditures from tax revenue, despite the difficulty of finding good subcentral taxes; it must be said, though, that the scheme gives the *Länder* no individual control over the level of their tax receipts.

As it happens, most of this chapter is concerned with the second of these three purposes. Altogether, some twelve criteria are examined. Four of these criteria relate to the yield of each of the taxes: the size of the yield; the ease with which it can be decided which authority is entitled to the yield from any particular taxpayer; the evenness of the distribution of the potential yield between the various authorities in any tier; and the elasticity of the yield, although some doubt is cast on the importance of this criterion. Two of the criteria relate to the burdens the taxes may impose in addition to the yields they exact: the level of administration and compliance costs; and the degree of 'excess burdens' or distortions they may impose on economic activity, especially if tax rates vary between areas. A further four criteria relate to how far the taxes assist in causing optimal levels of services to be provided: the degree to which taxes can be exported from residents on to non-residents; the degree to which tax payments by electors are perceptible to them; the degree to which authorities may be tempted to have suboptimal service levels in order to set low tax rates and thus attempt to attract industry; and the degree to which the incidence of the taxes is spread widely among an authority's citizens, in particular the extent to which it will fall on median voters. The next criterion concerns the extent to which taxes may impinge on the central government's macroeconomic policies. The final criterion turns from efficiency considerations to equity ones, and considers in particular the extent to which the various taxes are likely to escape criticism on equity grounds; the discussion of this issue is necessarily somewhat subjective. These criteria are considered in separate sections.

It cannot be claimed that this list of criteria is exhaustive. Other criteria that have been put forward for consideration by the United Kingdom government (HMSO, 1981b, pp. 6–7) include the points that a good local tax should have a yield that could be predicted with reasonable precision, to avoid the need for large balances, and that it should not have a lumpy yield (that is, one where the minimum practicable change in the tax rate produces a large variation in tax revenue). As it happens, it was argued

(pp. 24 and 35) that local sales and personal income taxes did badly on each of these points. It seems unlikely, however, that factors such as these would weigh very heavily on any final decision to allow subcentral authorities to levy a particular tax, and this chapter concentrates on what are probably more important matters.

It is arguable that some of these twelve criteria are concerned more with the question of whether a tax is a good or satisfactory tax at all than with the question of whether or not it has any specific drawbacks that make it unsuitable for use by subcentral authorities. For example, it might be felt that the inelasticity of a poll tax, the excess burden of an income tax that affected the supply of labour, the imperceptibility to voters of a tax on natural resources used by industry, the possibility that median voters might pay no tobacco tax, and the various equity criticisms made against taxes on domestic property, would apply whether these taxes were levied by central or subcentral authorities. In turn, it might be argued that these issues need not be discussed in a work concerned with the issues that arise specifically in the context of a system of subcentral authorities. However, there seem to be two advantages in adopting a comprehensive list of criteria. First, it seems legitimate to suppose that if a tax is to be a good subcentral tax, then it should also be a good tax *per se*, and by adopting a broad approach this chapter covers more of the arguments that are raised when the choice of subcentral taxes is under discussion. Secondly, and perhaps more significantly, it is possible that the disadvantages of this sort that apply to some taxes may be more acceptable when they are levied centrally than when levied subcentrally. This possibility can arise because central governments typically levy a wide array of taxes whereas subcentral ones are often permitted to levy but a few, and, in many cases, just one. The drawbacks of a tax with low elasticity, or one that it was felt impinged unfairly on certain groups, might seem much greater to a subcentral authority that levied that tax alone than they would to a central government with many other taxes. The possible adverse effects of an income tax on labour supply might be more worrying if the tax was the sole one levied by subcentral authorities than if it was one of many levied by a central government; in the former case, increased subcentral spending would necessitate higher income tax rates, while in the latter case increased central spending could be financed from other taxes. In a similar way, it might be felt that an imperceptible tax, or one escaped by median voters, would be more worrying if it was one of a few taxes, or perhaps the only one, levied subcentrally, than if it was one of many levied centrally; in the former case, there would be little to discourage median voters from demanding vastly super-optimal service levels, while in the latter case such demands might perhaps be curtailed in the fear that other taxes would also be raised.

The case for a harmonised structure of subcentral taxes

Most of this chapter concerns the various criteria by which taxes may be judged in order to evaluate their suitability for use by subcentral government. Now it would be possible for a central government to use

an evaluation of possible taxes in two different ways. On the one hand, it could draw up a list of tolerable subcentral taxes, and allow each authority to decide which tax or taxes from this list to levy. On the other hand, it could attempt to pick out, say, one or two taxes that it considered best and require each authority to levy it or those alone; where more than one tax was involved, it could go further and regulate the degree to which authorities could choose to rely on one rather than another. The case for adopting the latter approach is considered in this section. It will be appreciated that this approach amounts, at most, to a harmonisation of tax structures in the different areas. It does not amount to full tax harmonisation, which would require identical tax rates in each area, thereby removing one means (and often the most effective one) whereby different areas could secure different budget levels.

In examining the case for harmonising tax structures, it will be necessary first to examine the problems that might arise without such harmonisation. These problems differ according to whether or not an equalisation scheme is in operation. It might be thought that no equalisation scheme could operate when authorities levied different taxes, but this is not so. Chapter 5 showed how equalisation schemes could be developed to allow for several subcentral taxes. If the authorities were allowed to choose to levy one or more taxes from a list of permitted taxes, then each could merely be supposed to set zero rates on the ones not selected. It will be recalled from Chapter 5 that an authority's grant receipts in such a scheme might depend on the ratio of its actual tax yield to the yield it would have with a set of standard tax rates determined by the grantor.

Table 6.1

Citizen	*Area A Income*	*Expenditure*	*Citizen*	*Area B Income*	*Expenditure*
$\alpha.1$	1,000	900	$\beta.1$	1,000	900
$\alpha.2$	2,000	1,400	$\beta.2$	2,000	1,400
$\alpha.3$	3,000	2,100	$\beta.3$	3,000	2,100
$\alpha.4$	4,000	3,600	$\beta.4$	4,000	3,600

The chief effects of an unharmonised structure can be explained with the help of some examples.[1] In each example it will be supposed that there are two areas (A and B), each with four citizens with the annual income and expenditure levels shown in Table 6.1 (in an unspecified currency unit); it will be further supposed that richer citizens tend to have more valuable residences, although the total value of residential property in each area will be assumed to be equal. It will be supposed, too, that each area provides services whose total costs, at given levels of service, are proportional to population, that preferences are similar in the two areas and that similar patterns and levels of services (costing 200 currency units

per household) are indeed provided. Finally, it will be supposed that each area is free to select one tax from a list of two decreed by some higher-tier authority. The examples below consider the effects of this list containing each possible pair of the following four taxes: a poll tax, an income tax on residents, a general sales tax and a property tax on residential property (taken to be owner-occupied).

Table 6.2

Citizen	*Area A Income*	*Poll Tax*	*Citizen*	*Area B Income*	*Income Tax*
$\alpha.1$	1,000	200	ß.1	1,000	80
$\alpha.2$	2,000	200	ß.2	2,000	160
$\alpha.3$	3,000	200	ß.3	3,000	240
$\alpha.4$	4,000	200	ß.4	4,000	320

Consider, first, the case where A and B can select a poll tax or an income tax, and suppose that A chooses the former and B the latter. It may be imagined that a major reason for B's choice is an intention that citizens with below-average incomes (ß.1 and ß.2) should pay less than 200 in tax while those with above-average incomes should pay more. In order for B to secure a total tax yield of 200 per citizen, it will need an income tax rate of 8 per cent, while A will exact a poll tax of 200 per head. The initial tax payments by each citizen will be as shown in Table 6.2. It will be seen that there will be horizontal inequity because the rich in A will pay less tax than the rich in B while the poor in A will pay more than the poor in B. This horizontal inequity will give an incentive for the rich in B to move to A and the poor in A to move to B. Such migration uses up resources and is likely to mean that labour relocates in areas where its marginal product is lower than before (assuming labour was, initially, generally located where its marginal product was the highest attainable) thereby causing a fall in output. As a result of the migration, the situation in each area might become like that shown in Table 6.3, where B needs a tax rate of 13⅓ per cent to maintain its total tax yield; the table ignores any fall in wage income that migrants might suffer. It will be seen that B has at least partly failed in its objective of securing low tax payments for poor citizens: citizens with incomes of 2,000 in B pay more than 200 and, indeed, more than the rich citizens in A. The situation for B's poor would worsen if $\alpha.2$ decided to return to A (as would now be in his interest) and ß.2 decided to join him, although what might be termed second-stage moves of this sort would be deterred by the fact that they would cause A's population to rise, and hence cause property prices there to rise, while the population and property prices in B would fall. Even the prospect of the first-stage moves reflected in Table 6.3 could cause B to decide not to select an income tax, for such moves would largely frustrate its attempt

Table 6.3

Citizen	*Area A Income*	*Poll Tax*	*Citizen*	*Area B Income*	*Income Tax*
$\alpha.3$	3,000	200	$\alpha.1$	1,000	133
$\alpha.4$	4,000	200	$\alpha.2$	2,000	267
ß.3	3,000	200	ß.1	1,000	133
ß.4	4,000	200	ß.2	2,000	267

to attain a more desirable incidence of taxes, although its decision has benefited the very poor. It follows that one effect of the unharmonised structure in this case is that it might not be feasible for one area, B, to achieve fully its objective of securing tax payments under 200 for citizens of below average incomes. This feasibility problem would not arise if there was an equalisation scheme, however; in that case B would be compensated for its shrinking base on one of the two available taxes. It is possible, of course, that citizens in other areas would be reluctant to pay equalisation grants to an area such as B that chose a tax that systematically reduced its tax base, although the influx of rich citizens into those other areas would enhance their ability to pay such grants.

Secondly, suppose that the areas can choose a poll tax or a general sales tax, and suppose that A selects the former while B, for distributional reasons, chooses the latter and sets a rate of 10 per cent to secure a total yield of 800. Suppose, too, that the areas are too big for shopping expeditions between them to be frequent. The results will be very similar to those of the first tax. There will be horizontal inequity because the high spenders in A will pay less tax than their equals in B (for instance, $\alpha.4$ will pay 200, whereas ß.4 will pay 360) and the low spenders in A will pay more than their equals in B (for instance, $\alpha.2$ will pay 200, while ß.2 will pay 140). Thus migration of households will be encouraged, consuming resources and possibly causing a fall in output. Further, B will find it necessary to raise its tax rate, and may decide its choice of tax is unfeasible, unless there is an equalisation scheme. If the areas were small, then shopping expeditions might replace the migration; indeed, all citizens in B would have an incentive to shop in A.[2] These expeditions, like the migration of households, would consume resources, but, unlike that migration, they would not affect total output. In such circumstances, B would very likely feel its tax was unfeasible unless there was an equalisation scheme.

Thirdly, suppose that the areas can choose a poll tax or a property tax, and suppose that A selects the former while B, for distributional reasons, chooses the latter. The effects would be very similar to those in the first two cases. There would be horizontal inequity because citizens with large

houses in A would pay less tax than their equals in B, while citizens with small houses in A would pay more tax than their equals in B. Once again, migration would be encouraged, with possible adverse effects on output and with a consumption of resources, and B might feel its choice of tax was unfeasible unless there was an equalisation scheme. One possible amelioration has been suggested here, and arises from capitalisation.[3] Initially, the price of large rich people's houses may well fall in B (and rise in A) and the price of small poor people's houses may fall in A (and rise in B), thereby discouraging moves. Such an expansion of the range of prices in A and contraction in B cannot persist, however. In the long run, the composition of properties will react: large houses in B will not be replaced, or will be split into small houses, and small houses in A will not be replaced, or will be knocked together to form big houses.

Table 6.4

	Area A			*Area B*	
Citizen	*Income*	*Tax*	*Citizen*	*Expenditure*	*Tax*
$\alpha.1$	1,000	80	ß.1	900	90
$\alpha.2$	2,000	160	ß.2	1,400	140
$\alpha.3$	3,000	240	ß.3	2,100	210
$\alpha.4$	4,000	320	ß.4	3,600	360

Fourthly, suppose that the areas can choose between an income tax or a general sales tax, and suppose that A selects the former and B the latter, perhaps believing that expenditure is a better tax base than income. Suppose, too, that the areas are too big for frequent shopping expeditions. In this case, A will start with a tax rate of 8 per cent while B will start with one of 10 per cent, and the initial tax payments by each citizen will be as shown in Table 6.4. There will be horizontal inequity again, since no citizen in A pays quite the same as his equal in B. This could induce migration, which would consume resources and would be likely to have an adverse effect on output. Citizens with high average propensities to consume (these are 0.9 for $\alpha.1$, ß.1, $\alpha.4$ and ß.4) would prefer to live in A, while those with low propensities (these are 0.7 for the other citizens) would prefer to live in B. The situation in each area might become like that shown in Table 6.5 (which ignores any fall in wage income that migrants might suffer), where B needs a tax rate just under 11½ per cent, whereas A is able to persist with a rate of 8 per cent, although with other numbers a slight rise or fall might have been necessitated. The situation is not quite analogous to the three previous cases, however. The rise in B's tax rate has been modest and it may well consider its choice feasible; moreover, there is no incentive for second-stage moves. Essentially, these results arise because it is assumed, plausibly, that there is a much smaller difference between the impacts of these two taxes on individual citizens

Table 6.5

Citizen	*Area A Income*	*Tax*	*Citizen*	*Area B Expenditure*	*Tax*
$\alpha.1$	1,000	80	$\alpha.2$	1,400	160
$\beta.1$	1,000	80	$\beta.2$	1,400	160
$\alpha.4$	4,000	320	$\alpha.3$	2,100	240
$\beta.4$	4,000	320	$\beta.3$	2,100	240

than there is between the impact of a poll tax and any of the others. It should be pointed out that feasibility will again be aided by an equalisation scheme that would compensate B for its shrinking sales tax base. Of course, if the areas were small, then the migration of households would be replaced by shopping expeditions, which would consume resources but not reduce output. In this situation, all B's citizens would be tempted to shop in A, and its tax choice might well be unfeasible unless there was an equalisation scheme.

It will be readily appreciated that the results in a fifth example, where A might choose an income tax and B a property tax, or a sixth example, where A might choose a general sales tax and B a property tax, would be very similar to those of the fourth example. In general, there would be some horizontal inequity and an incentive to migrate, causing resources to be used and making a fall in output likely but not causing feasibility problems unless, in the sixth example, the areas were small enough for shopping expeditions and there was no equalisation scheme.

Similar conclusions would apply if other taxes were considered. For example, if area A were to introduce a tax on business incomes while B did not, then it would be likely to encourage capital to leave for B, and perhaps (in the absence of other taxes) encourage labour to come in, possibly even for lower wage rates. These moves would consume resources. More importantly, the marginal product of labour in A would fall, relative to B, and the marginal product of capital would rise. These differentials mean that output would fall, for it would be possible to raise output again simply by shifting labour back to B and capital to A. Again, if A introduced a payroll tax on incomes earned in its area while B did not, then all workers in A might have an incentive to seek employment in B, even though this might be at lower wages; and if people worked in new occupations where their marginal products were lower, then output would fall. In each of these two cases, some horizontal inequity would be inevitable and there could be a feasibility problem unless there was an equalisation scheme.

There is likewise a case for harmonisation if the areas choose more than one tax. This means not only that the areas should, ideally, use the same taxes, but also that they should set similar relative rates. If, for instance,

A set a high rate on a poll tax and a low-rate income tax while B set a low rate on a poll tax and a high-rate income tax, then problems similar to those considered in the first example above would appear. Widely disparate tax rates might well arise if each authority sought to rely largely, if not exclusively, on the tax that, for a given total yield, impinged less on its median voter, and if the situation for median voters varied from area to area.

The situations just considered indicate that the absence of harmonisation means that there could well be horizontal inequity, and that the induced migration of households could consume resources, reduce output and, eventually, make some areas' choices of taxes unfeasible or untenable unless there was an equalisation scheme to compensate them for a shrinking tax base. All of these factors constitute a case for harmonisation. Nevertheless, it might be argued that the choice of taxes was something about which preferences might vary from area to area, and hence suggested that there would be gains from a lack of harmonisation as citizens chose the taxes they preferred, just as there are gains (discussed in Chapter 1) when citizens in each area choose their patterns of services; these gains need to be weighed against the losses. It is perhaps worth contrasting the possibility of subcentral control over the choice of taxes with the possibility of subcentral control over redistribution and allocation policies.

As it happens, different redistribution policies in different areas would have very much the same consequences as different tax choices. There would be horizontal inequity because the rich in the most progressive areas would have higher taxes and lower benefits than the rich elsewhere, while the poor would have lower taxes and higher benefits. Migration would be induced, consuming resources and perhaps reducing output and threatening the tenability of the policies in the most progressive areas. It will be recalled from Chapter 1 that the problem focused on in the literature is the question of tenability. This could be eased by an equalisation scheme, but taxpayers in general might well be unwilling to give ever-increasing grants to any areas that systematically reduced their tax bases. Arguably, then, the main objection to areas choosing their own tax bases is that some choices might be untenable without an equalisation scheme, and such schemes might well lose support if some areas systematically cut their bases.

It might seem surprising that those opposing subcentral control of redistribution policies have not concentrated more on the facts that migration would use up resources and could reduce output. Perhaps the answer is that it is assumed that the bulk of those costs will be borne by the migrants themselves (for they will pay their carriers and may take falls in wages if their marginal products fall), who presumably consider that the costs are outweighed by the gains. Arguably, a similar view should be taken of the same effects if they result from different tax choices. Of course, if migration does have external costs and benefits, as explained in Chapter 1, then it might be worrying unless these effects were internalised in some way.

The hypothesis that the key objection to devolving the choice of taxes is the possibility that some choices may be untenable might appear to be supported by the fact that all the other arguments against it could result from devolving resource allocation decisions. For example, even if two

areas levied the same taxes at the same rates, the net fiscal treatment of equals in each area could differ markedly if the patterns of services provided were not the same; migration could result, consuming resources and perhaps reducing output. It will be recalled from Chapter 1 that such migration is often seen as a distinct argument in favour of subcentral control of allocation policies since it enables people to vote with their feet.

This last point, however, hints at a fundamental problem with the migration caused by unharmonised tax structures. It will be recalled from Chapter 1 that the process of voting-with-the-feet could, under certain assumptions, result in citizens being optimally located. In particular, if authorities' service costs, at any given level, were proportional to population and if the services were financed by benefit taxes, so that each citizen paid in taxes the entire cost of providing him with services, then each would be tempted to move to areas with higher service levels until the extra tax cost of so doing (which would precisely equal the extra resource costs involved) equalled the marginal gain to him. In this way, citizens would end up optimally located in the sense that no one's welfare from subcentral services could be increased without making someone else worse off (perhaps by increasing tax rates in one area to raise service levels in another).

This process cannot work efficiently in the absence of benefit taxation, yet such taxes are never used; a later section of this chapter examines closely the inefficiencies that would arise with a harmonised structure of non-benefit taxes, but clearly inefficiencies would also arise with an unharmonised structure of non-benefit taxes. It is hard to compare two situations neither of which is efficient. However, the harmonised tax situation has the property that any citizen desiring to move in order to pay less in taxes will necessarily receive lower service levels, whereas in an unharmonised tax situation it would be possible (as in the examples of this section) to move to another area and pay less in taxes and yet not face lower service levels; in the real world, it might be possible to move to another area and pay less in taxes and yet enjoy higher service levels. It is arguable, if not demonstrable, that this means that unharmonised tax structures, and the migration caused by them, result in less efficient locations than would result from harmonised structures. This certainly seems to be a major reason why Oates (1972, pp. 145–7) favoured harmonised structures.

By way of contrast, it has been suggested by Foster *et al.* (1980, pp. 247, 505 and 545) that the main argument in favour of harmonising the relative rates at which different taxes are levied in different areas, and thus for harmonising tax structures, is to secure horizontal equity for its own sake. It is not clear, however, that horizontal inequity arising from unharmonized tax structures need cause any more concern than horizontal inequity resulting from different redistribution and allocation policies. It should be stressed that the horizontal inequity considered here arises from the conscious democratic and voluntary choices of A and B to do things differently. It is very different from the horizontal inequity that would arise if A were very much richer than B. In that case, even if A and B financed similar patterns of services from an identical single tax, there

would be horizontal inequity because tax rates in B would be higher than in A. It is this involuntary inequity that equalisation schemes seek to remove, as explained in Chapter 4.

It follows from this discussion that the principal objections to unharmonised tax structures are that the induced migration of trade, labour or capital could make some areas' choices of taxes unfeasible, while the migration of households could aggravate any tendency for citizens to locate non-optimally. Thus the strength of the case for a harmonised structure depends on how much migration might occur. Migration of households will in general be greatest when the impacts of the different taxes on equals are markedly different; for instance, it will probably be more problematic to have a tax loosely related to ability to pay in one area and a poll tax in another than to have different taxes loosely related to ability to pay (such as a personal income tax and a general sales tax). Migration of trade will be greatest if small areas are permitted to operate taxes on expenditure, and migration of labour will be greatest if small areas are permitted to operate payroll taxes; in these two cases, improvements in transport that facilitate commuting to work or shopping in other areas will aggravate the problems. Migration of capital will be greatest if areas are permitted to tax businesses, particularly perhaps with taxes on business incomes. Overall, the problems with small areas might be least if their choice was confined to personal income taxes or property taxes; for large areas, sales taxes and payroll taxes might also be acceptable.

It is worth concluding this section by noting an argument suggested by Lane (1974, pp. 132–6) for a rather different type of harmonisation from that considered above. Instead of pleading that all authorities in a given tier should levy similar taxes at rates in similar ratios, while implicitly permitting very different tax structures between tiers, this argument allows the mix of taxes selected by each authority in each tier to be very different provided that each authority is required to include one particular tax, such as a personal income tax. The authorities would be free, however, to choose how far they relied on this in relation to their other taxes, provided they set a positive rate. In this situation, electors would be enabled to compare the marginal costs of extra spending by each tier since each authority is likely to alter its rate of the common tax from time to time. This, it is claimed, will aid electors to form in greater knowledge their preferences on the balance of spending between tiers.

Taxes and the sizes of their yields

In any list of criteria for suitable taxes for subcentral authorities, the question of yield is likely to feature prominently.[4] These taxes, along with equalisation and revenue-sharing grants, are generally expected to finance that part of current spending not covered by charges or other grants, and may be expected to cover some capital spending as well. Thus in many countries the target yield for taxes along with tax-sharing and revenue-sharing grants exceeds 10 per cent of gross domestic product, and in some it exceeds 20 per cent. If the taxes are to cover a significant proportion

of this total, then they will have to include one or more of those taxes whose potential yield is high in relation to gross domestic product. In practice, the most common taxes concerned include: a tax on all personal incomes (paid to the area of residence); a payroll tax (a term that will be used here to cover any tax on earned incomes paid to the area of workplace, irrespective of whether the legal liability for payment lies with employers or employees); a corporation tax on incorporated businesses' profits; a value added tax;[5] a property tax on land or buildings; and a general retail sales tax. There are also some sales taxes on specific items where demand is sufficiently inelastic and high for high yields to be obtained, namely: a tax on motor fuel; a tax on alcoholic drink; and a tax on tobacco. In addition, there may be taxes on natural resources; these can take a wide variety of forms but will be assumed in this chapter to be based effectively on the sales value of the extracted resources. Another tax that might, in principle, produce a high yield is a poll tax. This tax is generally confined to the theoretical literature, but is considered along with the other ten in this chapter, partly because of its interesting properties, and partly because the government in the United Kingdom recently suggested that a poll tax on adults would be a suitable local government tax, albeit if levied at the modest rate of £30 per year in 1981 prices (HMSO, 1981b, p. 39). Of these taxes, it is only the personal income tax, the value added tax and the general retail sales tax whose yields are likely to exceed 10 per cent of gross domestic product; indeed, they are often the only ones whose yields exceed half that amount.

In addition to the taxes listed in the last paragraph, there are others that are sometimes encountered in this context and that should be mentioned. For instance, there are taxes on estates, on inheritances, on gifts *inter vivos* and on wealth; the first of these was used until recently by most states in Australia, and all but the last are used in various parts of the United States. There are also licence duties for motor vehicles.[6] In practice, these other taxes never produce yields comparable with those often obtained from the first ten taxes listed in the last paragraph, and they could at best be of modest use.

Having thus defined the eleven main candidates, it remains to consider the other criteria by which subcentral taxes may be judged and to see how they fare. The next three sections are concerned with three further criteria that relate to the tax yields.

The allocation of tax yields to different authorities

First, there is the point made by Maynard and King (1972, p. 57) that it should be clear which authority is entitled to the yield exacted from each taxpayer whether the tax is based on that payer's income, spending, property, or (in the case of a poll tax) existence. It is probably the taxes on property and expenditure that do best by this criterion, since the tax will generally be payable to the authority where the property is situated or in which the premises of the seller of an item sold is located at the time of transfer. Some difficulty will arise with a tax on personal incomes and a poll tax because the obvious system of paying tax to the area of residence

runs into problems with taxpayers who have residences in two or more authorities; However, apportionment rules can generally be devised to leave few ambiguous cases. In a similar way, problems could arise with a payroll tax in connection with citizens who work in more than one area; again, apportionment rules would be needed. The tax whose apportionment is most difficult is undoubtedly the corporation tax. In any country, there will be many businesses that operate in many authorities, and if subcentral authorities are entitled to tax their profits then the problem of allocating profits between areas, discussed by Maxwell and Aronson (1977, p. 117), may often be acute, particularly with small authorities. For this reason alone, this tax would seldom be considered a good one for subcentral use, although there are other objections to it as will become clear in later sections.

The geographical distribution of tax yields

A second criterion related to yields concerns their geographical distributions. There are four reasons why it is sometimes argued that a good tax for subcentral authorities should have a tax base that is widely and evenly distributed throughout the country.[7] (It is perhaps a little ironic that one country at least, namely the United Kingdom, uses a property tax for local authorities even though research – King, 1971, Vol. I, pp. 259–62 – has suggested that virtually all alternative local taxes would have more evenly distributed tax bases.) These reasons apply when a scheme of equalisation and revenue-sharing grants is in operation.

First, the grant payments to individual areas will generally be more unequal with unevenly distributed tax bases; insofar as grant payments may have conditions attached to them, some authorities will have less freedom than others. This problem is less serious with more evenly distributed tax bases.

Secondly, an unevenly distributed tax base means that a scheme of equalisation grants could produce large negative grants for some areas; even if the scheme also had a revenue-sharing purpose, some areas might still face negative grants. Negative payments will be very unpopular, particularly if these payments are effort-related so that each high tax resource area will be faced with a larger negative grant if it increases its tax effort, and the grantor may avoid using them. It might do this by ignoring the high resource areas concerned, in which case not all areas will be covered by the scheme; as a result, there will be horizontal equity problems between the areas inside the scheme and the others, and also between all those outside the scheme, which will have varying resource levels. Alternatively, the grantor might try to avoid negative grants by reducing the tax powers of the recipients to the level where even those with the highest resources merit grant receipts; in this case, the aggregate level of grant payments will rise, which may in time increase the level of grantor control over recipients. With a more evenly distributed tax base, the authorities can, in aggregate, raise greater proportions of their total revenue needs without negative grants becoming appropriate.

Thirdly, more evenly distributed tax bases reduce the importance of equalisation grants. Since the construction of the formulae and the methods

used for defining needs, resources and tax effort are often the subject of controversy, it might be expected that the controversy would be less heated when equalisation grant receipts were smaller and where, presumably, their absolute levels were less sensitive to changes in the rules.

Fourthly, it should be realised that equalisation grants transfer money from high tax resource areas to low tax resource areas. Unless subcentral authorities rely on an income tax, this will not necessarily mean that transfers are always made from high-income areas to low-income areas. Chapter 5 explained in more detail why transfers from low-income areas to high-income areas might occur and why they were justified. They might nevertheless be unpopular, and they will tend to be less important with more evenly distributed tax bases that require only small transfers of money.

The relative merits of different taxes on this criterion will depend very much on the sizes of authorities, although a poll tax will of course always be evenly distributed in relation to population. If the areas are small, then a tax on personal incomes will not be evenly distributed because the tax base will be high in the most fashionable suburbs and low in slum areas. The base will, however, be more evenly distributed than that of a payroll tax, which will be greatest in areas with labour-intensive industries and office blocks, or a general sales tax and a value added tax, which will each be greatest in areas with major shopping centres, or a corporation tax, which will yield little in residential areas. Taxes on alcohol, tobacco or motor fuel may be more widely distributed among the areas than general sales taxes since their outlets are widely distributed, but these too are likely to be more uneven than an income tax. Property tax bases may be very uneven since some areas have much more non-domestic property in relation to population than others, and the general level of property values varies greatly between areas (it should be recalled from Chapter 5, though, that there is a case for ignoring variations of the latter sort for equalisation purposes). Taxes on natural resources may be the most uneven of all.

With large areas, taxes will generally become more evenly distributed. It is even possible that a general sales tax or value added tax will be more evenly distributed than a personal income tax or payroll tax, since the average propensity to consume is likely to be lowest in areas where incomes are highest. Taxes on motor fuel, alcohol and tobacco will probably be less evenly distributed: the yield of the first will depend on many factors, such as population sparsity, which can vary widely between regions, and the yield of the others will depend on spending patterns, which can also vary widely between regions. A property tax will probably be quite uneven, too, since the ratios of domestic to non-domestic property can vary greatly even between large areas, and average property values can vary significantly between regions. A corporation tax may also be relatively uneven, being greatest perhaps in areas with the greatest levels of business activity. Once again, though, natural resource taxes may be the most uneven.

The elasticity of tax yields

The final criterion relating to tax yields concerns their elasticity or

buoyancy. This term is not always sharply or consistently defined. Sometimes, for example, it is regarded as a characteristic of a tax whose yield tends to rise at least as fast as costs and incomes without the need for changes in the tax rate (see the Layfield Committee, HMSO, 1976, p. 9), which perhaps suggests that the yield should keep pace with inflation, so that constant tax rates would produce at least a constant yield. At other times, and perhaps more usually, it is regarded as referring to a tax whose yield tends to rise at least as fast as total disposable income without the need for changes in the tax rate (see the Layfield Committee, HMSO, 1976, p. 261), which suggests that the yield should grow automatically in line with real increases in income as well as inflation. It is therefore appropriate to consider how elastic the yields of various taxes are in times of inflation and in times of economic growth.

To see how responsive yields are with respect to price changes, consider, first, elasticity in a period of inflation when real incomes are not rising, and suppose that inflation-elasticity is a measure of the percentage change in tax revenue divided by the percentage change in prices (on the assumption of constant tax rates). If the structure of income tax rates is progressive, and tax bands (in nominal terms) as well as rates are held constant, then the real yield from a personal income tax will rise as income recipients find their growing nominal incomes take them into tax bands subject to higher rates of tax; accordingly, an income tax will be highly inflation-elastic. With a payroll tax, a corporation tax and *ad valorem* taxes (such as a value added tax or general sales tax), constant tax rates will mean that the nominal yields from the taxes should rise roughly in line with prices and the real yield will remain roughly constant; such taxes have unit inflation-elasticity. With a poll tax, or specific taxes on expenditure (such as those usually levied on motor fuel, alcohol and tobacco) or a property tax (if the base is not revalued), constant tax rates mean that the nominal yields will remain constant and the real yields will fall; such taxes are inflation-inelastic.

To see how responsive yields are with respect to real income changes, consider how the taxes fare in a period of rising real incomes and stable prices (so that money incomes rise at the same rate as real incomes), and suppose that income-elasticity is a measure of the percentage change in tax revenue divided by the percentage change in real incomes (on the assumption of constant tax rates). Once again, income taxpayers will be moving into higher tax bands and so they will pay larger proportions of their growing real incomes in tax; hence the tax yield rises more rapidly than real income and a personal income tax is seen to be highly elastic. With a payroll tax, a corporation tax, a value added tax and a general sales tax, constant rates should result in yields rising in line with income growth; the taxes will have unit elasticity. A similar result would hold for taxes on motor fuel, alcohol and tobacco if the relevant income elasticities of demand were unity so that real spending levels rose in line with the rise in real income; spending on those items may, however, fluctuate in proportion to income as time passes, and it is arguable that taxes whose yields at constant rates may be erratic should be levied by the central government as it raises many other taxes, which helps to reduce the impact of these variations on its total tax receipts. The elasticity of the yield for a tax on the sales of a natual resource will depend directly on

the income elasticity of demand for the products that use it as a raw material, and inversely on the extent to which it is used more efficiently. The yield from property taxes will rise in line with any increase in the property stock, so the stock will have to grow at least as fast as real incomes for the tax to have unit elasticity. In practice, the property stock growth rate is probably not unduly different from the real income growth rate so that property taxes probably have roughly unit elasticity, at least in the long run. Poll tax receipts, by contrast, will not rise without changes in tax rates, and thus this tax is inelastic once again.

The predictions of the last paragraph are given some empirical support by separate findings for Australia, Canada, the United Kingdom, the United States and West Germany within the period 1965–71.[8] These studies related tax yields to personal incomes in different states or regions. The elasticity for the personal income tax ranged from 1.44 in Australia to 1.96 in West Germany, and always exceeded that of the business income tax, which ranged from 0.86 in the United Kingdom to 1.50 in the United States. Similarly, the elasticities of general (*ad valorem*) sales taxes (not analysed on this basis in the United States) ranged from 0.58 in West Germany to 0.94 in Australia, and were generally higher than those for (usually specific) excise and import duties, which ranged from 0.36 in the United States to 0.58 in West Germany.

It appears that a progressive personal income tax is the most elastic in both respects. In times of inflation, specific taxes on expenditure, a property tax and a poll tax will lack elasticity, while the other taxes will be reasonably elastic. All taxes other than a poll tax should be reasonably elastic in times of real income growth, except for any tax relating to an item where the income elasticity of demand is less than unity. These results indicate that if elasticity in both respects is required, then a progressive income tax is most satisfactory, with a corporation tax, a payroll tax, a general sales tax and a value added tax all perhaps acceptable, along with *ad valorem* taxes on items for which the income elasticity of demand is greater than or equal to unity.

However, despite the frequent discussions of the relative elasticity of different taxes and, particularly in a subcentral government context, discussions of the lack of elasticity or buoyancy of the property tax with respect to prices, the actual merits or otherwise of such elasticity are seldom fully explored.[9] For elastic taxes to be preferred, some sound reasons must be given for preferring subcentral tax yields to rise automatically in times of inflation or real income growth.

The main reason usually given is that electors will more willingly accept yield increases when tax rates are not altered. This suggests that electors suffer from some form of fiscal illusion, for what really affects them is whether the yield rises, not how the rise comes about. The illusion may arise because electors think rises in tax rates of unbuoyant taxes in inflationary times raise the real value of payments when this is not so. There might be a case against taxes that lead to electors drawing incorrect inferences, but it is not very satisfactory to justify any policy on the grounds of human irrationality. Nevertheless, it was this source of illusion that led local authorities in the United Kingdom to oppose the property tax on the grounds of the lack of

buoyancy. Apparently, they felt the lack of buoyancy and rising tax rates in times of inflation created a false impression in the minds of the public and caused them to draw false conclusions about the burden of the tax (HMSO, 1976, p. 10). However, the illusion may arise for another reason, namely, that electors are more aware of the rise when tax rates rise; in this case, it is arguable that the existence of this illusion constitutes a case against buoyant taxes, for in a democracy there is clearly a case for ensuring that electors are aware of tax yield changes.

A further argument against buoyant taxes stems from the fact that the costs to electors of acquiring information about their yields are higher than the costs of acquiring information about the yields of unbuoyant taxes. The cost differences arise because the publicity associated with tax rate changes makes information more readily available in the latter case. It is argued that electors might in principle oppose increases in all yields, but might in practice oppose increases only in tax rates simply because that is where they are better informed. Of course, electors could acquire information about buoyant taxes, but might believe the costs of so doing would be disproportionate to the benefits that might be received. This argument suggests that electors might not make a fuss about the yields of buoyant taxes but would nevertheless prefer the yields to be raised less than they are. This line of reasoning claims that there is a case for unbuoyant taxes where the yields follow electors' wishes more closely.

This discussion suggests that there may be some reasons for preferring taxes that lack buoyancy. Of course, not everyone will prefer these taxes. For example, businesses may prefer elastic taxes on the grounds that planning is easier when tax rates change infrequently, and the central government may prefer them on the grounds that they are more help in countering cyclical fluctuations in the economy. Any bureaucrats or politicians anxious to maximise the sizes of their budgets are also likely to prefer buoyant taxes. In an interesting study, Oates (1975, pp. 145–52) found that in the United States between 1960 and 1970 expenditures by state governments grew faster in the states with personal income taxes than in the states without. The study was unable to show whether this was because the possession of an elastic source of revenue stimulated high expenditure or because states anxious for high expenditures opted for a highly elastic source of revenue.

Administration and compliance costs

All taxes remove purchasing power from taxpayers and leave them worse off by the amount of the yield. For this reason, the principal burden of raising any given sum will not depend on which tax is used, though, of course, the distribution of the burden will depend on the choice of tax. However, raising taxes imposes other burdens, and, *ceteris paribus*, there is clearly a case for using taxes that minimise these other burdens. These other burdens are often divided into two main groups: on the one hand there are the resource costs incurred by collectors and payers, and on the other hand there are the less obvious costs, known as 'excess burdens', that

arise out of any tendency of taxes to distort the economic behaviour of taxpayers. These two extra types of costs will be examined in turn.

The resource costs incurred by collectors (that is, governments) are generally known as 'administration' costs. These will include the costs of inspectors and their offices and the legal costs of dealing with defaulters. The resource costs incurred by taxpayers may be termed 'compliance' costs, and will include the time (and any distress) incurred in completing tax returns, any time spent in devising ways of reducing tax payments, and all costs involved in employing professional advisers.

There is little that can usefully be said about these costs, which inevitably depend greatly on the precise nature of any particular tax. It has been suggested, for example, that states in the United States spend around 1.5–2.0 per cent of sales tax receipts on administration and a modestly smaller proportion of income tax receipts.[10] In the United Kingdom in 1975–6, collection costs for the property tax amounted to some 2.5 per cent of receipts,[11] although there was no general revaluation in that year so that these costs were lower than they might have been; it has been suggested by the Layfield Committee that a local income tax there could be administered for half the cost of administering the property tax (HMSO, 1976, p. 278). These figures hint that the differences between the major types of tax may be modest, in absolute terms, as far as administration costs are concerned. In fact, the costs of collecting any one tax in any area will probably fall in relation to yield as the yield rises, so that one conclusion is that it might well be cheaper to have a few high-yielding taxes than a proliferation of low-yielding taxes. It is also possible that collection is subject to economies of scale attainable in bigger areas, in which case costs will fall if administration is centralised as far as is possible consistent with varying tax rates between areas being permitted. Indeed, it is possible that the administration costs for some taxes would be unavoidably much higher if they were levied by subcentral governments rather than by the central government, and they might on this criterion be judged unsuitable in the former context yet suitable in the latter.

Compliance costs are probably highest with a value added tax, although they might also be relatively high with personal income tax. It has been suggested that the compliance costs of the national value added tax in the United Kingdom exceed 9 per cent of the yield, and that the compliance costs of a local income tax might equal the administration costs.[12] Generally, though, it is unlikely that variations in administration and compliance costs would be large enough to offset any other relative merits there may be between the ten major candidate taxes for subcentral use.

Excess burdens of subcentral taxes levied at uniform tax rates

The excess burdens of different taxes are generally well explained, if not quantified, in the public finance literature,[13] although the analyses are ordinarily based on the assumption that the taxes are levied on a uniform basis by a central government. It is useful to summarise these standard results, even if they apply equally to subcentral taxes only if the authorities

levy the same taxes at the same rates. In the next section, the possibility of different rates in different areas will be considered to see if this results in any further distortions or excess burdens. The discussion will reflect the literature by explaining the reasons for the distortions rather than attempting to quantify them. It might be wondered if the goal of minimising these distortions is appropriate at all, given all the other distortions in the economy, but it has been argued by Brennan and McGuire (1975) that neutrality is the appropriate goal if the aggregate size and net direction of the other distortions are unknown.

The object of this discussion is to see if any of the main taxes considered as possibilities for use by subcentral authorities should be regarded with especial favour or disfavour in this context. It is perhaps useful, for the sake of completeness, to recall two types of excess burden that can arise from subcentral taxation, whatever taxes are used; these distortions, together with some ways of largely, if not wholly, avoiding them, have been considered in detail elsewhere.

First, there is the possibility that problems will arise if different areas use different taxes, or use the same taxes but seek very different ratios of tax rates. This issue was discussed in an earlier section, and it was seen that the degree of excess burden – reflected by the degree of distortion in the geographical distribution of residents, labour, trade or capital – would depend greatly on exactly how the tax patterns differed and on the mobility of the relevant items. It was also seen that these problems might be much reduced if different areas harmonised their choices of taxes and their tax rate ratios.[14]

Secondly, problems could arise even if each authority used the same tax, if per capita tax bases varied from area to area. In such a case, there would be a tendency for tax rates to be lower, or service levels to be higher, in areas where the per capita tax bases were higher. This could stimulate migration into those areas, and the migration would consume resources and possibly have adverse effects on output. As discussed in Chapter 4, one way of largely, if not wholly, removing this excess burden would be to introduce an equalisation scheme.

The excess burdens to be considered now can arise even when different areas harmonise their taxes and even if there is an effective equalisation scheme. Initially, it will be assumed that each area sets the same tax rate. In that case, the excess burdens arising from any particular tax will be the same as if the tax were levied centrally. Each of the main taxes under consideration will be taken in turn and its major distortions noted.

Consider, first, a tax on personal incomes. The part of the tax relating to earned income would tend to distort the choice between income and leisure unless the supply of labour happened to be perfectly inelastic; if the supply were not perfectly inelastic, then the tax would give an incentive to take less leisure or more depending on whether the supply curve was backward-bending or not. The part of the tax relating to unearned income would effectively reduce the net rate of interest available to lenders and would thus distort the choice between present and future consumption unless the supply of saving was perfectly interest-inelastic. A tax on personal incomes might also cause households to try to increase their incomes from

sources whose returns were not taxed, or were taxed at preferential rates; thus they might purchase a larger home if the imputed income from living in it was not taxed, and they might seek to switch their savings from securities yielding high incomes to ones promising high capital gains. These effects all depend on the marginal rates of tax paid by households rather than the average rates, since the former affect their decisions at the margin.

With a corporation tax, there are two main distorting incentives. First, more capital may in future be directed to the non-incorporated sector, and less to the incorporated sector, and this will affect relative prices. Secondly, the tendency for net rates of return to funds in industry to fall could divert future capital into other assets, such as foreign investments, and this might cause home prices in future to be higher than they would otherwise have been.

A payroll tax, like the tax on personal earned incomes, would create an excess burden by distorting the income–leisure choice unless the supply of labour was perfectly inelastic. It should be observed that, even though liability to pay the tax might lie formally with employers, the incidence would fall wholly (or largely) on employees if the supply of labour were perfectly (or very) inelastic.

If a general sales tax were imposed on all items purchased by consumers, then it would formally be equivalent to a tax on personal incomes in which saving was exempt. Like an income tax, therefore, it would distort the choice between income and leisure, unless the supply of labour were perfectly inelastic, but it would not distort the choice between present and future consumption. A general value added tax that exempted purchases of capital goods alone would be equivalent to a general sales tax; if capital goods were included, then it would combine the effects of a general sales tax with the effects of a tax on all profits, thereby acting as a disincentive to invest at home (as opposed to abroad) and perhaps causing future prices to be higher than they would be otherwise. It should be added that, in practice, general sales taxes and value added taxes often exempt many items such as food; in such cases, the most significant distortions caused will probably be in the pattern of consumer spending.

An excise tax on a specific item, such as motor fuel, alcohol or tobacco, would create an excess burden unless supply or demand was perfectly inelastic, because only then would consumption remain constant and the losses to those on whom the incidence falls (producers if supply is perfectly inelastic and consumers if demand is perfectly inelastic) precisely equal the tax yield. In any other case, the effect of the tax to give a disincentive to the consumption of the item concerned would be manifested, and there would be an excess burden resulting from the distortion in the pattern of consumption. Similar considerations would apply to many taxes on natural resources. If these taxes applied only to home producers, however, and if these producers were in competition with foreign producers and had to accept world prices, then they would not be able to pass any of the tax on to consumers. Accordingly, there would be no distortion in consumer spending, but those natural resources that were previously only just worth extracting might now cease to be worth extracting, and thus there could be distortions in the pattern of domestic production.

With a property tax, there would be no excess burden if the supply of buildings were perfectly inelastic, like the supply of land, because then the incidence of the tax on both land and buildings would fall wholly on their owners, who would not alter the amount supplied, with the result that no distortions or excess burdens emerged. However, the supply of buildings in a nation is not perfectly inelastic in the long run; those who supply them can react to the property taxes by investing in other assets or abroad. Thus a property tax on buildings does cause distortions and has an excess burden by reducing the supply of buildings. Conversely, a property tax on land is often held to cause no distortions and to have no excess burdens, and is for this reason widely regarded as a neutral tax. These conclusions on a land tax apply only if the tax is applied to all land; if, for example, agricultural land were exempt, then the rate at which such land was converted to other uses could well be reduced.

A poll tax should create no distortions and generate no excess burden. It follows that, from an excess burden point of view, a poll tax and a tax on all land are the best of those under consideration. In principle, it would be possible to estimate the excess burden of each of the other taxes, though in practice this would be extremely difficult, and their relative burdens might well vary from country to country and from time to time since they depend on factors such as the elasticity of supply of labour and the elasticity of demand for tobacco and alcohol. In the absence of such estimates, it is not possible to say if any of these other taxes do particularly well or badly on this criterion. So, if subcentral authorities wished to levy one or more common taxes at uniform tax rates, then the only useful clear conclusion from an analysis of excess burdens is that a poll tax or a tax on all land would be better than any other.

Excess burdens of subcentral taxes levied at varying rates

In practice, subcentral authorities will generally set varying tax rates. These variations can generate incentives for taxed subjects to migrate to areas where the tax rates are lowest. If such migration does occur, then the tax rate variations will have succeeded in causing distortions and excess burdens. In this context, some taxes are far more likely to cause problems than others. For instance, different rates of taxes levied on corporate profits and non-residential buildings will give incentives for business assets to move from high-tax areas to low areas, thus distorting the geographical pattern of production. A similar result would arise from different rates of taxes on payrolls because, in the long run, the supply of labour in any area could be quite elastic, thus throwing much of the incidence of this tax on to employers. One tax that could be levied on businesses at different rates without generating such distortions would be a tax on non-residential land, if its supply in each area was perfectly inelastic;[15] this tax would cause falls in the values of non-residential land, the falls being greatest where the tax rates were highest. Distortion could arise with a tax on non-residential land, however, if the base related to the actual annual return

on the land and the rate exceeded 100 per cent of that return; in that case, it would be better for landowners to keep land idle and secure a zero tax base.

Problems will arise with varying tax rates for a general sales tax, a value added tax or taxes on motor fuel, alcohol and tobacco, if these variations cause price variations and encourage shopping expeditions to low tax rate areas. Similarly, varying rates of taxes on natural resources might lead to price variations, in which case excess burdens would arise if purchasers in high-tax areas bought from further afield and incurred greater transport costs. Moreover, insofar as taxes on selected items are borne by their producers, producers will be encouraged to move from high-tax areas to low-tax areas. All these distortions constitute an important objection to the subcentral use of taxes on expenditure, especially by small authorities.

Variations in the rates of personal income taxes, poll taxes and taxes on residential land or buildings may encourage households to migrate. Such incentives will, of course, be tempered by the fact that low tax rate areas will, typically, be offering their citizens low levels of services also.[16] However, even if service levels varied directly with tax rates, there could still be an incentive for citizens to make inefficient location decisions, and hence the tax variations could still generate excess burdens. This problem is best explained with the help of an example, derived from Foster *et al.* (1980, pp. 220–32).

Suppose that a tier of authorities is established to provide a single service, and suppose that ten of those authorities – named A, B, C . . . J in ascending order of service level (L) – provide the service levels indicated by OA, OB, OC . . . OJ on Figures 6.1 and 6.2. Suppose that the total cost (TC) of the service in any area is proportional to its service level and, for the moment, that it is also proportional to population (P), so that $TC = kLP$ where k is a constant. In this case, the cost per person is given by kL and is shown for each authority in Figure 6.1 by the line $O\alpha$. Suppose further, for the moment, that each area levies a poll tax, so that $O\alpha$ also shows the tax paid by each individual in each area; in this case, $O\alpha$ shows the total tax cost a citizen would face in each area in which he might live. This total tax cost would rise with service levels as the citizen moved from low-level areas to high-level areas. However, the marginal cost to the citizen of any unit increase in service levels (as might occur with a change of level from OB to OC or OG to OH) is constant. This constant marginal cost situation is shown by the line bb in Figure 6.2.

Next consider two citizens (α and ß) who reckon their marginal benefit schedules from the service in question to be those shown by MB_α and $MB_ß$ in Figure 6.2, and assume that α and ß can move from area to area without incurring mobility costs. In this situation, it is efficient for α and ß to live in areas D and G respectively. If they were to live in areas with lower service levels, then it would be possible for them to be provided with higher levels at the expense of extra service costs that would be lower than the extra benefits. Conversely, if they were to live in areas with higher levels, then the service cost savings incurred by shifting them to areas of lower levels would exceed the fall in benefits.

Now, α and ß will settle in areas D and G, as required for efficient

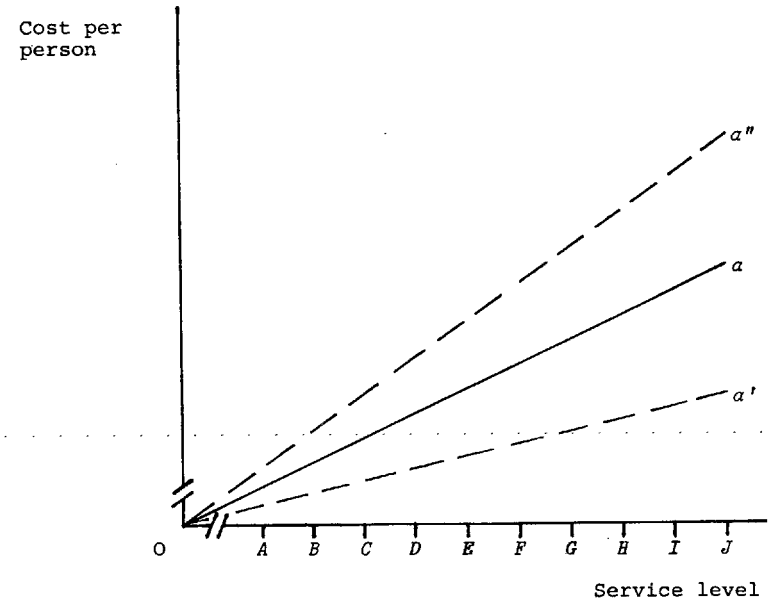

Figure 6.1

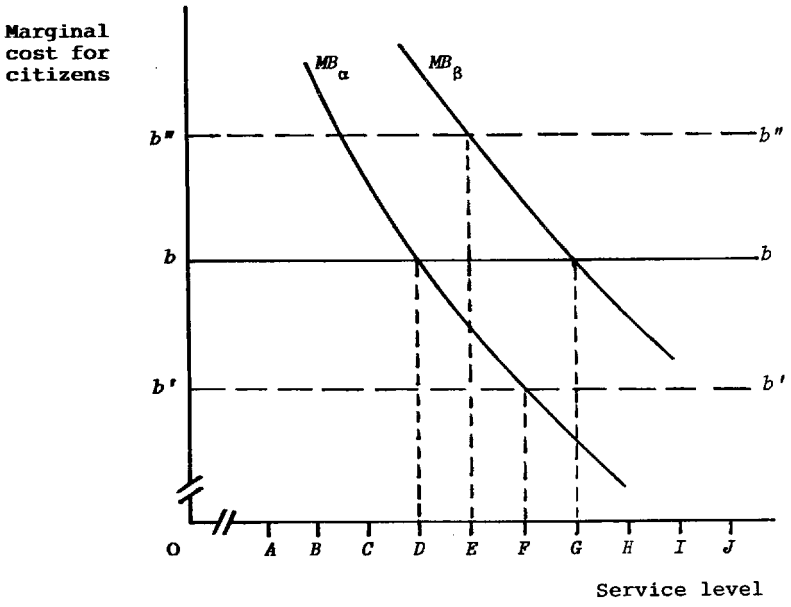

Figure 6.2

location, if the authorities levy a poll tax. With this tax, Oα in Figure 6.1 shows the tax payable for each person in each area, and *bb* in Figure 6.2 shows the extra tax to be paid when citizens move to areas of higher levels. α and ß will settle in D and G respectively, because, if they were to settle in areas of higher (or lower) service levels, then they would find moves to areas with lower (or higher) levels reducing (or increasing) their taxes more (or less) than their benefits.

It follows that a poll tax results in efficient locations for a service where costs are proportional to population. Suppose, instead, that the authorities levied a proportional tax on personal incomes. If total incomes per head were the same in each area (or if there were an equalisation scheme), then each area would need to set a tax rate proportional to its service level. In this case, the marginal costs to α and ß of moving to areas of different service levels would depend on their incomes. For instance, if α had a lower income that ß, then although moves from, say, B to C or H to I would give them equal rises in tax rates they would result in lower rises in tax payments for α than for ß; an individual with zero income would be in the extreme position of being able to move from area to area without any change at all in his (zero) tax payments and would always choose the area with the highest service levels. The lines $b'b'$ and $b''b''$ in Figure 6.2 show the actual marginal costs imposed on α and ß by this tax system: α will settle in F and ß in E, but these are not the efficient locations for them. The inefficiency arises because α and ß react to cost schedules that do not affect the true costs of varying the service levels for them. The inefficiency means that the tax has an excess burden in this context. A similar result would arise with a tax on residential buildings if α had a smaller property than ß, since $b'b'$ and $b''b''$ could represent the costs faced by α and ß with such a tax, but not the true costs of varying the service levels for them.

It might appear that a poll tax is the only one of the taxes on households that avoids such an excess burden, but it does so only if the service costs are proportional to population. Suppose, instead, that the costs were proportional to the income level (Y) and population (P) so that $TC = kYP$. In this case, the cost for any person would vary with that person's income, and Oa' and Oa'' in Figure 6.1 could show the costs that each area would incur by accommodating an extra low- or high-income citizen respectively. Thus the true marginal cost of varying service levels for a person on a low income (such as α) might be shown by $b'b'$ in Figure 6.2, and the true cost for one on a high income (such as ß) by $b''b''$. In this case, the efficient locations for α and ß would be F and E respectively, and would be secured by a proportional income tax that (as before) confronted them with the schedules $b'b'$ and $b''b''$. A poll tax that confronted them each with the schedule *bb* would result in the inefficient locations D and G. Similarly, if service costs were proportional to residential property values, then efficient locations would be encouraged by a tax on residential property and not by a poll tax.

This example shows that, when the tax paid by a household can vary according to the household's area of residence, as is the case with a poll

tax, a personal income tax or a residential property tax, then inefficient locations will be encouraged unless, by chance, one tax (or a combination of two or more) results in an incidence on all individual citizens that perfectly reflects their contributions to subcentral authority costs.

It should be noted that other taxes are even less likely to result in efficient location decisions. For example, the burden of taxes on corporate incomes and non-residential property on the citizens who finally bear the burden will depend much more on the area in which the relevant business or properties are located than on their area of residence; so all citizens would probably be induced to live in the areas with the highest tax rates and service levels because personal tax burdens may be no higher there than anywhere else. With taxes on expenditure, whether levied generally or on specific items, the effects depend to some extent on the sizes of the authorities. If the authorities are small, then all citizens will be encouraged to live in high tax and service level areas and shop elsewhere; if they are large, then high spenders may be encouraged to live in low tax rate areas and low spenders encouraged to live in high tax rate areas, and location patterns would be inefficient[17] (unless subcentral authority service costs were proportional to the relevant spending). Similarly, if payroll taxes were levied by small authorities, then citizens would be encouraged to live in high tax and service level areas, but work elsewhere; if they were levied by large authorities, then high earners would be encouraged to live in low tax rate areas while low earners would be encouraged to live in high tax rate areas, so that location patterns would most probably be inefficient.

In practice, of course, it is unlikely that any tax will be found whose incidence on citizens perfectly reflects their contributions to costs. It is therefore likely that some excess burdens will arise whenever tax rates vary. This burden must be seen as part of the price of allowing service levels to vary in accordance with variations in tastes between areas. Choosing a tax that minimises this excess burden may, in practice, mean that the tax with the least excess burden is regressive. If it were desired, on equity grounds, to have a progressive structure for taxes as a whole, then this could be achieved by complementing regressive subcentral taxes with suitably progressive central ones.[18]

It should be noted, though, that it would not be appropriate to attempt to finance any subcentral redistribution functions with taxes related to the cost of benefits received. Redistribution clearly requires that the poor should pay less in taxes than they receive in benefits while the rich should pay more in taxes than they receive in benefits. It was explained in Chapter 1 that different degrees of redistribution within different areas could create migration problems, and for this reason it is often argued that subcentral authorities should have little or no redistributive role. These migration problems, which would be caused only when relevant tax (and benefit) provisions varied, are similar in many respects to those considered in this section. It is not clear, however, how far such migration could be said to lead to inefficient location decisions since, as far as redistributive programmes are concerned, it is clearly not efficient for all citizens to locate where their tax payments equal the costs of their service benefits.

Tax exporting

In the last section, the discussion of excess burdens was concluded by considering the point that a good subcentral tax system should result in each citizen making an efficient choice about his area of residence if he was faced with a selection of subcentral authorities whose service levels varied. However, a good subcentral tax system should also help to ensure that the service and budget levels selected by any subcentral authority are themselves efficient. Ideally, this means that it should lead to an authority raising its output of any public good or service up to the point where the marginal rate of transformation between that good or service and any privately produced good or service equals the sum of all its citizens' marginal rates of substitution between them. If the price of each private good or service equals its marginal cost, then the optimality condition for the output of each subcentral service simplifies to rule that an authority should expand output until the aggregate marginal benefit equals the marginal cost.

It was explained in Chapter 1 that the optimal quantities of publicly provided services are not likely to be realised in practice because of the shortcomings of the electoral processes by which decisions on quantities are ordinarily made. Nevertheless, it is widely and understandably believed that these decisions will be improved if electors are, between them, required to meet the cost of the services provided by their subcentral authorities. If the tax system enables them to shift part of the cost of the services on to other people, then the services will, in effect, be subsidised, and the electors are likely to press for higher-than-efficient levels of service.

This criterion is sometimes modified and replaced by a weaker one arguing that it is acceptable for part of the cost of subcentral services to be borne by revenue-sharing grants, provided that electors are fully responsible for the costs of services at the margin.[19] In this way, electors are presumed to continue to seek subcentral services only up to the optimal point where marginal benefits and costs are equal. This weaker criterion is arguably inferior in two respects, however. First, the grants envisaged may, understandably, lead to control by the grantor, and thereby inhibit the electors' attempts to obtain the types and levels of services desired. Secondly, it is possible that the existence of general grants may lead to expenditure levels that exceed the optimal levels; this is the implication of some of the models of grant effects outlined in Chapter 3, specifically the Niskanen, Romer and Rosenthal, and Oates models, though it is not the implication of the other models discussed there.

It should be noted that neither of these two criteria would be met if a system of effort-related general grants were in operation. It was explained in Chapter 5, though, that the efficiency defects of such grants might be judged more acceptable in practice than the equity defects of the alternative lump-sum general grants.

There are two apparent ways in which electors might shift on to others the taxes that are needed for the services they enjoy. First, they might levy taxes that result directly in tax exporting – that is to say, that will be paid by non-residents.[20] This could perhaps be done most effectively by

imposing taxes on restaurant and hotel bills in tourist areas. Hogan and Shelton (1973, p. 556) have pointed out that the tendency of any area to have higher service levels when it shifts some tax payments on to non-residents may be partially (or perhaps more than wholly) offset by a tendency to provide lower levels as a response to income losses when residents pay taxes to other areas. However, widespread tax exporting does not alter the aggregate income for a group of authorities, although it does cut the relative price of their services to their voters, so it may well result in excess provision by the group as a whole. Secondly, they could set out to levy taxes on businesses, which typically have no vote and hence no say in tax rates. The incidence of business taxes nevertheless falls on people, and the electors will succeed in shifting taxes in this way only to the extent that the incidence of the taxes imposed falls on non-residents. Thus this method works only insofar as the taxes are, in the end, exported. It is arguable, of course, that subcentral authorities should be allowed to levy some degree of taxes on businesses on the grounds that their services, or at least some of them, are of direct benefit to businesses. Equally, it is arguable that, if some subcentral taxes are regarded as a proxy for charges on businesses, then it would be more equitable and efficient to impose charges that related to the level of services enjoyed by the individual businesses than, say, taxes related to the value of their properties, sales or profits. Where taxes on businesses are levied instead, there is perhaps a case for trying to ensure that the share of tax revenues raised from businesses is in some measure proportional to their share of the benefits from subcentral services. It is probable that, where authorities raise a property tax alone, businesses typically pay a larger than proportional share. An extreme case of this occurs in Shetland in the United Kingdom where in 1982–3 businesses accounted for 97 per cent of the property tax base (CIPFA, Scottish Branch, 1982, pp. 82–5).

It should be added that taxes may, in fact, be exported in a third way, without any explicit action by subcentral authorities, if subcentral taxes may be offset against liability for central taxes; in some countries, subcentral tax payments can be deducted or offset from taxable incomes for central income tax purposes. Thus a citizen whose marginal rate of central income tax was 40 per cent would lose only 60 cents if his subcentral authority raised his tax payments to it by $1. In this way, part of any rise in a subcentral authority's taxes is financed with smaller payments of central taxes by its citizens and therefore imposes a cost on citizens elsewhere who have to make good the reduced central tax receipts. This method of exporting taxes can readily be eliminated by not permitting subcentral payments to be offset against other taxes. One study by McLure (1967) in the United States (for 1962) suggested that, between them, these three methods of tax exporting resulted in an average of 20–25 per cent of state taxes being exported; in some states, the figure was around 40 per cent.

It should be noted that it may sometimes be thought appropriate for subcentral authorities to levy a tax some of whose incidence is exported. Some of the benefit of a subcentral authority's services will accrue to non-residents, and it was explained in Chapter 4 that in such cases the authority would be likely to provide too little of the services if its citizens had to

meet the full costs, since it would ignore these external benefits. It was argued that this problem could be overcome with the use of suitable matching grants specific to the services concerned. An attempt might instead be made to prevent underprovision by granting the authority a tax some of whose incidence fell on non-residents. This method is very crude compared with a set of specific grants, because it does not permit the degree of external finance to vary from service to service according to the degree of external benefit, and it encourages higher provision of services with no external benefits. However, the difficulty of measuring external benefits might be so great that, in practice, no great confidence would be placed in the results obtained by specific grants, and a partially exportable tax might be deemed a tolerable, albeit inferior, substitute.[21]

Whether or not some tax exporting is desired, it is important to consider the extent to which the various taxes can in fact be exported. This depends, of course, on the incidence of the taxes. It is generally accepted that establishing the incidence of taxes involves very great difficulties. For the present text, though, it seems appropriate to make the usual principal assumptions about incidence and not to explore the qualifications to those assumptions that are sometimes made.[22] It should be stressed that any tax could be exported if payments could be offset against liability for central taxes. In the following paragraphs, this possibility is assumed not to operate.

With a tax on personal incomes, a distinction needs to be drawn between the greater part accruing from wages or salaries (that is, from labour) and the smaller part accruing from interest or dividends (that is, from capital). The incidence of a tax on income from labour would fall wholly on the (resident) employees if the supply of labour by them were perfectly inelastic. To the extent that their supply of labour is elastic, some of the tax falls on employers and will result in higher prices and lower profits, thereby shifting part of the incidence on to owners of capital and customers, not all of whom will be residents. If the supply of labour is fairly inelastic, as it is considered to be, then little exporting of the greater part of a personal income tax (that on wage income) is likely. Of course, a high tax rate in one area may encourage residents to move to another, but in that case their original area ceases to tax them, and no tax exporting as such will arise directly. However, if this emigration from the original area causes employers there to raise wages in order to maintain output, then there will be a tendency for lower profits and higher prices, so that some of the tax could be exported indirectly on to any owners or customers who were non-residents. This exporting might cease in the long run if the firms relocated in other authorities. Incidentally, if the supply curve of labour were backward-bending, then the tax would reduce gross wage rates so that employers and, in turn, non-residents, would gain.

The incidence of the tax on income from capital would fall wholly on the (resident) suppliers of the capital if the supply of capital by them were perfectly inelastic; if it were not, then it might appear to follow that some of the incidence would be shifted on to the owners of capital, who would often be non-resident. If the demand were perfectly elastic, however, the incidence would fall wholly on residents, even if the supply were not

perfectly inelastic. For one subcentral authority, demand may be perfectly elastic since capital in other areas is typically a perfect substitute. Thus it is likely that little exporting occurs with the smaller part of a personal tax on income (that on capital income). Altogether, then, a tax on personal incomes should fall mainly on residents.

A tax on payrolls (that is, a tax on income from labour paid to the area of workplace) is similar to the tax on income from labour just considered in that its incidence falls wholly on employees if their labour supply is perfectly inelastic, and partly on employers if it is not. The supply of labour in a country may be fairly inelastic, but it is likely to be much less inelastic in one area of a country, particularly in the long run, so that part of a subcentral tax on payrolls seems certain to fall on employers, and thereby tend to increase the prices and reduce the profits of businesses in the area. Some of this part of the tax falling on employers will be exported because some of their customers and owners will be non-resident. Likewise, some of the part of the tax that falls on employees will be exported, especially in small areas, as some of them will be non-residents; perhaps the classic example of this situation is an urban area that levies a payroll tax much of whose incidence falls on those who commute there to work from adjacent suburbs. Altogether, it seems likely that tax exporting could be more serious in this case than with the income tax, particularly in small areas. Thus McLure (1967, p. 55) noted that 7 per cent of the Delaware payroll tax was paid by non-residents, although the analogous figure was under 1 per cent in most states.

The incidence of a corporation tax falls wholly on business owners (that is, on capital) if the supply of capital is perfectly inelastic. If the supply of capital is not perfectly inelastic, then the tax will cause the quantity supplied to shrink, thereby reducing both the demand for labour and the supply of products and shifting some of the incidence on to employees and consumers. The supply will be perfectly inelastic in the short run, but may not be in the long run, especially in a single subcentral authority's area. From the point of view of a subcentral authority using such a tax, part of the incidence falling on business owners will be exported because some owners will be non-residents, and (a perhaps smaller) part of any incidence falling on employees and customers will be exported if some employees and customers of those businesses are non-resident. Overall, this tax is likely to result in a substantial degree of incidence being exported.

To consider the incidence of a property tax, it is useful to consider separately the part of the tax on land and the part of the tax on buildings. Land is in perfectly inelastic supply, and the incidence of a tax on all land falls wholly on its owners; more precisely, the tax falls on the owners of the land at the time the tax is introduced (or increased), for land values will fall sufficiently for the net return to equal the net return available from other assets. A land tax levied by a subcentral authority can be split up into a tax on the land belonging to owner-occupiers of residential property, a tax on the land occupied by tenanted residential property and a tax on land used for non-residential purposes. The tax on the first category will fall wholly on the resident owners and not be exported. The tax on the other categories of land will be exported to the extent that the land is owned by non-residents. It should be noted that, if the tax exempts

some land, such as agricultural land, then it might reduce the rate at which such land was developed for other purposes and thereby make the relative values of agricultural and other land different from what they would otherwise have been. Any resulting changes in prices and profits could affect the degree to which the tax was exported.

A tax on buildings will fall wholly on owners in the short run because, in the short run, the supply of buildings is perfectly inelastic. Thus, the tax on owner-occupied residential buildings will fall wholly on the resident owners, but the tax on tenanted residential buildings[23] and non-residential buildings will be exported to the extent that the buildings are owned by non-residents. In the long run, the supply of buildings may be very elastic, and so, in time, the incidence of the tax may be largely shifted on to occupiers. In that case, the incidence falls on residents for both types of residential buildings. For non-residential buildings, the incidence of the tax on occupiers will eventually be split between the owners, employees and customers of the businesses occupying them, and much of this incidence may be exported. It should be added that, if the aggregate supply of capital is perfectly (or highly) inelastic, then the reduced supply of buildings will be wholly (or largely) offset by increased supplies of other assets, which should reduce prices of items made from those assets. Thus, overall, the losses to a nation's citizens from paying more for occupying buildings may be substantially offset by paying lower prices for other items; indeed, Aaron (1975, p. 43) argued that, for this reason, approximately the entire burden of a tax on property will fall on capital owners, which is what would be expected if the aggregate supply of capital were fixed. From the point of view of an individual authority, however, the benefits from any increased supply of other assets created by its own property tax will generally be shared by all the nation's citizens (and perhaps foreigners too). These benefits thus do little to offset the increased price of occupying buildings in its area, and so the incidence of its tax on buildings will continue to fall broadly on occupiers.

The imposition of a general tax on all consumers' expenditure, to which a general sales tax and, if capital goods are zero-rated, a value added tax might approximate, would drive a wedge between the net and gross prices of consumer goods and services. If the costs of production remained largely unchanged (in money terms) then prices would rise by the amount of the tax and the incidence would fall chiefly on consumers. From the point of view of a subcentral authority, the degree of exporting would depend principally on the proportion of consumers who were non-residents; accordingly, it is likely that exporting will be more significant with small authorities than it will be with large ones. Exporting could be removed by exempting non-residents, but such devices would lead to excess burden problems if they encouraged residents in one area with a general sales tax to make shopping expeditions to another area where they were exempt.

In practice, of course, so-called general sales taxes generally omit many items, and the analysis of their incidence might best be handled by regarding each such tax as a collection of taxes on specific items and looking at the tax on each item separately. The incidence of a general sales

tax on any item covered, and the incidence of any tax on a particular item such as motor fuel, alcohol, tobacco or natural resources, is generally shared between purchasers and producers; the ratio of the shares equals the ratio of the elasticity of supply to the elasticity of demand. For motor fuel, alcohol, tobacco and natural resources, at least, these ratios are probably fairly high in most countries as a whole, although they may be lower in any individual subcentral authority – particularly perhaps in the case of natural resources, since a rise in prices there tends to divert custom elsewhere. While the incidence of any subcentral authority's taxes may be felt by both purchasers and producers, the producers will suffer less if they also own plants in the areas to which custom is diverted. A particular complication in the case of taxes on motor fuel and natural resources is that many purchasers in the area levying the tax may be businesses, and thus the ultimate incidence of the tax on them may be shared by their customers, employees and owners.

For a subcentral authority, the amount of the incidence on purchasers that is exported depends generally on the proportion of purchasers who are non-resident. This exporting could be avoided by exempting non-residents, but the widespread use of such a practice could create large excess burdens because all purchasers would be tempted to buy in areas other than their own. In the case of purchases of motor fuel and natural resources by businesses in the area, the degree of exporting depends on the extent to which the tax is felt by their owners, employees and customers and on the proportions of those groups who are non-residents. The incidence of the taxes on producers is likely to be spread between employees and owners of businesses in the area, and the amount exported depends on the degree to which they are non-resident.

The incidence of a poll tax on residents will fall wholly on residents. Thus, this tax, along with a tax on personal incomes and a tax on residential buildings, appears one of the most satisfactory if tax exporting is to be minimised. Conversely, a tax on business incomes is probably the least satisfactory. Some significant exporting is possible with all the other taxes, but the amount may be acceptable with many of them if used by relatively large subcentral authorities. This conclusion squares with the evidence of McLure's study (1967) of those taxes used in the United States in 1962; his main results were noted in Chapter 1.

Tax perceptibility

It should be noted that, even if subcentral services generated no external benefits and were financed wholly by taxes that could not be exported in any way, electors might still make inefficient choices over budget levels. This could occur if electors were unable to perceive clearly how much they were actually paying in taxes, and, therefore, were unable to perceive clearly how much subcentral services were costing. Thus another criterion for judging subcentral taxes is their degree of perceptibility, a point stressed by the Layfield Committee (HMSO, 1976, p. 184).

The perceptibility of most of the eleven major taxes under discussion could vary considerably according to how they were administered. A poll

tax might always be clearly perceptible, whereas a tax on personal incomes and a payroll tax would be much more perceptible if the tax on earning was collected from employees than if it was collected from employers and if the tax levied by a subcentral authority was collected separately from that levied by the central government. The perceptibility of the property tax on residential property will always be high in the case of owner-occupiers, whereas it may be high in the case of tenants only in those rare cases where the legal liability for payment lies with the occupier rather than with the owner.[24] Taxes on alcohol, tobacco and motor fuel (at least insofar as they are levied on sales to consumers), a general sales tax and a value added tax will be perceptible if retailers' price labels show the tax due on each item, though even then citizens may have much less idea about their total annual payments of these taxes than they have about poll taxes, income taxes, payroll taxes or residential property taxes. Even if retailers do not show the tax due on each item, these taxes on expenditure could probably be held to be a little more perceptible than corporation taxes, or taxes on business property, natural resources, or sales of motor fuel (to producers), because changes in tax rates with the former group might lead more perceptibly to changes in prices since their incidence may fall more on consumers in the area concerned.[25]

Tax competition

It is sometimes argued that there is a third way in which the choice of subcentral taxes can lead to inefficient levels of services; this could occur if subcentral authorities keep tax rates below the ones needed to secure efficient service levels, either in an attempt to attract new businesses into the area, or for fear of driving existing businesses away. Authorities are seen as taking part in a 'tax competition' for businesses and providing suboptimal services as a result. This line of reasoning was discussed at some length in Chapter 1.

It has been suggested by Grewal and Mathews (1977, pp. 84–5) that the taxes most likely to be used for competitions are a payroll tax, a corporation tax and taxes on expenditure, on the grounds that these taxes represent significant costs to businesses. In fact, a low payroll tax may have to be passed on in higher wages in the short run given the inelasticity of the supply of labour to a given area in the short run; but it will be to the benefit of businesses in the long run, when the supply of labour to any area (by both residents and non-residents) could be quite elastic. Business owners will clearly prefer low corporation taxes, since much of the incidence of these is likely to fall on them. They will prefer low rates for taxes on expenditure, even if their incidence generally falls chiefly on consumers, since a firm in a low-tax area can hope to attract customers from firms in high-tax areas. They will also generally be tempted by low taxes on natural resources.

Another tax that could be used in competitions is a tax on business property, but, in most countries, reductions in that tax have to be matched by reductions in the tax on residential property, and they will thus lead to substantial falls in revenue and service levels, which may not be thought

justified by the amount of any business likely to be persuaded to stay in or enter the area. If the authority levying the propery tax is also responsible for valuations, however, then it could perhaps seek to give businesses artificially low valuations and reduce their tax payments in that way. More recently, it has been suggested that a personal income tax might be used because businesses would then feel able to offer lower wage rates in low-tax areas (ACIR, 1981a, pp. 15–19). Similar reasoning would suggest that poll taxes or a tax on domestic property could also be used. If so, then it seems there may be no tax that could be entrusted to subcentral authorities in the certainty that it would not be used for tax competitions.

The spread of tax incidence

There is yet one other way in which a particular tax structure might make it unlikely that subcentral authorities will provide their services at optimal levels. This can arise, even when the incidence of subcentral taxes falls wholly on residents, if the incidence is spread very unevenly among those residents. Suppose each authority provides services at the levels desired by its median voter, and that it raises its revenues from, say, a poll tax on citizens whose heights exceed 1.80 metres. It is probable that citizens below this height would demand higher service levels than those above, and probable, also, that the citizen with median preferences would be in the former group. Now, it is unlikely that the median voter's preferred levels would precisely equal the optimal levels no matter what tax were used, but it is clear that his preferences could far exceed the optimal levels if the tax were one that he wholly or largely escaped paying. In general, it can be fairly safely said that his preferences will be closer to optimal levels with a tax whose incidence is well spread out among all citizens than with a tax whose incidence is largely confined to a small group of citizens. Thus, one criterion for a good subcentral tax is that its incidence should be widely spread.

As far as this criterion is concerned, it is fair to assume that a poll tax is most effective. A general sales tax or value added tax is likely to be spread out more evenly amongst an area's citizens than a personal income tax or payroll tax, since the distribution of consumer spending is typically less uneven than the distribution of incomes. The distribution of taxes on alcohol and tobacco might be much less even than that of taxes on general sales or personal incomes; indeed, the median voter in some areas might pay very little alcohol tax and no tobacco tax at all. The spread of the tax on motor fuel is hard to assess, for much of it is paid initially by business users; it may often be more even than the spread of alcohol and tobacco taxes but perhaps less even than the spread of a general sales tax or income tax. The spread of taxes on natural resources may be relatively small if these are borne chiefly by their producers. The incidence of a corporation tax may be concentrated substantially in those households with capital owners, and again the median voter may largely escape tax.

The situation with a tax on residential property depends on how the incidence is shared among voters in households with more than one voter.

If the incidence falls on one, say the head of the household, then the median voter might escape tax altogether and choose excessive service levels. If the incidence is shared equally, then the service levels chosen by the median voter could depend critically on whether he lived in a one- or two-voter household. The tax on business property has a very diffuse incidence; it may, in fact, be well spread out, but its perceptibility may be so low that the median voter assumes he escapes the tax altogether. It can be seen that none of the other taxes compares at all favourably with a poll tax. More research is needed to make careful comparisons of the other taxes, but a general sales tax and a value added tax are likely to be the best of them.

Taxes and their macroeconomic implications

Any change in any tax yield is capable of having some macroeconomic effect. Thus it might affect the level of aggregate demand, and so affect the level of unemployment and, perhaps, the rate of inflation. It could also lead directly to a change in the retail price index and so affect the observed level of inflation and, perhaps, exacerbate wage demands. Further, it might affect the level of investment and so the rate of economic growth. Finally, it might affect costs directly, thereby affecting the level of exports and, perhaps, imports. It is the purpose of Chapter 9 to consider how seriously changes in subcentral tax yields could in these ways affect the central government's macroeconomic policies. For the moment, it is sufficient to consider whether some taxes have more effects than others for a given change in yield.[26]

Changes in any tax will affect aggregate demand but, if the marginal propensity to consume for the rich is lower than that for the poor, then the biggest changes will occur with the most regressive tax. In this respect, then, a poll tax is almost certainly the least satisfactory and taxes that fall most on capital (such as a corporation tax) perhaps the most satisfactory. On the whole, though, taxes on personal incomes or payrolls and business property are likely to be more progressive, and hence more satisfactory, than any taxes on expenditure or residential property.

Taxes on expenditure, whether general or confined to specific items, and taxes on residential property will generally have an immediate impact on the retail price index. In this respect, they are most likely to have an immediate effect on inflation as well as result in higher wage demands, although taxes on business buildings (if not land) may also lead to price rises in time. Since taxes on personal incomes and payrolls both reduce net wages, they could also lead to higher wage demands. Indeed, a rise in a poll tax could have a similar effect. Much of any rise in corporation taxes is likely to be absorbed in lower profits, and this tax could be the most satisfactory in this respect, together with the tax on business land.

When it comes to effects on investment and growth, taxes on corporate incomes and on personal incomes are probably the least attractive because increases in them significantly reduce the net return on capital. There may be little to choose between the other taxes, which will have small effects

on the net return on capital by lowering demand and raising costs.

Taxes on expenditure could have an immediate direct impact on the competitiveness of exports. It is easy to overcome this effect in cases such as a general retail sales tax, a value added tax and a tax on alcohol or tobacco by remitting the tax on exported items, but it is harder, if not impossible, fully to overcome the effect in cases such as taxes on natural resources or motor fuel if they raise the costs of other domestic industries and make their exports more expensive. The tax on business property has an effect on business costs in the long run and so would have a direct effect, but this might be offset by the tendency for capital owners to shift from buildings to other assets, thereby reducing prices of other items. All tax rises can have a delayed indirect effect by raising the domestic level of prices, although, as noted, a corporation tax and a tax on business land are likely to have the least effects.

It appears that the tax on business incomes would undoubtedly be the most satisfactory tax in terms of its macroeconomic effects if it were not for its effects on growth. Overall, the differences between the taxes are perhaps not great enough to imagine that much weight would be attached to performance on this criterion when selecting a tax for use by subcentral authorities.

Taxes and equity

The previous criteria have focused chiefly on efficiency aspects of subcentral taxes. However, there is likely to be widespread support for the value judgement that subcentral taxes should also be in some sense equitable – that is, that they should result in individuals making payments that vary directly, and certainly not perversely, with their abilities to pay. Of course, this approach may seem irrelevant if the logic of the discussion of excess burdens is followed, for that suggested that an individual's payment of subcentral taxes should relate to the benefits he receives and not to his ability to pay. In practice, that course may be so hard to follow that ability-to-pay considerations have to be borne in mind. The discussion of this equity issue frequently stems from consideration of the tax on domestic property and the three main ways in which it is often held to be inequitable.

First, the tax may be unfair on those who live in areas where domestic property is relatively expensive. Suppose, for simplicity, that all citizens are owner-occupiers, and that domestic property is more expensive in area A than in area B, perhaps because A is more densely populated than B or enjoys a better climate. Suppose, too, that citizen α in A has the same income, family size and type of property as citizen ß in B. Now, α will have to pay more for his property, and he will have a higher property tax base. If the authorities in A and B each levy a tax on domestic property, and if the per capita tax base is higher in A than B, then A may well be able to levy a lower tax rate than B to secure similar services. In turn, α will pay a lower rate on a higher base than ß, and so perhaps pay the same total amount of tax for similar services. Problems will arise, however, if an equalisation scheme seeks to compensate B for its low tax base, because ß may well end up paying less tax than α for similar service levels. As explained in Chapter 5, equity problems of this sort can be avoided

by having equalisation schemes for property taxes that ignore the differences in domestic property values between areas.

It is clear that the problem here stems from having a property tax combined with an unsuitable equalisation scheme rather than from the property tax itself. Similar problems can arise with other taxes (although typically they would be less acute) if unsuitable equalisation schemes were used. Thus, it was suggested in Chapter 5 that equalisation schemes for taxes on expenditure should ignore differences in price levels between areas, and that equalisation schemes for taxes on incomes should ignore differences in wage levels between areas.

Secondly, domestic property taxes may be held to be unfair inasmuch as a married couple with two grown-up children living in an owner-occupied house may pay no more in property taxes, even if all four of them are working, than the elderly widow next door. This, of course, is a direct consequence of a tax on domestic property. It is equally possible that the widow would pay more in alcohol tax, tobacco tax or a natural resource tax if she consumed alcohol, smoked or used natural gas for her central heating while her neighbours were teetotal non-smokers who used log fires. Likewise, she might suffer more from a corporation tax if she had a few shares in some companies and they did not.

There are perhaps two main comments to make about this problem. First, it will be less acute with a general sales tax, a value added tax, a payroll tax, a tax on personal income and, arguably, a poll tax. Secondly, if it is decided to levy taxes where the problem may be acute, then, as explained in the introduction to this chapter, it may be better if these taxes are levied by central governments, which can raise many taxes, than by subcentral authorities, which can raise few. If it was desired not to raise the rate of such a tax for fear of hurting those on whom they impinge most, then the central government could still easily raise its expenditure by raising its other taxes, whereas a subcentral authority might have few other taxes to raise, and perhaps none.

The third main charge made against taxes on domestic property is that they are regressive. This charge is probably most sustainable where properties are typically owner-occupied, because some at least of the incidence on tenanted property may fall on owners, who may often be relatively rich. An illustration of the regressive nature of these taxes is given by some research conducted in the United Kingdom for the Layfield Committee (HMSO, 1976, pp. 159–62 and 425–34). This research indicated that, in 1975, the property tax payments due on domestic property amounted on average to over 7 per cent of gross household income for households where that income was under £20 per week, and fell to under 2 per cent for households where it was over £100 per week. The results would no doubt have been a little different if the tax base had related to capital values rather than estimated rental values, although the Committee felt (p. 172) that a switch to capital values might raise the payments due on both cheaper and more expensive properties while reducing the payments due on properties in between.

It has been argued by Aaron (1975, p. 29) that estimates of the impact of property taxes should not relate payments due to current incomes, as

was done by the Layfield Committee, but should relate them to normal incomes – that is, incomes averaged over a period of some years. This is because the low-income groups include some people whose incomes are temporarily depressed, so that using current incomes exaggerates the long-run burden on them, while the high-income groups include some people whose incomes are temporarily inflated, so that using current incomes understates the long-run burden on them. If normal incomes were used, then the tax should appear a shade less regressive. It would look even less regressive if tax payments due were expressed as a percentage of disposable incomes – that is, incomes after deducting direct taxes and adding in cash benefits.

However, a much more significant reason for believing that the property tax in the United Kingdom is less regressive than these Layfield Committee figures suggest is the existence of cash transfers (chiefly rate rebates) designed specifically to help low-income households meet their property tax payments. Thus it can be argued that attention should focus on net property tax payments – that is, total payments due less the value of these special handouts. In 1975, net payments were on average zero for households with gross incomes under £20 per week, and rose from around half a per cent for households with incomes between £20 and £25 per week to around 2 per cent for those with incomes over £100 per week that were not eligible for any specific handouts (HMSO, 1976, p. 162).

It seems, then, that objections to the regressivity of taxes on domestic property may be substantially reduced by means of programmes such as the rate rebates used in the United Kingdom, the whole array of circuit-breakers used in many parts of the United States, and similar measures used in some other countries (see OECD, 1983b, pp. 68–72). It may be added that the evidence on the impact of different taxes, although it depends on the precise nature of the taxes and the assumptions made, suggests that the only tax considered in this chapter that is always likely to be progressive in relation to original income for a wide range of incomes is the personal income tax,[27] though its progressiveness will be weakest where there is a proportional tax with a low exemption limit. An attempt can be made to have a progressive general sales tax or value added tax by exempting (or using low rates on) items that account for larger proportions of low-income budgets than high-income ones, but these attempts may not overcome the inherent regressivity of the taxes, which results from the fact that the average propensity to consume falls as income rises. A tax on motor fuel may also have an element of progressivity since low-income households may have no car, but it is unlikely to be progressive over the separate groups of non-car householders and car-owning households. Of course, these indirect taxes, like those on alcohol, tobacco and natural resources, and taxes on non-domestic property and business incomes, are all likely to be far less regressive than a poll tax.

Taxes in practice

The foregoing sections suggest that no tax is wholly satisfactory for use

by subcentral authorities. It is perhaps to be expected, therefore, that the extent to which subcentral authorities are able to raise tax revenue will vary from country to country, as also will the taxes those authorities elect or are permitted (as the case may be) to use. This section seeks to give some indication of how far these two expectations are realised.

Fortunately, this is one aspect of subcentral government for which there are some useful comparative data available, since the OECD (1983a) publishes relevant figures for all its twenty-three member countries except Turkey. An important limitation of these figures, however, is that they do not reflect the varying degrees to which subcentral authorities can determine their own tax rates, so that the subcentral authorities' shared taxes in West Germany (over whose rates individual *Länder* and local authorites have no individual control) are put on the same footing as taxes raised, for example, by the states of Australia or Canada. Table 6.6 shows the 1981 tax revenues of the state and local authorities in a number of those countries expressed as percentages both of gross domestic product (GDP) at market prices and of total tax revenue. The twelve OECD countries whose subcentral authorities account for under 10 per cent of total tax revenues are omitted along with Luxembourg. It will be seen that subcentral tax revenue is highest in relation to GDP in Sweden, at 16.46 per cent, and highest in relation to total tax revenue in Canada, at 43.60 per cent. It will also be seen that subcentral tax revenue falls below 25 per cent of total tax revenue in two of the six federal countries shown in the table (namely, Australia and Austria), yet exceeds that amount in four of the six non-federal countries (namely, Denmark, Finland, Japan and Sweden).

Table 6.7 shows the yields of the main individual subcentral taxes expressed as percentages of GDP.[28] At the state level, it will be seen that all states levy personal income taxes except those in Australia (which are now effectively empowered to do so but have chosen not to). It is chiefly this tax that helps states in Canada, Switzerland and West Germany to have substantial tax revenues, although all raise significant amounts from other taxes as well. The states of Australia rely chiefly on their payroll tax, which accounts for most of their revenue from 'other taxes', and the states of the United States are unique in raising more revenue from taxes on sales than from taxes on income or payrolls. Statements about subcentral taxation in the United States should be qualified, however, by pointing out that the array of taxes in use varies more from state to state than in other federal countries, and, as it happens, a few states do not levy income taxes and a few do not levy sales taxes. It will be noted that taxes on (immovable) property are generally relatively unimportant at the state level.

At the local level, it will be observed that taxes on individual incomes play a prominent role in the Scandinavian countries, which partly helps to explain why a number of them have a substantial proportion of tax revenues accruing to their local authorities. The table also shows that the tax on (immovable) property is the principal source of tax revenue for local authorities in Australia, Canada, the United Kingdom and the United States, but is not the major source of local tax revenue in any of the other

Table 6.6 *The relative sizes of state and local tax revenues in selected OECD countries, 1981*

	Tax revenues as a percentage of GDP			*Tax revenues as a percentage of total tax revenue*		
Country	*State*	*Local*	*State & local*	*State*	*Local*	*State & local*
Australia	4.70	1.20	5.91	14.91	3.83	18.74
Austria	4.35	4.75	9.10	10.23	11.15	21.38
Canada	11.97	3.18	15.15	34.45	9.15	43.60
Denmark	—	13.66	13.66	—	30.19	30.19
Finland	—	9.75	9.75	—	26.47	26.47
Japan	—	6.81	6.81	—	25.37	25.37
Norway	—	8.96	8.96	—	18.46	18.46
Sweden	—	16.46	16.46	—	32.09	32.09
Switzerland	6.96	5.28	12.24	23.00	17.45	40.45
United Kingdom	—	4.25	4.25	—	11.37	11.37
United States	5.32	3.36	8.68	17.06	10.77	27.83
West Germany	8.43	3.11	11.54	22.59	8.35	30.94

Source: OECD, 1983a, pp. 164, 169 and 171.

eight countries indicated.[29] This tax, which emerged in medieval England, seems more popular in countries later populated with English-speaking peoples than it does elsewhere. Finally, it will be observed that Japan's local authorities raise substantial sums from 'other taxes'; the principal one is the business tax, which is, in fact, closely related to a corporation tax (see Yonehara, 1981, p. 27).

Tax-sharing in West Germany

It was noted in the last section that some subcentral authorities acquire some (or all) of the revenue of certain taxes collected in their areas yet have no individual control over the rate of tax set there. As explained in Chapter 4, such an arrangement diminishes local autonomy over budget levels, and may or may not eliminate the threat of higher-tier interference that occurs with revenue-sharing grants. Such interference is least likely where the subcentral tax revenues have some form of constitutional guarantee. The most important examples of a situation where this form of tax revenue is accompanied by such guarantees are probably the tax-sharing schemes in West Germany.[30]

As far as the West German income taxes are concerned, the Constitution guarantees that a proportion of each of these taxes will be returned to the *Länder*, the *Länder*'s proportion being divided between them in relation to the money raised in each. Until 1969, the proportion going to the *Länder* varied from time to time, but it must be understood that proposals for changes in their share had to be passed by the upper house of parliament,

Table 6.7 *Subcentral tax revenues as percentages of GDP in selected OECD countries, 1981*

	Taxes on:							
	Individual incomes	*Corporate incomes*	*Immovable property*	*Wealth & inheritance*	*Financial & capital transactions*	*General consumption*	*Specific goods & services*	*Other*
State level:								
Australia	—	—	0.25	0.09	0.91	—	0.72	2.72
Austria	2.05	0.07	—	0.04	—	1.56	0.32	0.31
Canada	4.55	1.08	0.04	0.18	—	2.07	1.66	2.39
Switzerland	4.51	0.81	0.04	0.87	0.16	—	0.11	0.46
United States	1.45	0.50	0.10	0.08	0.03	1.65	0.95	0.56
West Germany	4.42	0.76	—	0.37	0.07	2.15	0.19	0.47
Local level:								
Australia	—	—	1.16	—	—	—	—	0.04
Austria	1.67	0.20	0.26	0.06	0.19	0.99	0.53	0.85
Canada	—	—	2.62	—	—	—	—	0.56
Denmark	12.10	0.19	1.35	—	—	—	0.01	0.01
Finland	8.87	0.87	—	—	—	—	—	0.01
Japan	1.93	0.81	1.40	—	0.12	—	0.95	1.61
Norway	7.76	0.55	0.16	0.25	—	—	—	0.24
Sweden	15.50	0.91	—	—	—	—	0.04	0.01
Switzerland	3.97	0.55	0.13	0.55	0.06	—	0.01	—
United Kingdom	—	—	4.25	—	—	—	—	—
United States	0.17	0.02	2.55	—	—	0.33	0.14	0.14
West Germany	2.05	0.40	0.39	—	0.09	—	0.01	0.17

Source: OECD, 1983a, pp. 85, 169, 171 and 174–95.

that is the Bundesrat, whose members comprise ministers of the *Länder* (or their nominees) who act in accordance with the decisions of their *Länder*; it was therefore constitutionally impossible for the *Länder*'s share of these taxes to be reduced without their consent.

Two changes were made to these arrangements with effect from 1970 as a result of amendments made to the Constitution in 1969. First, it was decreed that the main local authorities, that is the municipalities, should receive a proportion of the tax on wages and other income. The size of this share could be varied from time to time by legislation, and was 14 per cent in 1970. This share is divided between municipalities chiefly on the basis of the tax payments by their citizens. Secondly, it was decreed that the remaining part of this tax, and the yields of the corporation tax and the tax on interest and dividends, would be shared equally between the federal government and the *Länder*. Thus the *Länder* received 50 per cent of these last two taxes and 43 per cent of the much more significant tax on wages and other incomes. It should be appreciated that, although the legislation needed to alter the municipalities' share of the income tax yield would have to be ratified by the Bundesrat, it would not have to be ratified by the municipalities. Thus the municipalities might feel that their revenues contained no greater expectation of freedom from control than they would have with revenue-sharing grants.

These constitutional changes mean that the division of income tax yields between the federal government and the *Länder* can now take place only if the Constitution is further amended. Accordingly, it is less easy than it was to respond to changes in relative spending needs with changes in income tax shares. However, 1970 also saw the introduction of tax-sharing arrangements for the value added tax. The yield from this tax has subsequently been shared between the federal government and the *Länder*, whereas it was previously an exclusively federal tax. The share attributable to the *Länder* can be adjusted from time to time by legislation in accordance with the needs of each level of government. Such legislation must, of course, be ratified by the Bundesrat. In the 1970s, the *Länder*'s share fluctuated in the 30–40 per cent range.

However, it is arguable that the allocation arrangements for the *Länder*'s share of the value added tax mean that their benefits from this tax should strictly be seen as an example of general revenue-sharing grants rather than tax-sharing. This is because their share is not divided between them in accordance with the yield of the tax in each *Land*. Instead, 75 per cent of the yield is distributed between *Länder* in proportion to their populations, and 25 per cent is used to strengthen the position of *Länder* whose per capita tax yields are low. Even so, the arrangements differ from revenue-sharing arrangements in most other countries in that the *Länder*'s share cannot be altered without their consent.

It is useful to conclude this section by considering two other countries that have arrangements that are sometimes described as tax-sharing, but where the arrangements differ significantly from those in West Germany. In Canada, for instance, the term is sometimes used in relation to the personal and corporate income taxes. The federal government there collects these taxes in most of the provinces,[31] retaining some of the yield for

itself and returning some to the provinces. However, the provinces' revenues are collected for them essentially on an agency basis, and the provinces can determine the rates of tax that should be levied on their behalf. In essence, these provincial tax receipts are like any other tax levied directly by subcentral authorities.

In Australia, the federal government introduced a 'tax-sharing' scheme in 1976 whereby a specified percentage of its personal income tax yield (but, since 1981–2, its total tax revenue) would be distributed to the states. However, the distributions are made with reference to the states' needs and resources (as explained in Chapter 5) and not with reference to their income tax revenues. Thus, this scheme is essentially a revenue-sharing scheme rather than a tax-sharing scheme as defined here. It may be noted that the share allocated to states is not constitutionally guaranteed.

Summary

In some countries, all the subcentral authorities in a given tier levy the same tax or taxes, while in others the taxes used vary from area to area. It is suggested that there is a case for requiring all areas to use the same tax or taxes, and where they use more than one tax for all areas to attach the same relative importance to each. Unless tax structures are harmonised in this way, there could be horizontal inequity in that equals in different areas could pay very different total amounts in tax even though they might enjoy similar service levels. This inequity could stimulate migration, which would consume resources and, perhaps, have an adverse effect on output levels. It might be thought that a variety of tax packages would have the advantage of enabling people to vote with their feet for a preferred package, but in practice it is likely to aggravate any tendency for people to locate non-optimally because, for any individual, the relationship between the total taxes he would pay and the overall level of benefits he would enjoy in different areas could become very weak. A further limitation of unharmonised structures is that it might be hard for any one area to impose a more progressive structure than the others, since its choice might be untenable if the rich started to leave and poor people started to move in. This problem might be removed, however, if there was an equalisation scheme that compensated it for any falls in its tax base.

There are a number of taxes whose yields could raise a substantial part of the revenue needed by most subcentral authorities, and many criteria by which these taxes could be assessed. Altogether, eleven taxes are considered in this chapter. Of these, it is perhaps the corporation tax that has the most drawbacks: it requires the profits of each multi-area firm to be divided among the areas in which it operates; it is likely to have an unevenly distributed yield between areas; it could be used in tax competitions; its burden is likely to be substantially exported; and any burden that is not exported is unlikely to be very perceptible or widely spread. Its only attraction seems to be that variations in rates would have limited effect on aggregate demand, though they could affect investment and growth.

Taxes on specific items (such as motor fuel, alcohol, or tobacco) seem also to have disadvantages. For example, taxpayers may not be very aware of how much they pay, the payments may be spread unevenly between citizens, and the taxes may be used in tax competitions. They have further limitations that could be most serious if they were used by small areas: much of the burden might be exported; variations in tax rates could stimulate resource-wasting shopping expeditions; and there might be high bases in some areas and low bases elsewhere. Taxes on natural resources are likely to suffer to a high degree from at least some of these problems. More general taxes on expenditure (such as a general sales tax or a value added tax) would share many of these problems, but the geographical variations in their bases might be smaller and, more importantly, their incidence might be well spread out among all citizens.

Property taxes have fewer drawbacks than some others, but these drawbacks may still be regarded as serious. The most frequently mentioned are, perhaps, the weak relationship between the incidence of the tax on domestic property and ability to pay, the imperceptibility of the tax on non-domestic property, and a low elasticity in times of inflation and real income growth (although it is not really clear whether this should be seen as a disadvantage). The tax is likely to reduce the supply of buildings, and its base may be unevenly spread between areas.

Payroll taxes also have their problems. For example, some of their incidence is likely to be exported, especially if they are levied by small areas, they could be used in tax competitions, and they may not be very perceptible. A personal income tax should be better in all these respects; such a tax is also likely to have an elastic yield in times of inflation and output growth, although, as noted, this is a debatable advantage.

Of the taxes mentioned so far, a personal income tax might seem arguably the best. However, all these taxes are likely to suffer from one defect not yet mentioned, and the income tax may well suffer more than the others. The point is that all the taxes will encourage citizens to locate non-optimally unless they result in each citizen's tax payments reflecting the cost of providing him with services. While no tax is likely to do this perfectly, it is probable that taxes come closer to doing so the more regressive they are; hence the poor performance of a personal income tax. A general sales tax or value added tax would probably do better, as might a property tax, but a poll tax would probably be best. A poll tax would also do well in terms of being unexportable, highly perceptible, essentially free from excess burdens, and perfectly distributed between areas in relation to population; it might even be commended for its inelastic yields. Its non-existence stems undoubtedly from its regressive nature; perhaps a poll tax with rebates or circuit-breakers for the poor is a possibility worthy of further investigation.

Notes

1 These examples are related to those given by Oates (1972, pp. 131–40).

2 Hamovitch (1966, p. 413) has noted that for the city of New York, which was surrounded

by authorities with no sales taxes, each 1 per cent rise in the rate of sales tax between 1948 and 1965 led to a 6 per cent decline in taxable sales.

3 Oates (1972, p. 136) has argued that capitalisation is likely to mean that there is less migration in this case than with a poll tax and an income tax. However, the flight of the rich from B and of the poor from A occurs in those cases also, and should lead to similar effects on house prices. In each case, the damping effects of capitalisation are unlikely to be pronounced or long-lasting for the reasons noted in the text.

4 See, for example, Maynard and King (1972, p. 57) and the Layfield Committee (HMSO, 1976, p. 184).

5 The case for a subcentral value added tax is not often made, but Hicks (1968, p. 193) thought it might be easier to administer than a subcentral sales tax. The United Kingdom government of 1971 felt its administration would be hard, but this was because it assumed that a local tax would be worked on an 'origin' basis; there should be fewer problems with a tax on a 'destination' basis, as used in the European Economic Community (see HMSO, 1971b, p. 20).

6 These were considered for use in the United Kingdom by, for example, the Layfield Committee (HMSO, 1976, p. 186), and were recommended by the Royal Institute of Public Administration Study Group (1956, p. 128).

7 This criterion was noted, for example, by Hildersley and Nottage (1968, p. 15).

8 The figures in this paragraph are taken from the Commission of the European Communities (1977, p. 252). Separate figures for the general sales tax and excise duties were not given for the United Kingdom.

9 For a discussion of this issue, see Oates (1975) and Foster *et al.* (1980, pp. 495–8).

10 Maxwell and Aronson (1977, p. 113).

11 The Layfield Committee (HMSO, 1976) gave the yield as £4,000m. (p. 257) and collection costs as £100m. (p. 278).

12 See Sandford, Godwin, Hardwick and Butterworth (1981, p. 45) and the Layfield Committee (HMSO, 1976, p. 198) respectively.

13 See, for example, Boadway (1979, pp. 236–89) and Musgrave and Musgrave (1976, pp. 461–81).

14 Even with harmonised tax structures, there would still be some incentives for inefficient location decisions unless tax payments were related to benefits. This issue is discussed in the next section.

15 Oates (1972, pp. 142–3) discussed the merits of a land tax and concluded that it could be used more in the United States. However, it is arguable that non-residential land should not be taxed, because its inclusion in the tax base would mean that residents could shift some of the cost of services on to absentee non-resident landowners, and thereby be encouraged to vote for excess levels of services. Unfortunately, the tax on residential land does have an excess burden; the reasons are the same as those applying to a tax on all residential property and are explained shortly in the text.

16 The incentive and counter-incentive to migrate in the property tax case arise clearly in the case of owner-occupiers, for they will find taxes lowest in the lowest service level areas. In the case of tenants, the inelasticity of supply of buildings could suggest that property taxes are, in the short run, borne wholly by property owners (see Aaron, 1975, pp. 38–9) so that tenants need only consider service levels and could ignore property tax rates. However, a tendency for tenants to migrate to areas with high service levels will raise the demand for property in these areas and hence raise rents too. In the long run, the housing supply becomes elastic, and variations in property taxes may always be borne by tenants (see Aaron, 1975, pp. 39–40 and 55).

17 Foster *et al.* (1980, pp. 220–32) also considered whether capitalisation effects would prevent the inefficiency expected with various non-business taxes. In general, the tendency of those with high incomes, high spending or large properties to leave high-spending areas could depress up-market property values there, while the tendency for those with low incomes, low spending or small properties to arrive in high-spending areas could boost down-market property values there. These effects, and the converse effects in low-service areas, might dampen the enthusiasm for moving, but they will not be substantial. Large properties in the high-level area can be split up, thereby moderating the rise in small property values and the fall in large property values, while small properties in the low-level area can be knocked together, thereby moderating changes in relative values there also.

18 For a discussion of this point see Brennan (1977, p. 7). Oates (1972, pp. 121–5) argued that subcentral taxes should ideally have an equitable incidence, and suggested that this criterion is likely to conflict with efficiency criteria; but he also argued (p. 150) that the conflict is best resolved by letting subcentral taxes be determined by efficiency criteria and leaving any consequential distributional problems for resolution by the central government.

19 See, for example, the Layfield Committee (HMSO 1976, p. 184), which stressed that the most important aspect of this problem is that electors should be aware of the costs of changes in service levels; and see Oates (1972, p. 124).

20 For a discussion of tax exporting, see Browning and Browning (1979, pp. 420–3) and Oates (1972, pp. 140–2). For an analysis of tax exporting in the context of a general equilibrium model, see McLure (1969).

21 For a discussion of tax exporting in a situation where subcentral authorities generate spillovers, see Sandler and Shelton (1972).

22 For a detailed discussion of tax incidence theory, see Musgrave and Musgrave (1976, pp. 375–458) and Brown and Jackson (1982, Chapters 10–11).

23 Foster *et al.* (1980, p. 324) noted that the incidence would fall wholly on tenants in the short run if rents were controlled and maintained at sub-equilibrium levels, and if they were allowed to rise in line with property tax increases.

24 See OECD (1983b, p. 67). This report included a survey of twenty-two property taxes, in eighteen of which the legal liability for payment usually fell on owners rather than occupiers. Note that the incidence between owners and occupiers is not in fact dependent on who has the legal liability to pay. The widespread preference for putting the liability on to owners may result from a feeling that electors will believe the incidence lies with those liable and prefer that to be with owners.

25 The Layfield Committee (HMSO, 1976, pp. 186–8) considered that the lack of perceptibility of taxes on business incomes and payrolls, and of all taxes on sales, made them unsatisfactory for subcentral use.

26 The importance of this point is implied strongly in HMSO (1971b, pp. 15–23).

27 See, for example, Musgrave and Musgrave (1976, p. 391) for estimates of the burden of different taxes in the United States (in 1968) on different income groups, and O'Higgins and Ruggles (1981) for a similar analysis for the United Kingdom (in 1971).

28 For discussions of state and local taxes see, for example, Prest (1977) on Australia, Hayes (1977) on Canada, Maxwell and Aronson (1977, Chapters 5–7) on the United States and Spahn (1977) on West Germany.

29 For a detailed study of these property taxes, see OECD (1983b).

30 For outlines of these arrangements, see Hunter (1973, especially Chapter III) and Spahn (1977, especially pp. 215–17).

31 In fact, Quebec collects its own income tax and corporation tax, and Ontario collects its own corporation tax (see Hayes, 1977, p. 148).

7

Charges

Introduction

Charges by subcentral authorites often account for a significant proportion of their revenue. For example, current charges accounted for almost 11 per cent of the total tax and current charge revenue of states in the United States in 1980, while the figure for local governments was over 24 per cent (United States Department of Commerce, 1981, p. 279). The corresponding figure for local authorities in England and Wales in 1981–2 was around 17 per cent (HMSO, 1981a, p. 6). The early sections of this chapter are concerned with the principles that should govern the decisions on what subcentral activities should be financed partly or wholly by charges, and consider how the appropriate charges might be determined in each case. This analysis takes in turn the various types of function that subcentral authorities might perform, as discussed in Chapter 1.

In recent years, a number of writers have suggested that subcentral authorities should greatly increase their revenue from charges, perhaps by levying charges on services currently provided free (here defined to mean at zero prices) or perhaps by increasing the level of charges on services that are presently priced but at subsidised rates. It would not be appropriate in a theoretical text to attempt to examine every service that is somewhere provided by a subcentral authority to see how far it might be desirable to finance it by charges, but some of the most commonly found subcentral services are considered in the later sections of this chapter.

It should be observed that, although the debate on charges is essentially one about applying the principles of this form of finance to actual subcentral services, arguments that charge income should be raised are necessarily also arguments in favour of a reduction in income from taxes, tax-sharing schemes or revenue-sharing grants. As a result, there is likely to be less need for equalisation grants to help poor or needy areas, although there is likely to be more need for transfer payments to help poor or needy citizens. Moreover, any reduction in subcentral taxes should weaken the incentives for citizens to locate themselves non-optimally; it was explained in Chapter 6 that these incentives would arise if the authorities used different taxes, or, indeed, if they used the same tax, unless they used a tax where each taxpayer's payment reflected his contribution to subcentral authority costs. It may be added that the general tendency for revenue-sharing grants to grow, (observed in many countries in recent years) and the difficulty of finding wholly admirable subcentral taxes, as explained in the last chapter, add interest to the debate on charges.

The finance of subcentral authority enterprises

The discussion in Chapter 1 began by considering allocative functions. It noted that a perfectly competitive economy would, under certain assumptions, secure a Pareto-optimal or Pareto-efficient allocation of resources. Governments tend, therefore, to be most concerned with markets that are far from perfectly competitive or where the necessary conditions do not hold. These two situations will be considered in turn.

In practice there are few, if any, markets operating under conditions of perfect competition, but most concern is given to those at, or near, the opposite end of the spectrum, namely monopolies and oligopolies. It may be decided to control these in ways that attempt to ensure that their behaviour, particularly in respect of price, does not exploit their market position. Such control may be operated indirectly, by issuing legally binding regulations, or directly, by taking the firms into public ownership. Sometimes, as for instance with postal services, a free market is likely to produce a nationwide monopoly, and this suggests that it might not be sensible to have subcentral regulation or ownership, so that central regulation or a nationalised industry results. On other occasions, as for example with ferries, a free market may lead to local monopolies, and in such cases it might be equally feasible for regulation or ownership to lie with a subcentral government or with a central government. There is little explicit discussion in the literature about whether central or subcentral intervention would be preferable in such cases. However, most of the arguments in Chapter 1 favouring subcentral provision of local public goods could be used to make a case for subcentral intervention with local monopolies; indeed, such intervention has itself the essential characteristics of a local public good. In practice, subcentral intervention occurs quite commonly, but perhaps takes the form of ownership rather more often than it takes the form of regulation.

Now there is nothing in this analysis of why subcentrally owned enterprises may exist to suggest that their products should be offered free, and it would be expected that positive prices would be set. There is an extensive theoretical literature on the issue of public enterprise pricing, which has led to the prescription that prices should generally be set to equal marginal costs, as is the case in a perfectly competitive industry.[1] This literature is generally related to the case of nationally owned enterprises, but there is no reason to suppose that a different policy should apply to ones owned by subcentral authorities. However, there are three types of occasion where a little more discussion is needed.

First, there is the possibility that the enterprise concerned operates under conditions of increasing returns to scale.[2] This means that it would be faced by a long-run average cost (*LRAC*) curve similar to that shown in Figure 7.1. The associated long-run marginal cost curve is labelled *LRMC*. Suppose that the demand or average revenue curve faced by the enterprise is as shown by *AR*. Marginal cost pricing requires that, if capacity is optimal, then an output be set that can just be fully sold for a price equal to its long-run marginal cost. This occurs at an output level OQ_1, which must be given the price OP_1. The price is lower than the average

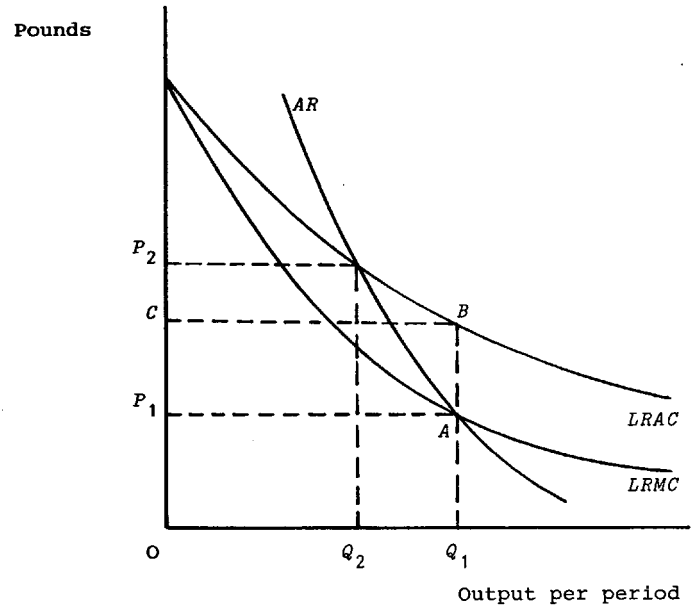

Figure 7.1

production cost OC, so that losses are sustained; these are represented by the area $ABCP_1$. Nevertheless, the total benefits to the consumers (represented by the area under the demand curve) exceed the total cost (represented by the area under the marginal cost curve), so that a positive output is desirable.

It follows that the subcentral authority will have to consider giving the enterprise a subsidy or requiring it to depart from marginal cost pricing. The latter course results in allocative inefficiency whether the enterprise is asked to operate average cost pricing or to operate some form of price discrimination. With average cost pricing, the enterprise would set the price OP_2 and so produce the non-optimal output OQ_2. With price discrimination, the enterprise might charge some types of consumer a higher price than others, which means that some consumers, if not all, would pay a price in excess of marginal cost. Alternatively, it might operate a multi-part tariff whereby all consumers had to pay a high unit price on the first units bought to help defray costs, and a low unit price equal to marginal cost on subsequent units, an arrangement that might drive some consumers away altogether. It is probable, though, that the multi-part tariff is the best of these non-marginal cost pricing options. By contrast, a subsidy that enabled output to stay at OQ_1 might seem altogether preferable, but against this it must be pointed out that the subsidies would have to be financed through taxes, and taxes themselves distort the allocation of resources. Subsidies may also be opposed on the grounds that it seems unfair to tax all citizens in order to lower the price of an item

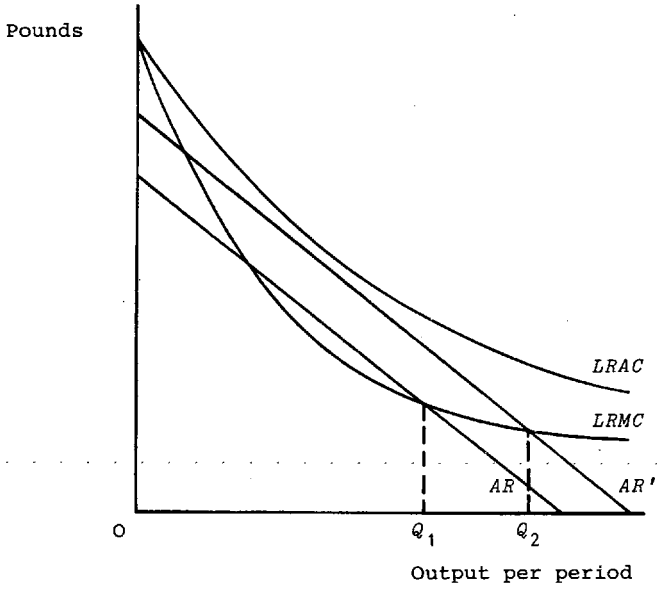

Figure 7.2

that may not be consumed by all citizens. Thus no general rule can be given and it is necessary for the subcentral authority to consider the options for each case on its merits.

Secondly, there is the possibility that the production or consumption of the enterprise's product yields external benefits or costs. If subcentral authorities are responsible for the correction of externalities, then they should approach the externalities created by enterprises owned by them in the same ways as those created by other producers, as outlined in the next section. In general, this means requiring the enterprises to set prices equal to marginal social costs rather than marginal production costs.

Thirdly, there are cases where demand is too low for total benefits to exceed total costs, even at the usually optimal output level where price equals marginal cost.[3] Figure 7.2 shows how the situation would arise for a firm that faced increasing returns to scale, although a similar situation could arise if the firm operated under conditions of decreasing returns to scale. In Figure 7.2, demand is represented by the curve *AR* and marginal costs by the curve *LRMC*. Marginal cost pricing would result in the output OQ_1 though it is clear that the total benefits at this point (represented by the area under the demand curve) are less than the total costs (represented by the area under the marginal cost curve). The implication in this situation is generally that production should be discontinued. In some cases, however, this conclusion may be felt to be inapplicable. For instance, a subcentral authority may own a local monopoly or oligopoly that it regards as part of an infant industry, or it may own one that produces a merit good that it feels should be subsidised

anyway; merit goods are discussed further in a later section of this chapter. It should be observed that the situation discussed here does not necessarily arise just because the demand curve fails to intersect the average cost curve (*LRAC*). For instance, had the demand curve for the firm in Figure 7.2 been AR', then the optimal output level would have been OQ_2, not zero, because at OQ_2 total benefits (given by the area under AR') exceed total costs (given by the area under *LRMC*).

It may be noted that subcentral authorities should generally make decisions about the prices of any private monopolies or oligopolies subject to regulation on similar principles to those made in respect of the prices of enterprises that they happen to own. This means that subsidies may sometimes be paid. If the subsidies were large, however, then public ownership might eventually be sought.

The finance of externality correction and public good provision

The second type of justification for government interference in resource allocation occurs when the remaining assumptions needed for an efficient allocation are not met. Principal among these are cases where there are externalities, or one or both properties of public goods, or where the costs of price-exclusion exceed its benefits.

It is perhaps helpful to regard cases with externalities as intermediate between pure public goods and private goods.[4] With private goods, all consumption is rival and excludable, while with pure public goods it is all non-rival and non-excludable. Where externalities arise, the production or consumption of an essentially private good gives rise to unintended by-products that are typically non-rival and non-excludable. In most, if not all, cases where externalities arise, the private benefits are sufficiently large that production will take place without government intervention; in the absence of intervention, though, the level of output will probably not be optimal. The general objective of governments will be to seek to secure output levels where prices equal marginal social costs – that is, marginal private costs (to the firm) plus marginal external costs or minus marginal external benefits. This objective will generally involve trying to increase output where there are external benefits and trying to reduce it where there are external costs. Note, however, that where production by a firm not in perfect competition creates external costs it may actually be desirable to stimulate an increase in output. The reason is that such firms may well produce where marginal revenue equals marginal private cost, which is at a lower output than where average revenue (or price) equals marginal private cost, and so may be lower than the optimum, where average revenue equals marginal private cost plus marginal external cost. In general, governments are likely to seek to reduce output by means of pollution taxes or charges or by means of regulations, and they are likely to seek to increase output by means of subsidies or public provision of the services concerned.

Since the objectives of externality correction are clearly defined, it is a function for which the central government may assume most

responsibility. If subcentral authorities do have any responsibility for dealing with externalities, then they will incur substantial expenditure only where subsidies are given to stimulate the production or consumption of items whose production or consumption generate external benefits or where these items are publicly produced at subsidised or zero prices. In principle, the authorities could seek to finance these subsidies by levying charges (or perhaps special taxes) on those who benefit. This might be held to be fair on distributional grounds, although there is no need for such charges on efficiency grounds.

It is possible that public goods will not be produced in the absence of intervention; if they are produced, then it is probable that the quantities will be well below the optimal levels. It is for this reason that governments are in practice likely to consider producing the goods themselves rather than subsidising private producers. In principle, too, public provision is preferable to subsidised private provision because, unless the subsidy meets the whole cost, those particular private citizens who are stimulating private provision will not face a zero price, yet, as explained in Chapter 1, a zero price is the efficient one for goods that are non-rival. If the subsidy were to meet the whole cost, then clearly there would be no reason why the original purchasing citizens should continue to have any more say in the level or type of production than other citizens. So, while subcentral authorities generally play an important part in the production of public goods, there is no efficiency case for supplying them at any price other than zero, although, as noted in Chapter 1, it might be possible to oppose free public provision on fairness or equity grounds.

Subcentral authorities may also be concerned with providing services where price exclusion is possible, but where the benefits of exclusion are felt to be less than the costs. Clearly there is an efficiency case for zero prices for all such services, although, again, equity considerations might result in no public intervention at all.

The finance of stimulating merit good provision

Another reason why governments might interfere with resource allocation results from their concern with merit goods and demerit goods. The resulting action is likely to include taxes or regulations to discourage or ban the consumption of demerit goods, and publicity campaigns or regulations to encourage or force the consumption of merit goods. Consumption of merit goods can also be stimulated by subsidies, however. This approach is often used when it is felt that underconsumption is a result of many households having low incomes, and government subsidies of merit goods, or government provision of such goods at subsidised (or sometimes zero) prices, can generally be seen as part of the redistribution function rather than the allocation function, and the former function must now be considered. It should be stressed that, while redistribution is usually considered as a policy to benefit the poor, it may also be used to benefit certain specific groups, such as the old, the disabled or armed service veterans.

The use of subsidies as part of a redistribution policy is itself rather

curious. Musgrave and Musgrave (1976, p. 65) have pointed out that, if the purpose of the subsidies were merely redistributive, then it could be secured better by making cash transfers, allowing the recipients to use the cash in line with their own preferences. They suggested that the decision not to have cash transfers is made because redistribution is a voluntary process, and donors are unwilling to make cash grants that may be used without the donors' control. Foster *et al.* (1980, pp. 512–13) have expanded on this view that redistribution is intended to increase the utility of the donors. They argued that there may be certain states among their fellow citizens that offend donors. One of these states may be poverty itself, and this can obviously be tackled by cash transfers. Other states might include illness or slum dwelling, but attempts to alleviate the offence caused by these conditions may fail if they involve cash transfers, since recipients may spend the transfers on other items. Arguably, the conditions are more likely to be eased if redistribution is tied to the specific conditions that upset the donors.

There are, in fact, at least three ways in which redistribution can be tied to specific conditions. Suppose, for example, that the condition concerned is rented slum property. First, the relevant central or subcentral authority could provide higher-quality housing at subsidised (conceivably zero) rents. Secondly, the authority could subsidise landlords with a view to encouraging a fall in rents and an increase in the quality of housing. Thirdly, the authority could issue subsidies to the relevant consumers on condition that these were spent on housing; the most frequent suggestion is that recipients would be given vouchers worth specified sums that could be given to housing suppliers in part or full payment of rent, and redeemed by these suppliers in cash from the authority.

It is important to discuss the relative merits of cash transfers and these three alternatives, because decisions about the subcentral authority provision of merit goods and the charges set for them depend critically on which option is selected. In the case of housing, for instance, the authorities might feel obliged to supply publicly owned properties for rent only if the option of subsidised public housing alone was used, and in this case the charges or rents set would obviously be below market rents. If the authorities used either of the other options, then they might supply no housing themselves; if they did supply some, perhaps to increase competition in the housing market, then it would be expected that they would set market rents if cash transfers or vouchers were being used, and subsidised rents if private landlords were also being subsidised. The following paragraphs contrast the different policies, but it should first be stressed that, although each system could be used in an attempt to concentrate aid on the poor, each could instead be used in an attempt to help all consumers. Thus public provision of housing could attempt to provide homes to satisfy all income ranges, or just homes of the sort demanded at the lower end of the market. Subsidies could be given on all property in equal proportional (or absolute) amounts, or they could be concentrated on cheaper property. Vouchers or cash transfers could be given in equal values to all citizens, or they could be taxed or means-tested (perhaps with reference to income tax returns) to discriminate in favour of the poor.

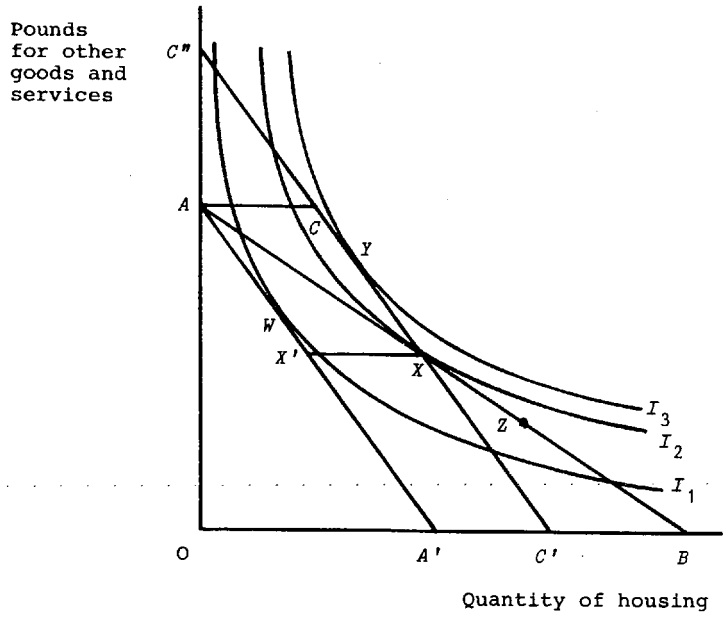

Figure 7.3

A major factor in deciding which of these methods to use will be how effective they are in stimulating consumption of the service concerned. One common way of comparing subsidies to producers and vouchers is shown in Figure 7.3, which uses the example of housing (and relates to a particular period of time). Suppose that without government intervention a citizen could purchase any combination of housing and other goods and services shown by the budget line AA', and would select point W on indifference curve I_1. A subsidy to producers, or subsidised subcentral provision, might produce the budget line AB and take the recipient to X on I_2. By contrast, a voucher whose value equalled the value of the subsidy ($X'X$) would create the budget line ACC' and take the recipient to point Y on I_3; in other words, the voucher here results in less consumption of housing, although it produces a higher level of satisfaction for the citizen.

It is important to realise, however, that these results might not apply to all citizens. To understand the reasons for this, and to obtain a clearer comparison of the effects of subsidies to producers and of vouchers, it is useful to imagine a community with two citizens eligible for housing aid of some sort. These citizens (α and ß) may be supposed to have equal incomes and family circumstances. Their situations are depicted respectively in Figures 7.4 and 7.5, which are analogous to Figure 7.3. In the absence of aid, each would face the budget line AA'. Each would select the point on AA' touching his indifference curve labelled I_1, but it will be seen that α consumes far more housing than ß.

If housing producers were subsidised, then each citizen might face the budget line AB and move to the point where this touched his indifference curve labelled I_2. As drawn, the figures suggest that α will end up

consuming more housing than ß, and also that his consumption will increase rather more than ß's. These results are certainly plausible. A consequence of them is that the subsidy paid in respect of α (shown by $X'X$ in Figure 7.4) is greater than the subsidy paid in respect of ß (shown by $X'X$ in Figure 7.5).

The critical question is whether the subcentral authority concerned would give a larger voucher to α than ß if it replaced its subsidies with vouchers. If it did not, then its vouchers would presumably vary with individuals' spending levels and might formally be very similar to the subsidies. In fact, though, the authority might well decide that voucher values should depend on need rather than spending levels. Accordingly, if α and ß had similar family circumstances, as well as similar incomes, then each would qualify for a voucher with the same value. Figures 7.4 and 7.5 assume for convenience that this is midway between the two $X'X$ subsidy values shown in them. These vouchers put α and ß on the budget lines ACC'. α and ß consume at the points where these touch their indifference curves labelled I_3. As drawn, the figures once again show α consuming more than ß. The most interesting results, however, are that α prefers the subsidy to the voucher, while ß consumes more housing with the voucher than the subsidy.

Figures 7.4 and 7.5 are drawn for particular individuals with particular indifference maps. They suggest, though, that it cannot be proved *a priori* that the replacement of a scheme of subsidies with a scheme of vouchers

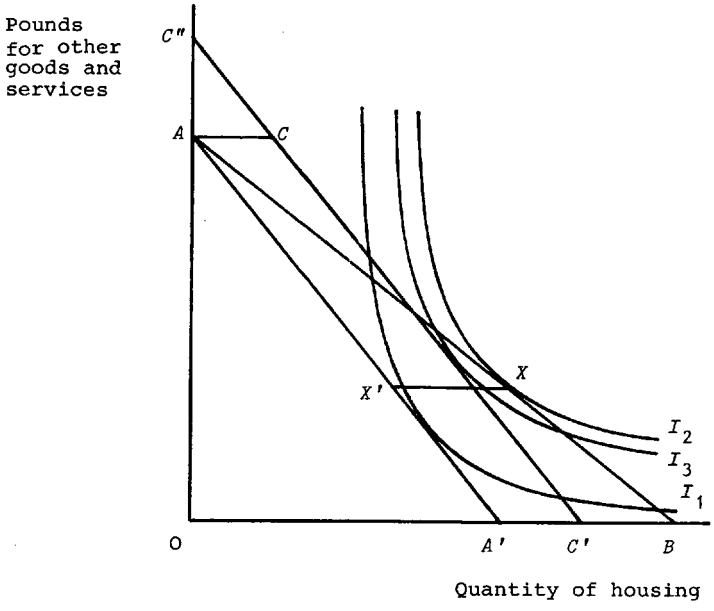

Figure 7.4

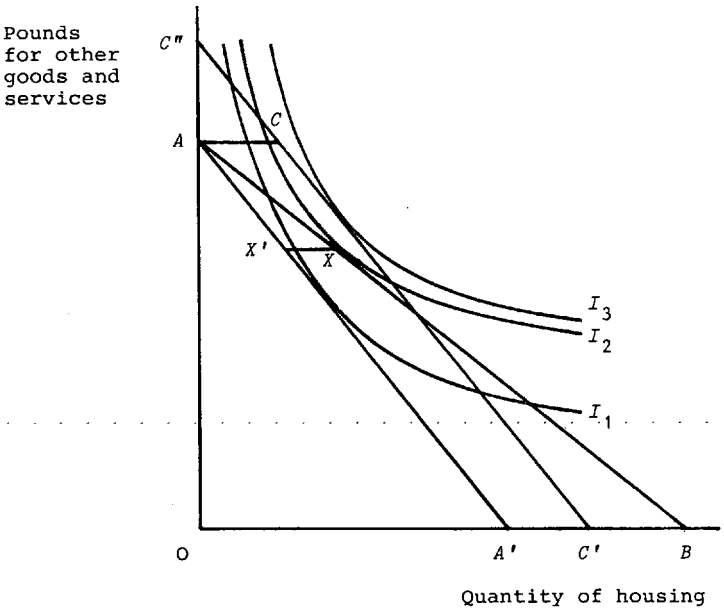

Figure 7.5

at equal total cost to the public authority concerned would necessarily lead to an aggregate fall in housing consumption. More importantly, they suggest that even if the policy switch did lead to lower aggregate housing consumption it might nevertheless increase the consumption by those such as ß who previously consumed little, albeit at the cost of lower consumption by those such as α who previously consumed much. Thus, if the primary objective of intervention is to reduce or eliminate the number of people who are inadequately housed, then the voucher system might seem the more attractive policy, for it is likely to do more to help those who, like ß, start off with the lowest consumption levels.

The effects of cash transfers on recipients may often be the same as those of equal value vouchers. In Figures 7.3, 7.4 and 7.5, vouchers produced the budget lines labelled ACC', because they would not enable citizens to consume more than OA pounds of other goods and services. Equal value cash transfers would produce the budget lines $C''CC'$, but in each case would result in recipients choosing the same position as the vouchers, namely where CC' touches I_3. Cash would have a different effect from vouchers only in the situation depicted in Figure 7.6, where AA', $C''CC'$ and indifference curve I correspond to AA', $C''CC'$ and curve I_1 in the previous three figures. In Figure 7.6, the chosen point with cash is at Y, which is on the $C''C$ section of $C''CC'$ and also on indifference curve III. It is therefore preferred to the best point available on ACC' with vouchers, namely C, since C is on the lower indifference curve II; in

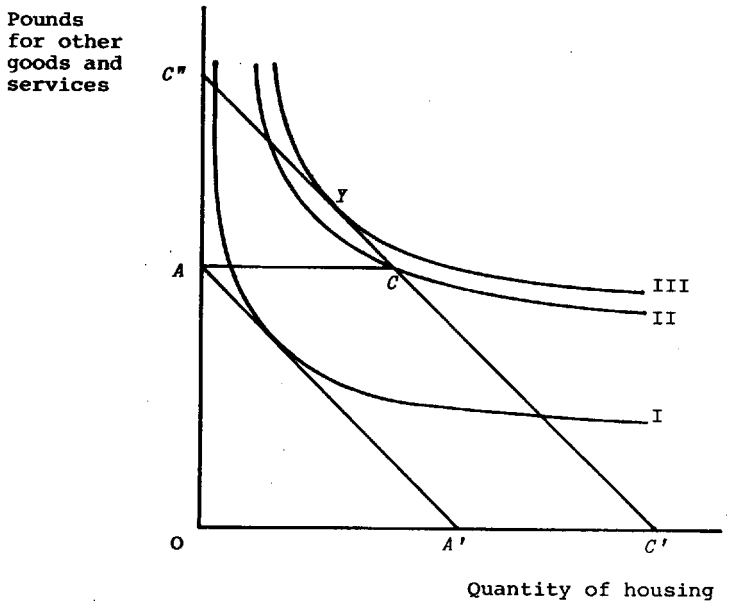

Figure 7.6

this perfectly possible situation, vouchers increase housing consumption more than equal value cash transfers do.

It appears, then, that vouchers may be a surer way of stimulating consumption than equal value cash transfers. It is not certain whether a system of vouchers whose values depended on need would stimulate aggregate consumption more or less than a system of subsidies whose total value was the same as the total value of the vouchers. Even if, in a particular case, subsidies were more effective in raising aggregate consumption, vouchers would be likely to stimulate low consumers more, and so result in lower minimum consumption levels. One way of trying to make subsidies more effective, especially for citizens whose tastes induce them to be low consumers, is to confine them to publicly provided houses and make them available only if the recipients consume stipulated quantities. Both Robinson (1979, pp. 117–18) and Stafford (1978, pp. 53–4) have considered the case (shown in Figure 7.3) where subsidised housing at the rate reflected by AB is available only upon the condition that the recipient consumes the quantity indicated at Z rather than the preferred quantity indicated by X. Provided Z is outside the indifference curve I_1, the quantity Z will be consumed. In this case, the constraint increases consumption beyond the level that would be selected at X with an unconstrained subsidy, even though the consumer would prefer position X to Z. Of course, the consumer gets more subsidy at Z than he would

at X, and so constrained subsidy schemes are likely to cost more in aggregate than unconstrained ones with similar subsidy to rent ratios. Whether a constrained subsidy scheme will secure higher minimum consumption levels than an equal value voucher scheme is a question that could only be answered by empirical work. A point to stress is that the constraint ought, ideally, to be low enough to ensure that no low-taste consumers preferred to ignore the subsidy altogether.

One final point that is sometimes made[5] when comparing these various systems is the allegation that vouchers and cash transfers may in fact give little benefit to the recipients in terms of increased consumption of housing because the elasticity of supply of housing is low; accordingly, their chief result in the housing market may be an increase in rents. This effect is applicable only in the short run, however, and it may be added that subsidies, too, may not be able to result in increased quantities of housing in the short run.

There are perhaps five further criteria by which the various policies may be judged. First, there is the certainty with which they will help those whom they are intended to help. The risk with subsidised prices is that in some cases, such as subsidised school fees, they may still be beyond the means or the desires of the recipients, whereas vouchers might always be spent. On the other hand, the administrative arrangements used to distribute vouchers or cash transfers may be imperfect and omit some of those entitled to them. This risk is greatest if the vouchers or transfers are subject to a means test and if some of those eligible prefer not to be so tested; it is perhaps smallest if the vouchers or transfers are tied to the consumption of a service such as primary school education that is made compulsory.

Secondly, there is the choice of service available to the recipients. This will probably be most constricted when free or subsidised public provision is used. Seldon (1977, p. 40) has argued that there are two reasons for this. On the one hand, politicians and bureaucrats can never fully know the diverse preferences of thousands of different citizens. On the other hand, the citizen who is unhappy with the authority's services, such as a hospital or school, may be unable to change things except by persuading thousands or millions of other citizens to share his view. Note that the citizen could, of course, move to another area, although this would be costly. It might nevertheless be the only option – even with subsidies to private producers, vouchers or cash transfers – if the service concerned exhibited increasing returns to scale and resulted in local monopolies.

Thirdly, there are the possible benefits that will occur if services are supplied by competitive producers as opposed to monopolistic government agencies. In the case of schools, for example, Peacock and Wiseman (1964, p. 46) have argued that the near monopoly position enjoyed by the public education authorities in the United Kingdom has given little impetus to cost saving or product improvement. This argument suggests that subsidies to private suppliers, vouchers or cash transfers would be better than subsidised or free supplies by governments.

Fourthly, there is the question of fairness. Bird (1976, pp. 35 and 11) has suggested that fairness requires that those who consume the largest

amounts of a service should pay the most. This is clearly not the case with free provision and, though it is the case with subsidised provision, it might be argued that differences in consumption levels are not fully reflected by differences in payments. Fairness in this sense is secured by vouchers or cash transfers, provided there is no means test. If, however, fairness were restricted to a view of horizontal equity that argued that two citizens with equal incomes and family circumstances should receive equal financial help to encourage them to consume merit goods, then fairness would be achieved by vouchers or cash transfers whether means-tested or not; it would not be achieved by subsidised or free provision since the degree of financial help would depend on the quantities chosen and thus on tastes. An implication of this point is that, if subsidised or free provision of any service were replaced by unsubsidised services and vouchers or cash transfers, then some redistribution would take place in favour of low users at the expense of high users.

The fifth consideration is which method is most likely to lead to an efficient quantity of the relevant services being provided. There are perhaps four main reasons why free provision by governments may not be successful here. First, the government concerned will have great difficulty identifying what levels and types of services people want. Discussing this point, Bird (1976, p. 34) has noted that, without prices, public sector officials are deprived of an important objective indicator of how resources should be used. Secondly, as Foster *et al.* (1980, p. 509) have pointed out, the absence of a price means that the consumer cannot take into account the resource costs of the alternatives open to him when he decides to consume. Bird (pp. 34–5) has suggested that in such a situation there will be a constant pressure for more expenditure on the activity, since potential beneficiaries perceive little or no direct cost to themselves in such expansion. Foster *et al.* (p. 509) have argued that the quantity of the services demanded will continue to increase until the marginal utility derived is zero, and Seldon (1977, p. 27) has contrasted the thoughtless use made of unpriced services (such as an employer's stationery or the United Kingdom's National Health Service) with the thoughtful shopping done in supermarkets. In answer to this point, it is sometimes claimed that some price elasticities are zero, and hence quantities demanded will not always be higher under free provision. Prest (1982, p. 12) has noted that this is an empirical matter, and it may be observed that there is, to say the least, no evidence that it is often the case. Thirdly, the absence of prices means in Seldon's view (1977, p. 25), that when people differ about how resources should be used then they argue and debate and, in political democracies, organise political parties to gain power and compel minorities (or even majorities) to accept their views. Pricing, by contrast, is a peaceful way of resolving argument and conflict, and therefore, perhaps, results in more satisfactory outcomes. Fourthly, while free provision may stimulate quantities demanded, it is likely to constrain supplies. Harris and Seldon (1976, pp. 77–8) have argued that individuals are more willing to pay directly for services that benefit their families than they are to pay indirectly through taxes for similar services. Such supply constraints offer a counteracting force to the demand stimulus, and indeed

Seldon (1977, p. 148) has suggested that they often cause public services to seem shabby and scrimped in contrast to private services. Even if the forces were to combine, by chance, to secure an approximately optimal level of provision in any particular service, congestion or queueing would be likely.

All four problems outlined in the last paragraph explain why free provision is unlikely to lead to an efficient level of output. It remains to see how far these strictures apply to subsidised provision (by public or private producers), vouchers and cash transfers. With the first problem, subsidised provision and vouchers (but not cash transfers) will also lead to difficulties for governments in deciding how far to stimulate service provision, since demands are made by consumers whose payments may not reflect the real values they place on the services concerned. As for the second problem, subsidised provision, vouchers and cash transfers are all likely to stimulate the quantity demanded; with the last two, however, consumers who spend more than the value of the voucher or cash transfer will pay the full price of extra purchases and should stop buying at a lower level than they would in a system where prices were subsidised. With the third problem, there may also be arguments about the level of the subsidy, voucher or cash transfer. However, these three problems may all be less acute than they would be with free provision, and they should be least serious when the subsidies, vouchers or cash transfers are relatively small compared with the total expenditure on the merit good concerned. Note, too, that allowing citizens to spend more to benefit themselves should largely circumvent the fourth problem of supply constraint and shabby services, although some voucher proposals might prohibit 'topping up', in which case this advantage compared with free provision would not apply.

Taken together, the various criteria discussed in this section suggest that cash transfers or vouchers that can be presented to private producers may often be the best methods of stimulating the consumption of many of those merit goods whose consumption by people on low incomes would otherwise be felt to be too low; of the other two methods, subsidised provision may often have fewer problems than free provision. It is probable that the same conclusions would apply also to any merit goods whose consumption was thought to be too low in the absence of intervention because of consumer ignorance or because of a dichotomy between the consumer and the purchaser.

It is worth concluding this discussion by recalling from Chapter 1 that it is not generally considered appropriate for subcentral authorities to play a major part in the distribution function. Indeed, if redistribution were carried out wholly by taxes and cash transfers, then subcentral participation would probably be very modest. If redistribution were carried out in part by stimulating the consumption of relevant merit goods by vouchers or subsidies to private producers, then, again, the role for subcentral authorities would be likely to be modest. Only if it was finally decided that the consumption of some merit goods should be stimulated by having subsidised or free public provision would there be a case for a substantial subcentral role, principally so that variations in tastes could be catered

for, although the central government might well wish to ensure that the degree of redistribution was reasonably similar in each area.

The finance of subcentral stabilisation policies and law-making policies

It was noted in Chapter 1 that subcentral authorities might participate in the stabilisation function by handling measures designed to reduce unemployment in areas where it was relatively high. They might also have some law-making functions. In neither of these cases, though, would charges appear to be very relevant. Rather, it appears that these policies should be financed through taxation.

The main arguments in favour of increased charges by subcentral authorities

The discussion in this chapter so far suggests that those campaigning for increased charges by subcentral authorities are likely to rely principally on one or more of the following arguments. First, they may claim that subcentrally owned enterprises have charges that are too low. This argument could be used wherever it was believed that the prices set by an enterprise were below its marginal costs (both private and external combined), but it is perhaps more likely where the enterprise is in receipt of a subsidy that is believed to exceed that warranted by a situation of increasing returns to scale or external benefits. Secondly, they may claim that some services provided free on the grounds that they are public goods are not really public goods. In this case, the appropriate action may be to operate them as trading services with suitable prices, or, perhaps, move to a situation where the service was provided by private firms. The latter option would probably be recommended unless it was likely to lead to the promotion of a monopoly, since it might increase consumer choice and lead to competitive cost saving. Thirdly, they might oppose free provision of goods with public goods characteristics on the grounds that the efficiency gains were outweighed by equity losses, in that budgetary finance would require all citizens to help finance services from which not all might benefit. Fourthly, they might argue that charges should be imposed on items presently supplied free in the belief that the costs of price-exclusion would exceed the benefits. Such arguments could be based on the view that the costs need not exceed the benefits or on the view that it is fairer for any item to be paid for by the particular citizens who consume it than by taxpayers in general. Fifthly, they may claim that a number of merit goods are being supplied by subcentral authorities free, or at subsidised prices, and believe that if it is desired to stimulate the demand for these particular items then the best mechanism is one of cash transfers or vouchers paid for directly by the central government, so that subcentral provision of such items should either be terminated in favour of private provision, or at least compete with private provision on equal terms. Alternatively, they may accept that subsidised public provision is the best arrangement, but feel that present subsidies are too high.

It might be thought that most of these issues, if not all, would have been sorted out long ago in all countries. However, there are reasons why regular reappraisals are needed. For instance, some services may be provided free on the grounds that price-exclusion would be not so much impossible as so costly to operate that any benefits from pricing would be less than these costs; in time, changing technology could reduce these costs to the point where private provision was judged to be preferable. Some free or subsidised services may have been introduced for merit good reasons in preference to cash transfers or vouchers on the grounds that there would be large administrative costs with the transfers or vouchers and that many of those entitled to them might not receive them; changing technology and a trend for more and more citizens to file income tax returns might greatly reduce these problems. Then there is the possibility that some services, or, more usually, parts of some multi-aspect services such as education and police, might have been introduced in the belief that they were public goods or merit goods, whereas subsequent appraisal suggests that their benefits are rival or excludable or that they have negligible redistributive impacts in the ways intended. Many of the proposals for raising charges for specific services stem from considerations such as these.

A number of these proposals for raising charge income are considered in the remaining sections of this chapter. It is not possible in the present text to explore in great detail the extent to which charges are appropriate for each service that is somewhere provided by a subcentral authority. Rather, the following sections consider some of the main arguments in favour of using charges to finance part or all of some of the most common and most important subcentral services, and they focus on cases where suggestions that reliance on charges be increased are most widespread.

Perhaps the most radical set of proposals is the one that was put forward by Foster *et al.* (1980, pp. 528–36) for the United Kingdom, although their arguments would apply to subcentral authorities in many countries.[6] They argued that if redistribution was intended to encourage the consumption of certain items, then it should rely on vouchers. They also claimed that other arguments for free or subsidised provision, including those relating to public good considerations, apply to no more than 10 per cent of local authorities' current expenditure. Since vouchers would be exclusively part of a redistribution policy, it would be expected that they would be financed by the central government. They argued that the existing local authority tax would be more than sufficient to cover the cost of the remaining services requiring tax finance, and that the need for revenue-sharing grants would disappear.

Charging for education

The most costly service for many subcentral authorities is education, and the most expensive parts of this service are primary and secondary schools. Education in such schools is frequently offered free. This might suggest that education is seen by governments as a public good. Foster *et al.* (1980)

did not include it in their list of public goods, however, and many other writers in this field do not even give this possibility any thought.[7] Seldon (1977, p. 62) considered the possibility, but argued that education can be supplied to individuals separately, implying that it does not exhibit non-rivalry, and that it can be refused to people who do not pay, implying that it is an excludable service.

Another possibility is that education is a merit good provided free because some parents might not purchase it for their children, either because of ignorance or lack of care, or because of low incomes. As Ziderman (1980, pp. 149–51) has noted, however, the most effective remedy for ignorance or lack of care is to make schooling compulsory up to a certain age, for there is no guarantee that all parents will send their children to school as opposed to work merely because schooling is free; such compulsion is, of course, widespread. The only action that can be clearly warranted on the grounds that some parents are too poor is redistribution through cash transfers, vouchers or subsidies; if this redistribution were sufficiently generous to enable recipients to purchase reasonable-quality education without needing to use much of their own incomes, then there would be no need on merit good grounds for schools to be run by the state.

Another explanation for free state provision might be that it is a response to possible external benefits in a situation where demand has a low price-elasticity. It would be a remarkable coincidence, though, if zero prices were always the ones needed to stimulate precisely the target level of consumption. Blaug (1967) distinguished between economic spillovers, in which the extra education of any one individual would increase output by more than it would increase his own earning power, and other 'atmospheric' spillovers, in which education would lead to gains in national solidarity and political stability, and, in Ziderman's view (1980, pp. 147–9), a more enlightened, literate and cultured society. Blaug claimed that no one has demonstrated that externalities of the former type exist. In general, if spillovers do exist, then Seldon (1977, p. 62) argued that the correct procedure is to stimulate production. He suggested that this be done by vouchers or cash transfers (though subsidies could be used instead), possibly combined with a mandatory school leaving age. He did not consider that the possibility of externalities justifies monopolistic state provision.

It is arguable, however, that one particular atmospheric externality – namely, social unity or cohesion – would be created more effectively if such a monopoly existed. Ziderman has explained (1980, p. 151) that it can be argued that a uniform state education system is a powerful tool for creating social cohesion through common values; but, like Peacock and Wiseman (1964, pp. 55–6 and 24), he felt that this may be much more important in countries like the United States with large immigrant and racial minorities, or in poor countries rent by linguistic or tribal divisions, than in countries such as the United Kingdom. Moreover, Peacock and Wiseman pointed out disparagingly that public provision of education makes the imposition of values easier, citing Nazi Germany as an example of a country where public education was used to further the ends of a ruling group, and they found it hard to see what justification there could

be for the continued existence of a society that cannot withstand the impact of a diversified private education system, provided the public authorities have some overall responsibility for standards.

Blaug (1967, pp. 43–6) has suggested that the main need for cohesion in the United Kingdom is between social classes, and suggested that uniform provision would ease the problems of class distinction. However, Peacock and Wiseman (1964, pp. 56–7) felt that if vouchers could be used at all schools then the intake of the existing independent schools would be much broader and they would lose their 'separate' character; similar consequences might follow if cash transfers or subsidies were used instead. If a uniform system were desired, though, then it would seem curious to entrust it to subcentral authorities rather than to the central government, for the latter could much more easily ensure uniformity. It is impossible to say how effective a state monopoly is in promoting cohesion, as compared with other systems, without conducting experiments on alternative systems in different areas, but it is possible that the idea of supplementary controls on parental choices, noted by Katzman (1972, p. 381), would ease the problems that private supplies might create.

Another possible reason for free state provision is that other systems could create an external cost – namely, inequality of opportunity – if some parents spent more on educating their children than others. Ziderman (1980, pp. 151–6) regarded this as the central issue, but it is possible that this problem could be largely circumvented if private schools were required not to set fees in excess of a specified limit, a possibility noted by Seldon (1977, p. 70). However, if equality of opportunity were to be pursued more rigorously than this, then central provision would be needed, since subcentral provision might well lead to different standards of schools in different areas.

It is clear that there is considerable disagreement about the need for state intervention in the provision of education. The implication of many of the arguments just considered is that the main need is for compulsory schooling up to a certain level or age. In addition, it is arguable that maximum school fee levels should be set, and it is arguable that producers should be subsidised if education can be shown to have external benefits, or if it is decided to help the poor in this way rather than with vouchers or cash transfers. There appears to be little case for subcentral authorities to provide schools, but if they do, perhaps to increase competition, then it seems that they should receive subsidies on the same basis as other schools and rely on charges for most of their income. However, if it is believed that the social cohesion or equality of opportunity arguments warrant state provision of all schools, then the provision might well be made free, but should arguably be a central responsibility.

Charging for roads

A common approach to the provision of roads is to argue that where roads are uncongested then consumption is non-rival, and marginal costs are zero, so that it would be inefficient to levy a charge that would curtail

their use.[8] It is widely recognised, though, that where roads are congested then a motorist who uses them imposes congestion costs on other users, and it would be efficient to impose a charge to equal the marginal congestion costs thus entailed. Various writers have considered ways in which motorists could be made to pay these costs, and their suggestions include substantial parking charges, or special licence fees in congested areas, or differential peak and off-peak tolls on toll roads.[9]

A more detailed analysis of the various costs of road usage has been given by Roth (1965), and more accurate ways of making motorists pay these costs have been examined by the Smeed panel (HMSO, 1964,). Roth argued (1965, p. 13) that the use of a road imposes private costs, direct road costs, community costs and congestion costs. Private costs are paid for directly by road users and do not give rise to any special problems. Direct road costs include maintenance, policing and road administration; it is arguable that these result in a positive marginal cost for each mile travelled, and Roth suggested (pp. 16–18 and 55) that they could most suitably be charged to motorists through a fuel tax, although the rate of tax suitable for private cars would not exact enough revenue from heavy vehicles, so he felt the latter should pay suitably high licence fees. Community costs arise from the noise, fumes, danger of accident and loss of amenity; Roth ignored these, partially (pp. 21–3) because of the difficulties of evaluating them and of evaluating the marginal community cost of any journey, but also out of doubts about how far road users should in principle bear these costs. Congestion costs arise when road users impede the movements of others; Roth (pp. 18–21 and 54–5) believed these would exist even on rural roads, and suggested that they could be charged to their users through an addition to the fuel tax, although he would have preferred them to be charged through an extension of the methods suggested for towns by the Smeed panel if this were possible.

This panel (HMSO, 1964, pp. 17–33) described no less than six meter systems (two manual and four automatic) that it thought could be made effective. One possibility is that each motorist would have to fit a road meter with a lamp to his vehicle. The power for the lamp would be fed through an electrolytic timer, which would eventually run down and have to be replaced. In congested areas, the lamp would have to be switched on, and motorists would need to buy new timers from time to time. These would be bought at garages or post offices at prices well above their production costs, the excess revenue being paid to the public authorities. Roadside equipment could be used to control the meters when vehicles entered congested areas, and this equipment could arrange for timers to be used up more quickly in the most congested zones. Although such proposals seemed rather science-fictionish to Bird (1976, p. 61), it is probable that recent developments in technology would make them more feasible today than they were when suggested by the Smeed panel in 1964.

In the absence of adequate data on whether the transport facilities provided by roads were produced under conditions of constant, increasing or decreasing returns to scale, Roth (1965, p. 30) assumed that the returns were constant, which would mean that average and marginal costs would be equal. In such circumstances, a system of marginal cost charging for

the direct road costs imposed by vehicles (collected through a tax on motor fuel and licences on heavy vehicles) should cover all the current costs incurred in making roads satisfactory for vehicular use. Roth (pp. 33 and 56–63) felt that pedestrians and cyclists should make a contribution to road costs through local taxes, and, by considering the expenditure on roads in the days before the advent of the motor car, he estimated (for the United Kingdom in 1964) that a 'reasonable' contribution from this source would just exceed half the total costs of roads and so exceed the contribution required from motorists.[10] On this basis, a tier of subcentral authorities that provided roads and sought to levy appropriate fuel taxes and licence duties on heavy vehicles would still be expected to meet a large slice of their current expenditure on roads from taxes. If they also levied congestion charges, however, then they could well find that the total charge income exceeded their current expenditure on roads. It must be said, though, that there is no particular efficiency or equity case for permitting any congestion charges that are introduced to be levied by the same authorities that provide the roads. Indeed, it might be felt that such an arrangement would discourage some authorities from undertaking road improvements designed to reduce congestion.

Charging for housing

The renting of publicly owned property at subsidised rents is probably best seen as part of a redistribution policy that seeks to stimulate the consumption of certain merit goods of which housing is one.[11] It is worth stressing that the subsidies may be offered chiefly for redistribution reasons and not solely because seeing people in slum property offends the rich; if this external cost were the only problem, then Friedman (1962, p. 178) argued that the appropriate remedy would be a tax on those living in the property who generate the externality! The main arguments for and against the various methods of stimulating the consumption of merit goods such as housing have been outlined earlier in this chapter, and little more need be said here.

Gray (1968, pp. 47–9 and 79) has noted that any cash transfers or vouchers used need not be of sufficient value to enable recipients to pay market rents without using their own funds, because they would not be likely to replace a system of free housing, and he argued that assistance should vary from area to area to reflect variations in local rent levels. He also thought that new measures could be introduced to enable those on low incomes to purchase their own homes. Bird (1976, p. 231) has argued that replacing of subsidies on public housing with subsidies to people would have the twin advantages of helping those who would otherwise be waiting in the queues that often exist for public housing and of providing a considerably wider range of choice to those whom the policy was designed to help. The existence of queues is not surprising if the public sector offers housing at lower rents than the private sector, because this will increase the demand in the public sector, which it may take a long time to meet. The existence of queues, though, forms another problem, noted by Seldon

(1977, p. 91), in that it intensifies the immobility of labour, since moving to another area may entail joining another queue or moving into unsubsidised accommodation.

If subsidies on publicly owned houses were replaced by cash transfers or vouchers, then the money needed might well come from the central government which is generally regarded as the authority which should assume prime responsibility for distribution policy. Initially, this would leave many subcentral authorities owning properties which would be let at unsubsidised rents. However, it is arguable that there would be no special case for continued public ownership of these properties, and that they should be sold, in which case subcentral authorities would have little direct involvement in housing provision. However, they might still have an important role to play in stimulating the replacement or improvement of properties in slum areas, since individual landlords might be reluctant to improve or replace their properties unless others do also; but while the authorities might need to subsidise the cost of improvements (as these may generate external benefits if they improve a property's appearance) and might need to acquire and clear sites where rebuilding is required,[12] it is not at all clear that such policies require them to own property in the long term or to subsidise rents in the short term.

Charging for police services

Police services are generally financed from taxes. Foster *et al.* (1980, pp. 529–30) considered that the benefits of most police services (namely, the protection of life, limb and property) that result from reductions in the crime rate are non-rival, and this alone makes a case for free provision on public good grounds, although it is also arguable that excludability would be hard. Certainly, if some citizens were to purchase police services while many others did not, then there could be substantial external benefits that it would be hard to handle efficiently. They also argued, however, that a small proportion of police activity (such as attending football matches) benefits a clearly defined set of people and should be priced; they noted that, in England and Wales in 1976–7, charges met over 6 per cent of police expenditure.

Reviewing some studies in the United States, Bird (1976, pp. 129–30) concluded that perhaps 10 or 20 per cent of police expenditure financed services for identifiable beneficiaries who could be charged, and that as much again was spent on traffic services. He argued that motorists as a group should pay for the latter, perhaps through licence fees or parking fees (although presumably fuel taxes could be used instead). Bird also mooted the idea that special police fees could be levied on enterprises (such as banks) that give rise to a high incidence of crime, particularly if they failed to take adequate security measures.

Overall, it seems that charges for police services might be increased. It is nevertheless clear that most police services must be financed through taxation.

Charging for fire services

Fire services, too, are often financed almost entirely out of taxation. Foster *et al.* (1980, pp. 528–31) found no public good characteristics in these services, while Bird (1976, pp. 135–9) and Seldon (1977, pp. 110–11) have justified public intervention on the grounds of externalities. In essence, a citizen whose property is ablaze often confers an external benefit when he has the flames extinguished because they might otherwise set fire to other property. On this basis, most property owners should subsidise the costs of neighbouring owners in insuring themselves against the need to employ a fire brigade, but most should simultaneously be subsidised by their neighbours. Bird concluded that these subsidies might largely cancel each other out, and argued that each citizen should pay the full costs on his property and, for the benefit of others at least, be obliged to take out the appropriate cover.

Now, if all buildings have to be insured, it could be argued that the necessary payments could be taken out of tax revenue and used to finance either a publicly produced service or one produced by one or more private suppliers on a contract basis. Bird and Seldon pointed out, however, that there would be advantages in levying a separate fire protection charge for each building. Such charges could relate more closely to the potential benefits by being based on the value of each building and its contents. They could also discriminate against thatched houses and wooden buildings and, more importantly, discriminate in favour of establishments where substantial fire precautions (such as fire doors, fire alarms and sprinkler systems) were installed and, thereby, hopefully encourage such installation and reduce the damage from fires. It may be added that neither Bird nor Seldon explored the extra costs that might be entailed in financing fire services from charges – the principal costs being the valuation of each building and its contents and an assessment of its flammability and its fire precautions. It is possible that such an analysis would indicate that the case for charges was weaker for domestic property than for business property.

Charging for refuse services

Seldon (1977, pp. 111–13) argued that refuse collection and disposal services are rival and excludable, although he conceded that one household's consumption usually generates external benefits to its neighbours by diminishing smells, unsightly piles of rubbish and the risk of disease. As with fire services, the benefits are reciprocal, which suggests that if costs were covered by charges then households might as well just pay their own costs. The main effects of using charges instead of taxes would be that households with the most rubbish would pay the most and, as Seldon pointed out, that the resources used in refuse collection and disposal services would fall, since households would have an incentive to separate combustibles from non-combustibles and burn the latter. In addition, Bird (1976, p. 116) suggested that charges might enable individual

preferences (such as more frequent collection) to be accommodated. One problem he noted (p. 117) with charges is that administering the efficient charge for each individual household might generate more costs than benefits. Another trouble with charges is that they might encourage the hoarding of rubbish, with the associated externalities; charges could also generate externalities in the forms of smoke and hazard as a result of encouraging bonfires. Feldman (1972, pp. 230–1) implied that it is these considerations that lead to tax finance in many areas, but charges are common in parts of both the United States and Europe. It is possible that there is a weaker case for using charges for refuse from domestic properties than for refuse from other properties. Whether an authority levies charges or taxes, it could use the proceeds either to finance a publicly produced service or to purchase services from one or more private contractors.

Charging for other services

Some remaining services can be considered more briefly. Kahn (1961) has argued that the purpose of welfare expenditures is clearly redistributive and concluded, therefore, that the relevant services cannot be charged. Conversely, Foster *et al.* (1980, p. 512) argued that free services are not needed to help specific groups such as the blind, diabetic, old and disabled because these people can be helped by entitling them to specific cash transfers. Taxpayers may, however, be prepared to finance such transfers only if they are spent on merit goods, and it is not clear whether Foster *et al.* would plea for guide-dog or insulin vouchers. If transfers or vouchers were to replace free provision, then subcentral authorities might well not be responsible for financing these services, and any supplies they made of them might have to compete with those of private suppliers.

Foster *et al.* (1980, pp. 533–4) argued, too, that prices should be used for art galleries, museums and parks, although in practice the benefits from charging admission to parks may not always warrant the costs of fences, walls and tickets. They wanted prices that equal the marginal costs, including both resource costs and congestion costs. It might be thought that marginal resource costs would be zero, but it is claimed that this is not so; in a museum, for instance, it is argued that increases in the number of users could raise expenditure on keepers, administrators and exhibits. It is nevertheless arguable that these facilities operate under conditions of falling marginal costs, even when congestion costs are included, and that some subsidy is needed that should be financed through taxation; indeed, efficiency requires subsidies to be paid in all such cases, whether privately or publicly owned. On the other hand, it might be argued that there is an equity case for finance through charging so that the costs of these facilities are met by those who enjoy the benefits.

Finally, consider the case of water supply. Water consumption is generally rival and excludable, and so water would doubtless be supplied without any public intervention. However, it is an item whose production could result in local monopolies and so need to be controlled or undertaken by subcentral authorities. The authorities might wish to subsidise the

supply of water if they think its consumption generates external benefits when it is used to keep people and property clean. They might wish to meet the whole cost of supplying water through tax revenue if they think the benefits of price-exclusion would fail to offset the costs. The relevant issues here were discussed in Chapter 1, and it will be recalled that the case for free public provision will be strongest in areas where exclusion costs are relatively high and demand is relatively inelastic. It may well be that water is presently supplied free in some countries where charging would be preferable, and vice versa.

The possibility of experiments with charging

This chapter has attempted to cover the main issues concerned in cases of common subcentral services where a greater reliance on pricing has been advocated. It is not the place for exhaustive discussion, and it has not sought to reach definitive conclusions. The increasing discussion of charges makes pilot experiments – advocated by many[13] – increasingly likely. Where these suggest that tax finance be replaced by charge finance, then the tax-revenue needs of subcentral authorities might, in time, decline. This chapter suggests also that, in some cases, the termination of tax finance may, sooner or later, be followed by the termination of public supply and a reduction in the number of subcentral authority functions.

Benefit taxation and earmarked revenues

It is appropriate to conclude this chapter with references to two methods of financing public services that have much in common with the charges considered here. First, there is the possibility of benefit taxation. Foster *et al.* (1980, p. 526) implied that, if a tax system were to impose levies that reflected the uses made by citizens of public services, then it would be a system of benefit taxation or, in effect, charging, although the levies would have to relate to the marginal costs of the services for them to be a perfect substitute for an efficient pricing system. However, such a system is a full substitute for a situation where services were efficiently priced only where those services were provided monopolistically by public authorities. Benefit tax finance of publicly provided services does not create the choice and competitive forces of private provision. Irrespective of the theoretical merits of benefit taxation, there remain great difficulties in designing any tax (as opposed to a charge) where levies correspond closely with the use made of any public facility; the most commonly cited example is that of motor fuel tax payments, which bear some relation to road use.

Secondly, there is the possibility of earmarked revenues. This term covers many possible situations. Bird (1976, pp. 23 and 26) argued that it would be most effective where the entire proceeds of an earmarked tax, and only those proceeds, must be spent by the authority that collects it on some particular activity, and where the rate of tax is variable. As Buchanan has explained (1963, and 1967, Chapter 6), such a system can help voters to make more intelligent decisions about the operation of the public sector

because it makes the costs of different services clearer. However, such a system will be most helpful in securing efficient levels of publicly provided services when the tax levies on individuals are closely related to the benefits received. As noted in the last paragraph, it is difficult to devise taxes that meet this criterion.

In principle, therefore, there are some good arguments in favour of citizens making payments to subcentral authorities that are linked to the benefits secured and used to finance the services concerned. In practice, it is much easier to devise charges that satisfy these requirements than it is to devise taxes that satisfy them.

Summary

There are a number of grounds on which subcentral authorities could seek to justify subsidised or free provision of goods and services. These include items produced under conditions of increasing returns to scale, items whose production or consumption yield external benefits, items with either or both characteristics of public goods, items where the efficiency benefits of price-exclusion would fall short of its costs, and merit goods.

Where there is a choice between finance from charges, finance from taxes or a mixture of the two, it should be recalled that the burden of charges will be apportioned between citizens in relation to their consumption levels whereas the burden of support from taxes will be apportioned in relation to tax liabilities. For some items, at least, some people may so strongly prefer the burden to be related to consumption levels that they feel any efficiency arguments for subsidised or free provision are more than fully offset by this preference.

The case for subsidised or free provision is perhaps most controversial with merit goods. The principal argument for replacing such provision of any particular merit good with a system of vouchers or cash transfers is that the change should enable recipients to reach preferred spending patterns. The principal argument against replacement of this sort is that aggregate consumption levels for the merit good concerned might well fall. This chapter suggests that such a fall is not inevitable, and, more importantly, it shows that the consumption levels of those least disposed towards purchasing the merit good might actually rise if the replacement policy were pursued.

Of the common subcentral services, it is perhaps free school education and subsidised housing where the arguments for moving towards market pricing are most keenly debated. A free state school system might be defended on the grounds of promoting social cohesion or equality of opportunity, but the logic of this defence points towards uniform central provision rather than varied subcentral provision.

Subcentral authorities might be able to levy charges on motorists for road use, but they would presumably make a significant contribution towards road costs from their tax revenues on behalf of pedestrians and cyclists. It might be possible for charges on road congestion to be levied as well, but it might be best for them not to be levied by the authorities

providing the roads in case they were deterred from making improvements that reduced congestion and, in turn, their revenue from this source.

Notes

1 See, for example, Webb (1976) for a discussion of the theory. The prescription that prices should equal marginal costs assumes that there is perfect competition in all private industries and that there are no external costs arising from any industry. Webb has shown (pp. 17–21) that, where capacity is of the optimal size, short-run and long-run marginal costs will be identical and prices may be set equal to both. For a discussion of how this principle may be applied to some typical local authority services, see Vickrey (1963).
2 For a fuller discussion of this situation, see Boadway (1979, pp. 154–8).
3 See Boadway, (1979, pp. 156–7) for a discussion of this situation. It is assumed in the text and in Figure 7.2 that the average revenue curve includes any subsidies that may be available to cover any external benefits created by the firm concerned.
4 See, for instance, Boadway (1979, pp. 31–6) and Peston (1972).
5 See, for example, Robinson (1979, p. 119) and Stafford (1978, pp. 69–70).
6 For other general surveys, see Seldon (1977, Chapters 4–9) and Mushkin (1972, Chapters 6–16). See, also, Bird (1976, Chapters 5–14, and, for a bibliography of relevant works, pp. 245–59).
7 See, for example, Peacock and Wiseman (1964) and Maynard (1975).
8 See, for example, Musgrave and Musgrave (1976, p. 51), who present this analysis in the context of an uncrowded bridge.
9 See Bird (1976, p. 60), Brownlee (1961, p. 428) and Wohl (1972, p. 263), respectively.
10 For the United Kingdom in 1964, Roth (1965, p. 57) suggested figures of £163m. for the fuel tax, £32m. for licence duties and £205m. as a contribution from local taxpayers. Such revenues would have closely matched the road costs incurred that year. By contrast, he estimated that congestion charges would have raised a further £624m.
11 See, for example, the conclusions of Musgrave and Musgrave (1976, p. 65) and Seldon (1977, pp. 91–2).
12 See Robinson (1979, pp. 98–115) and Stafford (1978, pp. 58–67) for further discussions of these points.
13 See, for instance, Prest (1982, p. 21) and Harris and Seldon (1976, p. 56).

8

Loans

Introduction

The literature on borrowing by subcentral authorities seems chiefly to relate to three major debates. Each of these debates puts forward reasons why the borrowing activities of these authorities might need to be subjected to some controls. One debate concerns the possible macroeconomic implications of subcental spending financed by loans. That debate is considered in the next chapter. The present chapter concerns the other two debates.

At first sight, these two debates may appear to be unrelated, and certainly they seem to have been discussed quite separately. Ostensibly, the first concerns only the question of equity over time, and has focused in particular on the issue of inter-generational equity by considering whether or not subcentral authorities should be required to finance investment projects with long lives by bonds rather than taxes in an attempt to spread some of the burden of payment on to future generations, which will also benefit from the projects. The second, by contrast, began with a definition of optimal levels for subcentral spending and debt, and has focused on whether there may be systematic forces at work encouraging the authorities to issue super-optimal or excessive debt in the pursuit of super-optimal or excessive spending levels; in general, this debate has implicitly assumed that the benefits from the spending concerned and the burden of repaying any loans incurred will accrue to a single generation.

This chapter examines both these debates, and demonstrates that they are, in fact, closely related. More importantly, this chapter shows that when the issues raised by each debate are considered fully then each debate produces identical policy implications. These are that subcentral authorities should abide by two rules when financing their expenditures: first, current expenditure should be financed by taxes and charges alone, and, secondly, each capital project should be wholly financed by loans that should be paid off over the lifespan of the project concerned in such a way that the proportion of the total interest and redemption costs falling in any one year equals the proportion of the project's total benefits enjoyed during that year. These rules preclude the use of a current surplus as a source of revenue for capital projects, although, of course, a current surplus would be needed subsequently to repay the loans raised to pay for capital projects. Whether any subcentral authorities are, in practice, obliged to abide by these rules is doubtful. It is true that current account surpluses often provide only modest contributions to capital outlays, amounting for example to only 6 per cent of local authority capital expenditure in England and Wales in 1981–2 (HMSO, 1983, pp. 14–15), and it is true

that subcentral authorities are often subject to tight borrowing rules. Nevertheless, small contributions are not equivalent to zero contributions, and, much more importantly, borrowing rules are usually more concerned with preventing the authorities from running into financial difficulties than with the issues discussed below. Such rules seldom require each capital project to be financed with loans whose amortisation periods equal the project's lifespan, and seldom, if ever, require the amount of debt paid off each year to follow the procedure proposed here.

This chapter suggests that the differences between the two debates are essentially ones of approach and emphasis. In the case of the first debate, it is clear that if the present generation can use bond finance escape some or all of the costs of investment projects when they expect to enjoy a proportion of their benefits, then they may well attempt to escape all their costs and opt for excessive spending levels; optimal levels are likely to arise only if the electors concerned are confronted with the full costs of the benefits they themselves stand to gain. Thus the first debate does relate to the question of optimal debt and spending levels, and it implies that excessive levels may arise if subcentral authority finances do not aim to secure inter-generational equity. By contrast, the second debate implies that excessive levels can arise if these finances do not aim to secure intra-generational equity. Here, interest focuses on authorities whose median voters expect to emigrate to other authorities at a future date, and who therefore urge that projects with benefits that will be largely short-lived and enjoyed by them should be financed by loans that need to be largely repaid after they have moved. Such voters hope to escape part of their fair share of the costs of the benefits they expect to enjoy (at the expense of future immigrants who may derive little benefit from the projects but will be required to pay taxes to help repay the loans), and they may demand excess spending levels as a result. This chapter shows that the rules suggested in the previous paragraph should secure inter-generational equity, intra-generational equity and optimal debt levels.

Although the two debates are closely related, they are discussed separately in the following sections of this chapter. This is partly because their approaches are different and need to be explained separately. It is also partly because the possibility of migration between authorities is important to each debate yet affects them differently. Broadly speaking, such migration is likely to reduce the likelihood of inter-generational inequity but increase the likelihood of intra-generational inequity.

Before proceeding with the analysis, three simplifying assumptions – common to most, if not all, the literature on this topic – will be made. First, it will be supposed that no tax exporting occurs. Although the possibility of tax exporting is likely to exacerbate any tendency towards excessive spending levels, it is unlikely to affect the choice made between taxes and loans. This is because an authority that, for example, replaced a tax structure where there was no exporting with one where 25 per cent of the burden was exported would reduce by that amount the sacrifices its citizens must make to repay and service loans as well as the sacrifices needed to finance expenditure directly from taxes. Accordingly, the more attractive of the two revenue sources beforehand will remain the more

attractive one afterwards. Of course, the citizens might well be tempted to demand that taxes replace charges in the new circumstances.

Secondly, it will be supposed that voters in general, and median voters in particular, are wholly lacking in altruism. In formal terms, their welfare functions do not take the welfare of other citizens, except those of their own families or households, into account. If some altruism were present, then, of course, voters would be less inclined to seek benefits at the expense of future generations or future immigrants into their areas. This would weaken the case for the rules suggested in this chapter, but it would not in general destroy that case.

Thirdly, it will be supposed that there is no inflation. The existence of inflation, and the possibility that it might exist during the lifetime of some subcentral authority assets and loans, would not lead to any change in the rules put forward here. It would, however, make those rules a little harder to follow. One possibility would be for subcentral authorities to issue bonds whose value would rise in line with prices and that would attract a fixed real rate of interest. In these circumstances, the number of bonds due for repayment each year could be calculated at the outset on the assumption of zero inflation, and the burden of repayment in each year would be unaffected by any subsequent general changes in prices. Moreover, the real burden of paying the interest on any outstanding debt would also be independent of price changes.

Loan finance and inter-temporal equity

When any government spends money on current account, it is generally the case that the benefits from the expenditure are confined to the time when it is made. When money is spent on capital account, on the other hand, assets are acquired from which benefits will be expected to flow for a number of years. There is likely to be widespread support for the value judgement that inter-temporal equity will be attained if the present generation meets the entire burden of financing current expenditure and if present and relevant future generations share the burden of financing capital expenditure. Consequently, there is considerable interest in designing revenue arrangements where these conditions are met. Most of the analysis of this issue relates to central governments, and their situation will be considered first. Subsequently, the discussion will consider what modifications to the analysis are needed in the context of subcentral authorities. One difference between the case of a central government and that of a subcentral one is that the budget policies of the former are far less likely to lead to migration into or out of its area. Indeed, discussions of the central government case generally assume, if only implicitly, that migration can be ignored.

Suppose, then, that the budget policies of the central government will not lead to migration. Suppose, too, that this government has contrived to fix the level of aggregate demand at the level it deems appropriate for its stabilisation policies. Now, if this government wishes to increase its spending on goods and services, whether on current account or on capital

account, then it will have to take steps to cut other items of spending; in a narrow sense, therefore, the immediate cost of the increased government spending must be met in full by the citizens of the day who will have to release these resources. However, as Musgrave and Musgrave (1976, p. 605) have pointed out, it makes an important difference whether the resource withdrawal is from consumption or investment. If the withdrawal results in reduced consumption, then the full cost falls on the citizens of the day. If it is from investment, however, then these citizens make no such sacrifice. The real cost in this case is a smaller capital stock and hence lower living standards for citizens of the future. Thus inter-generational equity can be secured by ensuring that current spending is financed in ways that cut consumption and that capital spending is financed in ways that cut investment. Ideally, any government capital project should be financed in a way which cuts other investment projects of a precisely similar life span.

The situation is very similar if the central government wishes to let aggregate demand increase by the amount of its desired spending increase. In this case, there will be no cost to either the present or future generations in terms of reduced consumption or investment, since no reductions will be needed; but there will be an opportunity cost to them in terms of the increased consumption or investment that the central government would have allowed had it not wished to raise demand by raising its own spending. Considerations of inter-generational equity suggest that increases in government current spending should be offset by forgone increases in consumption, and that increases in government capital spending should be offset by forgone increases in private investment.

The actual burden of any increase in government spending will be distributed between present and future generations in proportions that depend on how it is financed. Musgrave and Musgrave (1976, p. 605) noted that tax finance tends to fall on consumption (although it has some effect on interest rates and investment) and so it appears broadly suitable for the finance of government current spending. Conversely, domestic borrowing tends to fall on investment (although saving may increase when interest rates rise, thereby having some effect on consumption) and it thus appears broadly suitable for the finance of government capital spending. It is likely that charges tend to fall chiefly on consumption (of other products) and thus have similar effects to taxes and are broadly suitable for the finance of government current spending. A fourth form of finance is external borrowing, which, in the case of a central government, means borrowing from abroad. This does not cause any fall in other spending, but it imposes a burden on future generations who have to repay the loans, and therefore it appears suitable only for the finance of government capital spending. Musgrave (1959, p. 574) has pointed out that if the government wished to maintain a stable level of aggregate demand, then it would probably be necessary to supplement domestic loan finance of capital spending by restrictive liquidity measures to obtain the proper release of resources from private investment, and, of course, it would probably be necessary to raise taxes or charges by more than the increase in current spending since any increase in taxes or charges might be met partly

by reduced saving and not wholly by reduced consumption; external borrowing would probably not be used in this case.

The situation for a subcentral authority differs in three principal respects from that for a central government. First, the authority will have recourse to only three forms of finance: taxes, charges or external borrowing. Internal borrowing is not really an option since most of such an area's debt issues are normally held by residents of other areas. Secondly, the authority will not be responsible for managing aggregate demand. It will raise taxes or charges, or issue debt, only to an amount sufficient to cover any increase in its spending, so that further action will be needed by the central government to neutralise aggregate demand fully against the increased subcentral spending. As far as inter-temporal equity is concerned, Musgrave and Musgrave (1976, pp. 606–7) have argued that, as with the central government case, the burden of taxes will fall principally on the present generation (and the burden of charges should be similar) while the burden of external borrowing will be spread among future generations. Thus, it seems that taxes (or charges) should be used to finance current spending and loans used for capital spending; ideally, loan repayment should be spread out over the life of the asset concerned.

One exception to these rules may be mentioned. It is arguably acceptable for loans to be used to finance current spending if the interest and repayment costs of the loans are met from a property tax on domestic property alone. In this situation, the expected extra future property tax liability in an area would depress its citizens' demand for housing and so diminish property values. As a result, the present generation would face the burden of the current spending in the form of a capital loss. Property values would subsequently rise steadily as the loan was paid off, so that future generations would receive capital gains to compensate them for their extra taxes. One problem with this exceptional case arises with rented property, because the benefits of the current spending financed by loans may accrue to tenants whereas the cost, in the form of a capital loss, may accrue to landlords. In principle, landlords could seek to compensate themselves by raising rents for the year in which the current spending was made, but in practice the decision to undertake this spending, and the spending itself, might occur while existing tenancy contracts were in force, thereby precluding any rent rise.

The effects of migration on inter-temporal equity

Daly (1969, pp. 46–8) has argued that this type of analysis ignores a third difference in the position of subcentral authorities – namely, that citizens can readily immigrate and emigrate. Suppose that all citizens are owner-occupiers of property and can move costlessly between areas in response to fiscal differentials. In this situation, Daly has claimed that, in any area, both a rise in taxes (or charges) and the issue of debt will have the same effect, and that both will fall wholly on the initial citizens. This is to be expected with taxes (or charges) for the reasons explained earlier. It applies with loans because residents in the area are saddled with the tax liability necessary to serve and repay the loan. Whatever (unexportable) tax is in

use, this situation will encourage emigration and so lead to property values in the area falling by the discounted value of the extra future tax liability to be encountered there. This means that, while the burden from taxes (or charges) falls on the initial citizens in the form of reduced scope for consumption and saving, the burden from loans falls on them in the form of reduced property values. Thus, either method of finance can be used for current spending, whose benefits are confined to the initial citizens, without producing inter-temporal inequity.

More surprisingly, perhaps, Daly claimed that either method can also be satisfactorily used for capital spending. While initial citizens will feel the full burden of the cost, whatever source of finance is used, they will also get a benefit in the form of an increase in property values equal to the discounted value of the expected future benefits from the project concerned. Consider a marginal project, whose future benefits to the members of each household precisely equal its cost to them. Suppose, first, that the project is financed by loans whose repayments are spread out over the life of the project. In this case, the expected benefits from the project will tend to raise the value of any household's property by an amount precisely equal to that by which the future extra tax liability incurred as a result of loan finance tends to reduce it. Thus property values do not change, and each generation benefiting from the project shares in the cost through extra taxes. Suppose, instead, that the project is financed by taxes or charges. In this case, the expected future benefits from the project will succeed in raising the value of each household's property, and they will do so by an amount precisely equal to the taxes or charges the household has paid for the project. The value of the property will fall as time passes, and thus the financial burden of the project is shared by each generation, since each generation makes a capital loss on its property.

Oates (1972, p. 157) has argued that a critical assumption of Daly's hypothesis is that there is sufficient mobility of citizens in response to fiscal differentials between areas for these differentials to become fully capitalised. If mobility is very costly, then property values will not rise in response to the benefits from capital projects; thus, if these projects were financed by taxes or charges, then the whole burden would fall on the initial citizens. By analogy, the future tax liability resulting from loan finance might not generate emigration, and so might not lead to a fall in property values; thus, if current spending were financed by loans, then there would generally be no burden on citizens at the time. There is one exception to this latter proposition, and this occurs if domestic property taxes are used alone. In that case, as noted in the last section, the extra future property tax liability resulting from the use of loans to finance current spending would depress citizens' demands for housing, and so depress property values, even if no emigration occurred.

In practice, of course, migration is costly. It follows that there is an inter-temporal equity case for using taxes or charges to finance current spending (unless the loans are repaid and serviced by domestic property taxes) and for using external borrowing to finance capital spending. More specifically, the loan needed for any project should be paid off during the lifetime of the project so that the burden of servicing and repaying

the loan is shared amongst all those benefiting from the project in a manner that places no reliance on capitalisation adjustments in property values. In other words, the loan should be repaid in a way which means that the proportion of the total interest and repayment charges due in any one year equals the proportion of the total benefits enjoyed during that year.

Another assumption of Daly's hypothesis is that all citizens are owner-occupiers. The existence of rented property does not affect the burden imposed on each generation, because the value of such property will also fall if current expenditure is financed by loans and rise if capital projects are financed by taxes or charges. Problems can still arise, however. Suppose, for example, that subcentral taxes and service benefits fall chiefly on tenants. In that case, if current spending is financed by loans, then tenants would expect to escape their share of the present generation's burden that takes the form of a fall in property values, but, if capital projects are financed by taxes or charges, then tenants would expect to receive no share of the rise in property values that compensates the original generation for its tax payments. In principle, equity between landlords and tenants could be secured by suitable temporary changes in rents. A fall in property values resulting from using loans to finance current spending could be passed on to tenants if rents rose in the year of the spending, while a rise resulting from the use of taxes to finance capital spending could be passed on to the taxpaying tenants by a fall in rents in the year when the necessary tax payments were collected. In practice, tax and spending decisions may well be made and implemented during the period of existing tenancy agreements so that the necessary rent adjustments cannot and do not take place. Accordingly, it may be felt that equity between landlords and tenants would be served best if there was no loan finance for current spending and no tax or charge finance for capital spending. Again, the loan needed for any project should be paid off during the lifetime of the project in such a way that no capitalisation adjustments need occur to property values and, in turn, no intra-generational equity problems between landlords and tenants arise.

A further limitation of Daly's argument is that it considers a single area. Property values in one area may well rise if it alone finances capital spending from taxes and becomes relatively attractive, or fall if it alone finances current spending from loans and becomes relatively unattractive. If all areas financed capital spending from taxes or current spending from loans, however, then maybe none would become relatively attractive or unattractive. Accordingly, the forces causing capitalisation could disappear and so inter-generational equity problems could reappear. In such cases, the only circumstances in which Daly's arguments would remain unscathed would be where loans were used to finance current spending and where a tax on domestic property alone was used in future years to serve and repay the debt. Such a tax would initially depress property values, which would rise steadily as the loans were paid off so that future citizens would be compensated for their tax payments. If inter-generational equity is to be preserved, then it seems that loans should not be used for current spending under other circumstances, and that taxes should never be used to finance capital spending.

It should be noted that, even if Daly's argument was always valid, the choice between taxes or charges and loans might still not be entirely a matter of indifference. In particular, there are two practical reasons why loan finance for capital projects might still be preferred to tax (or charge) finance. First, for cash-flow reasons, the initial citizens might not be very happy at paying a large tax bill with the thought of compensation when, maybe years later, they sell their houses; loan finance avoids any large lump-sum bill.[1] Indeed, citizens might simply vote against expensive projects that it was proposed to finance from taxes. That citizens do not regard tax and loan finance with indifference is suggested in a study by Pogue (1970, p. 47) which found that restrictions on local authority borrowing in the United States in 1958 and 1962 resulted primarily in less spending rather than a switch to tax finance.[2] Secondly, tax finance for capital projects could lead to substantial changes in tax rates over time and between areas; these effects could cause uncertainty and disruption, and might (depending on the taxes used) lead to large temporary flows of trade or labour from one area to another.

In practice, of course, Daly's assumptions are unlikely to be wholly valid. Consequently, inter-generational equity will obtain only if current expenditure is financed from taxes or charges and if capital projects are financed by loans that are paid off in the manner suggested. It should be added that such requirements are needed not only to secure equity, but also to discourage excessive spending and debt levels. For, if much of the burden of repaying loans falls on future generations, then the present generation might well be tempted to finance current spending and short-term capital projects out of long-term loans, and would be tempted to raise such spending up to the point where the marginal benefits equalled the marginal costs to them, even though the marginal benefits might be well below the real marginal costs.

Loan finance and intra-generational equity

In the second debate concerning loans as a source of subcentral income, the possibility that subcentral authorities may issue more than an optimal or efficient amount of debt is focused on at the outset. The optimal level of borrowing is held to occur when the marginal gains from loans to the citizens in the borrowing authority equal the marginal costs to them. It might be argued that, if authorities are free to choose, then they will borrow up to this level, and thus that there is no case for having controls over the amount of debt they can issue. However, it is sometimes argued that such controls are needed, and this chapter considers the argument in the light of a paper by Wagner (1970) and the discussion stimulated by it.[3]

Wagner's analysis begins by considering the similar cases of an individual citizen and an individual subcentral authority.[4] Initially, it is assumed that no migration occurs between authorities. In each case, spending in one period can be increased by borrowing, although interest payments and loan repayments will reduce spending in future periods below the levels that would otherwise have been possible. Typically, loans will be considered

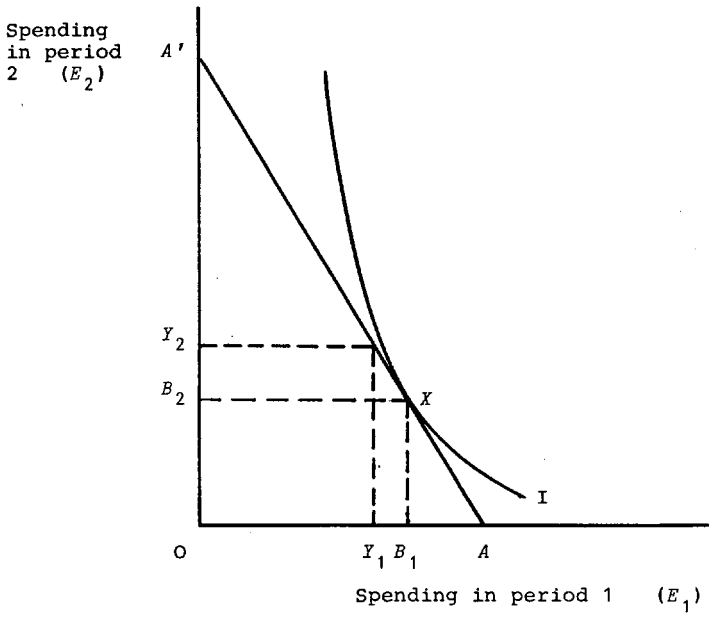

Figure 8.1

for capital expenditure on consumer durables or public investment, principally because loans enable payment to be smoothed out over a number of time periods. It is up to the borrower to decide how much to spend in each time period and, in turn, how much to borrow in any one period. The situation can be readily illustrated for an example with two time periods, as in Figure 8.1.

Consider, first, the case of an individual citizen. In that case, the line AA' in Figure 8.1 represents the possible ways in which the individual citizen can spread his spending out over the two periods concerned. At A, all income is spent in period 1. Spending thus equals income in period 1 (say Y_1) plus the maximum amount that can be borrowed and repaid from income in period 2 (of, say, Y_2), that is $Y_2/(1+i)$ where i is the rate of interest; hence $OA = OY_1 + OY_2/(1+i)$. At A', all income is spent in period 2. Spending thus equals income in period 1 and the accrued interest obtained by lending it, that is $Y_1(1+i)$, plus income in period 2 (Y_2); hence $OA' = OY_1(1+i) + OY_2$. It follows that $OA' = OA(1+i)$, and that the slope of AA' is $-(1+i)$.

Consider, next, the case of an individual subcentral authority. For simplicity, assume initially that this authority provides a single service (cash handouts) financed by a single tax, and suppose that the handouts and taxes for any citizen are proportional to his income. In this situation, the line AA' in Figure 8.1 can be taken to reflect the spending possibilities available to the median voter if it is assumed further that he undertakes no private borrowing and that his incomes in the two periods equal OY_1 and OY_2 respectively. He could, for example, be placed at A if his

authority levied no taxes in period 1 but gave him a handout of $Y_2/(1+i)$, which it could afford to do if it borrowed this sum at an interest rate i and repaid it with interest in the next period by taxing his entire income of OY_2 in that period and giving no handouts then. The authority would not be able to pay the median voter more than $(OA - OY_1)$ in period 1 because it could not borrow any more on the strength of expected incomes in period 2. Alternatively, he could be placed at A' if his authority taxed his entire first period income of OY_1 and had no handouts in that period, but by lending this sum at the interest rate i it was able to give him a handout of $Y_1(1+i)$ in the second period and levied no taxes in that period. It follows that the spending combinations available to the median voter, as depicted by the line AA', reflect the varying borrowing and lending strategies which that authority could undertake in period 1 along with the consequences it would face in period 2. As before, the slope of AA' will be $-(1+i)$. Note that relaxing the simplifying assumptions made in this paragraph complicates the analysis but does not alter its general implications. In particular, these would not be seriously affected if the process outlined here were supplemented by one in which, when these income taxes and handouts were completed, further subcentral taxes were levied to provide other subcentral activities.

The choice between spending in periods 1 and 2 will depend for an individual on his preferences; for an authority, it may be assumed to depend on the median voter's preferences. The relevant preferences are embodied in an indifference map containing the curve I. If spending in period 1(E_1) and spending in period 2(E_2) were rated equally, then the curves would be symmetrical about a line $E_1 = E_2$; they would have a slope of -1 where they intersected that line, indicating that when $E_1 = E_2$ the citizen would be indifferent between an extra pound of spending in period 1 and an extra pound of spending in period 2 – or, in other words, would have a zero rate of time preference. If there is a preference for spending in period 1, then the curves will be generally steeper (with slopes < -1) where they intersect the line $E_1 = E_2$. This latter situation occurs in the figure, and the individual concerned selects position X with expenditure of OB_1 in period 1 and OB_2 in period 2. At X, the slope of AA' (that is, $-(1+i)$), equals the slope of I, which in turn equals $-(1+r)$, where r is the rate of time preference at that point, so that $i = r$. If income in period 1 were OY_1, then it can be seen that loans of $(OB_1 - OY_1)$ would be borrowed in that period. In the case of an individual, there is no sense in which he is getting any unfair gains at anyone else's expense; in the case of an authority, there is no sense in which the median voter is getting any unfair gains at anyone else's expense, and so no case for suggesting that there is any intra-generational inequity between him and anyone else. Thus there seems no case for limiting the debt of either the individual or the authority.

The effects of migration on intra-generational equity

For an authority where migration occurs, however, the analysis needs some

extension. The situation in Figure 8.1 will continue to hold if the median voter estimates the probability of remaining a resident in period 2 (denoted by λ) as 1. If the probability is estimated as less than 1, then the voter's preference about public spending by the authority might be expected to change in favour of period 1. This change in preferences occurs because the possibility of migration gives rise to a possibility that he will escape any taxes needed in period 2 to repay money borrowed by the authority to raise its spending in period 1. In other words, the possibility of migration creates a possibility that some spending in period 1 can be enjoyed without cost to the voter, and this changes his views on how much authority spending there should be in that period.

It should be stressed that the value of λ does not affect the spending combinations shown by AA' in the figure, and between which the median voter must choose. This line shows the spending combinations available to him, and anyone else on a similar income level, if they should be a resident in each period, and these combinations reflect only the varying possible borrowing or lending strategies that the authority might follow in the first period. The value of λ does, however, affect which combination along AA' the median voter will select; that is, it affects the strategy he will wish the authority to follow. For instance, if λ were 0, then the voter would have a vertical indifference curve such as I_1 in Figure 8.2, where AA', X and Y_1 are the same as in Figure 8.1.[5] In this case, he would select point A, where the maximum possible spending occurred in period 1 requiring loans to finance handouts to everyone, including one of $(OA - OY_1)$ to him; the voter's share of the repayment and interest on these loans would be escaped by emigrating before the

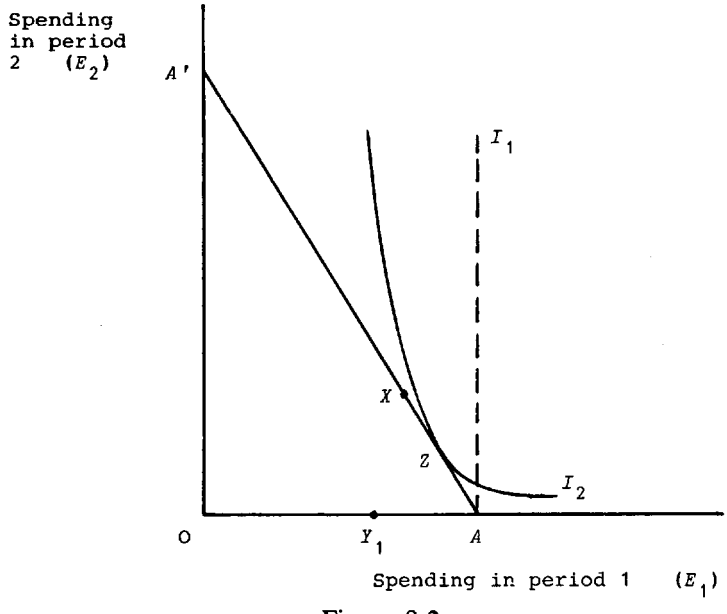

Figure 8.2

taxes needed to repay them were levied. With λ between 0 and 1, say ½, the relevant indifference map might include the curve I_2 in Figure 8.2 and lead to the point Z being selected. However, provided λ is less than 1, the median voter has an incentive to desire greater spending in period 1 than would be found at X because there is a chance that the benefits of some spending can thereby be enjoyed for nothing. In effect, the median voter votes for excess spending and excess borrowing in period 1 since he makes his choice on the basis of an expected private cost that is less than the social cost.[6] In this analysis, the median voter votes for excess borrowing to finance large handouts in the first period. Some, at least, of these handouts might then be used to pay higher subcentral taxes to enjoy higher levels of subcentral services than would otherwise be provided. If, in practice, subcentral authorities do not operate cash handouts, then similar results could be secured by having them borrow large sums to finance higher service levels at lower tax rates than would otherwise apply. In effect, median voters are seen to be demanding that a large part of public spending be financed through public borrowing in the hope that they can enjoy the benefits of that spending and yet escape, through migration, a portion of their due share of its cost. Wagner (1970, p. 300) noted, therefore, that at first sight 'local debt choices which are based on positive rates of migration would seem to generate an excessive amount of local debt', but he argued that forces are at work to prevent this excess.

The point is that the median voter can escape part of his due share of the costs of public spending in his area only if others have to pay more than their due share of the costs. Those most likely to come into the second category are the immigrants who arrive at the start of period 2, for they will have missed out on the benefits from the spending in period 1 but will have to pay taxes to help repay the loans incurred in that period; this suggests that there is a possible situation of intra-generational inequity between those who emigrate and those who immigrate. However, if the median voter is an owner-occupier who expects to sell his house to an immigrant when he departs, then he must expect to find any undue burden falling on the incomers reflected in, and offset by, a downwards capitalisation of his property. Thus it will not be possible to shift any burden on to incomers. If the median voter has perfect knowledge, then he will be aware of this and realise that he cannot gain from voting for the issue of excess debt, and therefore he has no incentive to do so. If he has imperfect knowledge, then he may vote for excess debt. However, such excess debt issue is likely to be temporary, for, as Wagner (1971, p. 109) pointed out, once it is issued, the median voter will see property values falling in his area and learn that he cannot escape the consequences of his debt choice, and so he will stop voting for excess debt. Thus the case for debt controls appear to be negligible.

Wagner nevertheless believed that this analysis understates the case for controls. He argued (1970, pp. 301–3) that it is more appropriate to consider subcentral authorities as a group, rather than an individual one on its own. The principal difference that arises when considering authorities as a group is that if all authorities issue excess debt then downwards capitalisation of property values may not occur, or, if it does, it may

not be complete. The reason is as follows. When a single authority issues excess debt, it encumbers future residents with a relatively poor future tax-to-benefit position; accordingly, it is a relatively unattractive area in which to live and domestic property values will fall. When all authorities issue excess debt, however, none becomes relatively unattractive, and accordingly it is not at all clear that domestic property values will fall anywhere. Indeed, it is quite possible that domestic property values would not fall at all. Such falls would be caused only by the taxes needed to pay off the excess debt. Suppose, for example, that these taxes did not affect the total level of saving, or, in turn, the total demand for housing and other assets, and suppose that they did not in any way discriminate against housing vis-à-vis other assets. In those circumstances, there is no reason to expect house values to decline. Broadly speaking, the degree of capitalisation will be greater, the more saving falls in response to the taxes and the more the taxes discriminate against housing. Complete capitalisation would be most unlikely unless the tax(es) fell on housing alone. Few, if any, subcentral authorities levy such a tax, so presumably capitalisation in practice would be partial at best. However, as Aronson (1971, p. 108) has noted, anything less than full capitalisation means that issues of excess debt may be made because some of the cost of subcentral services in any area could be shifted on to future immigrants.

It is nevertheless still not immediately clear that excess debt will in fact be issued. Suppose, initially, that each area has an efficient level of debt. If the median voters each have perfect knowledge, then none will vote for excess debt because each knows that if his area alone issues excess debt then capitalisation will prevent any gains being secured as a result. Wagner (1971, p. 109) discarded the assumption of perfect knowledge, however, and without it excess debt may arise. This is because many median voters may, initially, vote for excess debt and will find that the gains are only partially offset by capitalisation, since, as explained in the last paragraph, complete capitalisation is unlikely to occur if excess debt occurs widely. Thus median voters are likely to conclude that their areas have gained as a result, and so they will not vote against excess debt in the future. If excess debt were issued widely, then citizens who moved would derive gains by avoiding some of their share of the cost of the services enjoyed in their original areas, and they would derive losses by shouldering an excess share of the cost of the services provided in their new areas; for many citizens, such losses and gains might roughly balance. Awareness of this situation would not, however, deter median voters from voting for excess debt issues, for it is these excess issues that produce the gains they would enjoy if they moved. Wagner (1970, p. 303) concluded that there is a case for statutory debt limits.

Commenting on Wagner, Aronson (1971) has suggested two further reasons why subcentral debt may be excessive; unlike Wagner's they do not rely on market failure through imperfect knowledge. First, if the median voter who is a potential emigrant is not a property owner, then he may vote for excess debt in the hope of avoiding some of the taxes needed to repay it without fear of any capital loss on property. It would be quite rational for him to vote for excess debt even if other areas had

efficient levels. Secondly, there is the factor of life expectancy. A median voter would have a natural desire to vote for debt finance to be used for projects in circumstances where he expected to live long enough to enjoy most or all of his share of the benefits, but not long enough to pay his full share of the costs; arguably, this point suggests that excess debt may arise if median voters are able to exploit the possibility of inter-generational inequity rather than intra-generational inequity. Wagner (1971, pp. 109–10) has added a third possibility, namely that median voters may ignore the effects of capitalisation because they may choose to consider only those events that they expect to occur in, say, the next two or three years, and they may not expect to move so soon. In that case, Wagner suggested that the effects of excess debt on their house values will be ignored since the effects will not be felt within the relevant time horizon; on the other hand, though, these median voters might be expected to regard λ as 0 and select optimal debt levels after all.

It seems possible, nevertheless, that subcentral authorities will be tempted to issue excess debt and, as a result, spend at levels higher than the efficient ones. Wagner (1970, p. 304) implied that mitigation of this problem justifies broad intervention policies to regulate the total amounts that subcentral authorities can borrow and the maximum amortisation periods of their debt issues. The problem would not arise at all, however, if subcentral authorities were required to finance their activities in accordance with the specific rules suggested to secure inter-generational equity – that is, to use taxes and charges for current spending and to use loans for capital spending, and to repay the loans raised for any project in such a way that the proportion of total interest and repayment charges due in any one year equalled the proportion of the project's total benefits enjoyed during that year. Provided they followed these rules, the authorities could borrow as much as they pleased and have amortisation periods for the loans raised for their most durable projects that were as long as those projects' lives; nevertheless, projects with modest lifespans would not be financed by loans whose repayments were all due at some distant date, so median voters who had a positive probability of moving would not have any expectation of being able to enjoy some of those projects' benefits while escaping any liability for loan repayment. Moreover, it would not be possible to finance current expenditure with loans due for repayment at a time when median voters might have some chance of being no longer resident. Accordingly, median voters would have no incentive to vote for excess debt.

Summary

This chapter focuses on three consequences of possible concern that might arise if subcentral authorities' borrowing was not controlled. First, the authorities might be able to use loan finance to throw some of the burden of financing services benefiting present citizens on to future generations, thereby causing inter-generational inequity. Secondly, intra-generational inequity could occur if loans could be used to throw some of the costs

of benefits enjoyed in one year on to citizens a few years hence; some of those who enjoyed the benefits might leave the area before their costs had been fully met, while citizens who moved into the area in years to come might have to meet some of the cost without enjoying the benefits at all. Thirdly, as a result of escaping part of the costs of the services they enjoy, present citizens might vote for subcentral spending to be at excessive levels.

It can be shown that the forces giving rise to these consequences might be tempered if subcentral authorities levied a tax on domestic property alone and if there was some altruism amongst voters. The situation will also be affected by factors such as the ease of inter-area migration and the presence of rented housing. Nevertheless, it appears that a sufficient, and probably necessary, condition for avoiding these consequences is to require subcentral authorities to use tax and charge receipts alone for current spending and loans alone for capital spending; the loan for any capital project should be paid off over its lifespan in such a way that the proportion of the total interest and repayment costs due in any one year equals the proportion of the project's total benefits enjoyed in that year. These rules are very different from, and probably less restrictive than many of those currently in existence.

Notes

1 In principle, citizens might be able to borrow privately to pay the tax bill and thereby attempt to put themselves in the same position they would enjoy had the authority used loan finance. However, as Oates (1972, p. 158) has explained, government units can typically borrow at interest rates significantly below those available to private citizens, so that there are real gains to be obtained by undertaking borrowing collectively through subcentral bond issues.

2 It should be noted, though, that the analyses of the present chapter suggest that the avoidance of large lump-sum bills may not be the only reason why loan finance is sometimes preferred; present citizens may hope that using loans means they will escape some of the cost of subcentral authority spending.

3 For a quite independent attempt to give a rationale for debt controls, see Breton (1977).

4 The analysis in the text follows the reasoning in Wagner (1970) but is presented in terms of the conventional model for optimisation over time and ignores the modified approach used by Wagner and suggested by Thurow (1969). Thurow separated the rate of time preference from the individual's utility of income function, but, as noted by Hand and Mitchell (1971, p. 102), this is not necessary.

5 Wagner (1970, p. 302) and Hand and Mitchell (1971, p. 104) presented diagrams different from Figure 8.2 and different from each other. Each of their figures shows the possibility of migration leading to the median voter being presented with a new budget line. This suggests that their budget lines do not reflect the options open in the area of residence in period 1, for these are not affected by the probability of migration by the median voter. Nor do their budget lines show the E_2 expected to be enjoyed by the median voter for a given E_1; this would depend on the level of period 2 spending expected in the area or areas to which he might migrate. Thus it is not wholly clear what their budget lines do show. It is worth adding that, as the level of E_1 preferred by any citizen is likely to depend on the value of λ for that citizen, so the identity of the median voter, and hence the median preference, could alter when any one person's λ alters.

6 Note that the analysis in the text regards debt as excessive when it leads to overspending. It is not concerned with the optimal amount of loan finance for the optimal level of spending. Collins (1977, p. 123) suggested that the optimal debt level should be regarded as the one that minimised the cost of the optimal level of capital expenditure, although he noted that there would be no inefficiency effects on society from non-optimal debt in this sense; he did not allow for inter-temporal equity considerations, which, as argued in the present chapter, suggest that loan finance should equal capital spending. For a further discussion of this approach, see Sullivan (1979) and Collins (1979).

9

The Impact of Subcentral Government on Central Government Economic Policy

Introduction

In this chapter it is assumed that a structure of one or more tiers of subcentral authorities has been established, that their functions have been determined, and that they are financed by satisfactory taxes, charges, grants and loans. It is assumed that the functions are primarily ones of resource allocation, although the authorities may play some explicit part in redistribution – perhaps through the provision of some transfer payments and some merit goods – and a minor part in stabilisation – perhaps through the use of regional investment incentives. The question now arises whether the activities of this structure will in any way hinder the ability of the central government to succeed in its own economic policies. The central government, of course, has its own roles in resource allocation, redistribution and stabilisation, and it has been suggested (King, 1976) that the activities of subcentral authorities may impinge on each. The three policy areas will be considered in turn, although the third area is examined at greatest length.

Central government resource allocation policy

As far as allocation is concerned, the central government may generally be willing to permit the authorities to use resources in ways that secure the patterns of subcentral services desired by their electorates. Nevertheless, if it regards itself as having some overall supervisory role, then it may be concerned to see: (1) that the sizes of authorities established take account of the possibility of economies of scale; (2) that authorities take account of the external effects of their services; (3) that significant migration of households (other than to live in areas with preferred levels of taxes and benefits), labour, trade and capital is not caused by areas of high resources or low needs being able to set attractive tax rates or service levels, or caused by different tax structures or rates; and (4) that authorities use adequate expenditure control and project appraisal techniques. It is assumed here that concern over economies of scale is met by ensuring that this factor is considered when deciding on the sizes of authorities, that concern over externalities is met by considering them when sizes are fixed and resolving outstanding externalities with the use of appropriate grants, that concern over migration is met partly by using an equalisation scheme to neutralise the fiscal advantages of high-resource and low-need areas, and partly by

having a harmonised structure of taxes that individually minimise the migration risk, and that adequate expenditure control and project appraisal techniques are made obligatory.

A further resource allocation concern may be that of trying to ensure that the level of subcentral spending in each area is at or near the optimal level. It is sometimes argued that authorities in receipt of effort-related revenue-sharing or equalisation grants will tend to have excessive spending levels. This argument usually refers to grant schemes where grant receipts are proportional to tax effort (as with equalisation of fiscal potential grants) so that the apparent marginal costs of subcentral services to electors will be consistently below their true marginal costs. The argument is perhaps debatable in the case of revenue-sharing grants, because these are presumably specifically intended to facilitate subcentral spending. It is stronger in the case of a grant scheme solely concerned with equalisation, though in cases of equalisation of fiscal potential there will also be areas entitled to negative grants that find part of any increase in taxes having to be paid to the grantor and that thus have an incentive to underspend. Such problems of non-optimal service levels in individual areas might well be smaller if the new equalisation scheme outlined in Chapter 5 were adopted; with this scheme, an area entitled to grants would find the proportion of its income from grants declining as its tax effort rose progressively above a standard level, while an area obliged to make payments to the grantor would find the proportion of its tax revenue payable rising as its tax effort fell progressively below a standard level. It must be conceded that the only equalisation scheme that avoids all incentives for inefficient budget levels is the one that equalises fiscal capacity, since it uses lump-sum grants; but, as shown in Chapter 5, a switch to that scheme from one using effort-related grants would create a situation where horizontal equity problems arose.

Non-optimal spending may, of course, result instead from using subcentral taxes that are easily exported or hard to perceive or that may be used in tax competitions, in which cases the appropriate action is to find more suitable taxes. It would also accompany a tendency for subcentral authorities to issue excess debt for the reasons suggested in Chapter 8, in which case some rules over subcentral borrowing would be justified.

Central government redistribution policy

It will be assumed that the central government has prime responsibility for distribution policies and that it frames the details of these policies in the context of the behaviour of a typical or average authority. One possible worry is that the central government may have its intentions thwarted in subcentral authority areas where tax and expenditure mixes differ widely from the average. It has been suggested that it is poor households whose post-budget incomes – that is, incomes after deducting payments of all taxes and adding in any receipts of cash transfer payments and some value for items such as subsidised health care and zero-priced education services – will be most sensitive to subcentral authority fiscal behaviour. This sensitivity will arise more from changes in the levels of benefits, whose

value for the poor will be high relative to income, than from changes in the levels of tax rates, as tax payments should be low.[1] This worry could be used as the basis of a case in favour of some central control over subcentral fiscal behaviour. Of course, there is also a case against controls in that they would prevent geographical variations in tastes being fully catered for.

In practice, variations in subcentral budgets might be small. If so, the case for the controls outlined in the last paragraph would also be small, but there could still be grounds for central concern. It was explained in Chapter 1 that it would be hard for any one area to pursue policies with a more progressive impact than the others, and a narrow range of subcentral budgets could reflect a situation where each authority has had to conform closely to the policies set in the area whose budget has the most regressive (or the least progressive) impact. If this were so, then there would be a case for controls to make all subcentral budgets conform more closely to the median preferences in the country as a whole.

Central government stabilisation policy

In the case of stabilisation, a central government may have two very different fears. First, it may be worried that the behaviour of subcentral authorities will tend to result in a geographical distribution of output that differs from the one it prefers. This could occur in three main ways. For instance, some authorities might use any general powers they had to encourage industry to locate in their areas far more effectively than others; this possibility suggests that some control may be needed over the types of instruments used. Alternatively, some authorities might spend far more on such instruments than others; this suggests that some control over levels of spending may be needed. Finally, some authorities may encourage their service managers to purchase inputs from local suppliers irrespective, or at least largely irrespective, of whether inputs could be purchased more cheaply from suppliers in other areas. King (1981, pp. 11–14) has suggested that (for hypothetical regional authorities in the United Kingdom) each £1 of expenditure switched from another region could lead to a final increase in output of £1.20–£1.30 in the region concerned. This possibility suggests that some regulations may be needed to control or prohibit discretionary purchasing by authorities.

The second worry for the central government with stabilisation is that the actions of subcentral authorities might affect its ability to achieve its major macroeconomic goals.[2] For instance, it might fear that they could prevent it securing its desired levels of aggregate demand, output and employment. It might also fear that they could affect the levels of prices or disposable incomes, and thereby stimulate inflationary pressures. Further, it might fear that they could affect the rate of economic growth, perhaps by borrowing heavily and forcing up interest rates. In addition, it might fear that the balance of payments would be disturbed by any effects subcentral authorities were able to have on output levels, and so imports, or on interest rates, and so capital flows. It is this macroeconomic impact that occupies the rest of this chapter.

In the next section, interest focuses on what is usually regarded as the major issue in this topic, namely the effect subcentral authorities can have on the central government's demand management policies, particularly insofar as these affect the levels of output and unemployment. The following section considers the effect the authorities can have on other macroeconomic targets, particularly the rate of inflation, the balance of payments and the rate of economic growth. The final two sections examine the ways in which subcentral borrowing and budgets can be influenced to reduce any adverse macroeconomic effects.

The impact of subcentral authorities on demand, output and unemployment

A central government's concern with the impact that subcentral authorities may have on its demand management policies depends in some measure on whereabouts its beliefs concerning the nature of the economy lie on the spectrum of possible beliefs that has the extreme monetarist view at one end and the extreme Keynesian view at the other.[3] It seems prudent to consider separately the concern that would be felt by governments near each end of this spectrum.

Before considering these positions, however, five general points should be discussed. First, assuming the central government is the grantor, there is the question of whether its aggregate grant payments can be precisely predetermined. If they cannot, then the central government will lack control over one component of its own expenditure, and an inability to determine its own spending clearly affects its ability to manage aggregate demand. If fact, the possibility that aggregate grant payments cannot be precisely predetermined arises chiefly with matching specific grants and effort-related general grants. In some countries, grant payments in the former category are ordinarily so small that differences between the total grant forecast and the total grant realised can be ignored; in others, the problem is eased by having a number of grant programmes based on closed matching specific grants, so that a widespread increase in subcentral spending may not lead to excessive grant increases. More concern arises with total grant payments of effort-related general grants, which are generally used for equalisation or revenue-sharing purposes. It was explained in Chapter 5 that precise predetermination could be achieved, perhaps by relating payments to tax effort in a previous year, or by altering the grant formula parameters after all the grantee tax rates have been set. If a central government chooses to adopt a system where precise predetermination is not possible, then it must accept that some difference may arise between its expected and actual grant expenditures. However, as explained in Chapter 5, it can always change the parameters of the grant formula from year to year, so it may generally be able to keep the gap fairly small. It is important to realise that such gaps can exist, but it is reasonable to assume that when grants that produce them are operated then the gaps are generally felt to be unimportant. Accordingly, these gaps are ignored in the following discussion.

A further uncertainty over total grant payments could arise if grants were fixed in advance in nominal terms, because then the final real payment would depend on the subsequent rate of inflation and so could not be accurately predicted. Central governments concerned by this possibility could avoid the uncertainty by fixing grants in advance in real terms, and allowing the sums paid to take account of subsequent changes in the price level.

Secondly, it should be noted that the central government is likely to be concerned about the broad division of subcentral spending between purchases of new goods and services (whether by current or capital expenditure), subsidies, transfer payments and purchases of existing assets such as land, existing buildings or securities.[4] An increase in purchases of new goods and services necessarily increases spending on final goods and services by the same amount.[5] An increase in spending on subsidies reduces the prices of the items subsidised and hence raises the real incomes of consumers; the increase in aggregate consumer spending at factor cost will be less than the value of the subsidies to the extent that this rise in incomes causes a rise in saving. An increase in transfer payments will lead to a smaller increase in spending by consumers if the marginal propensity to save of the recipients exceeds zero. An increase in purchases of existing assets will lead to an increase in spending on goods and services, by the sellers, to the extent that their marginal propensity to consume out of the revenue from their sales exceeds zero. In general, it is probably the case that increases in subsidies and transfer payments have almost as much effect as increases in purchases of new goods and services, but that increases in purchases of existing assets have much less effect.

Thirdly, it should be noted that the central government will be concerned about how subcentral spending is financed. A given increase in spending (assuming grants and any tax-sharing receipts are fixed) can be financed by raising taxes, increasing charges, borrowing or reducing balances. An increase in taxes will generally lead to a relatively large fall in private spending on goods and services, although some of the increase may be met by a fall in saving;[6] indeed, the amount of the fall in saving will depend to some extent on the tax used, as explained in Chapter 6, being greatest perhaps with a corporation tax, much of whose incidence may fall on owners of capital who are widely believed to have relatively high marginal propensities to save. An increase in charges also reduces real incomes; the resulting fall in aggregate spending on goods and services (at factor cost) will be less than the amount of the increase in charges only to the extent that the fall in incomes reduces saving. By contrast, increases in borrowing or finance from balances create no direct offsetting falls in private spending.

Fourthly, there is the possibility that each citizen can deduct his payments of some subcentral taxes from his taxable income for central income tax purposes. Where this system operates, it may be harder for the central government to predict its own tax yield in any given year, and this may affect its ability to manage aggregate demand. Perhaps a more serious implication of these deductions or offsets is that they mean rises in subcentral taxes do not reduce disposable incomes by the full amount

of the rises. Accordingly, central governments may well be more concerned when subcentral authorities finance spending increases with taxes that can be offset than when they finance them with taxes that cannot be offset.

Fifthly, it should be appreciated that the central government may desire powers to secure rapid and sizeable changes in the aggregate size and broad composition of subcentral budgets, and yet also desire powers to limit the rate at which that size and broad composition can be varied autonomously by the subcentral authorities themselves. The central government may desire the first sort of power in order to facilitate changes in the level of aggregate demand, either to effect a change in its macroeconomic policies or in response to changes in the values of exogenous variables. As far as changes in subcentral budgets are concerned, there are arguably two separate aspects of concern. On the one hand, there is the possibility that, if the level of output in the economy were broadly stable, subcentral authorities might decide to alter their budgets and disturb the equilibrium. On the other hand, there is the possibility that, if the level of output in the economy were unstable and, more particularly, if it were to move in a cyclical fashion, then subcentral authorities might react to changes in incomes in a destabilising rather than a stabilising way. These two aspects will be considered in turn.

The impact on demand of changes in subcentral budget levels following a widespread policy decision for larger or smaller budgets, or for a switch in the mix of different types of revenue or expenditure, is clearly something that the central government will have to bear in mind. It needs to be said, however, that, in practice, the impact of such changes may often be small. Suppose, for example, that as a result of separate decisions in each authority there is a 10 per cent increase in aggregate subcentral tax and charge revenue used to finance an increased level of spending on goods and services.[7] Suppose, too, that the initial aggregate level of subcentral spending was some 20 per cent of demand and output. Now, there are two reasons why a 10 per cent rise in taxes and charges does not mean that total planned subcentral spending will rise by 10 per cent thereby causing a sizeable 2 per cent increase in aggregate demand. First, it is likely that a large proportion of subcentral revenue accrues from loans and lump-sum grants. Suppose, for convenience, that these two sources amount to three-eighths of subcentral revenue, so that they amount to 7.5 per cent of aggregate spending, while subcentral taxes and charges amount to 12.5 per cent. In that case, a 10 per cent rise in tax- and charge-financed expenditure will itself cause aggregate demand to rise by 1.25 per cent. Secondly, the rise in taxes and charges will cause a substantial offsetting fall in consumers' spending. The rise in taxes and charges cuts consumers' disposable incomes by 1.25 per cent of output, and if their marginal propensity to consume were, say, 0.9 then the offset would be some 1.125 per cent of output. Thus the net effect of the subcentral authorities' action would be to raise demand by 0.125 per cent. This is a small effect from a large initial change in budgets. It should be observed, though, that if the subcentral taxes concerned could be deducted from taxable incomes in respect of the central government's personal income tax, then disposable

incomes and consumption would fall by less. For example, if the average rate of central income tax was just over 33 per cent, then the fall in disposable incomes would be around 0.83 per cent of output, and the fall in consumption would be 0.75 per cent, so that the net effect on output would be to raise demand by 0.5 per cent. This is clearly a more serious effect.

As it happens, Foster *et al.* (1980, p. 375) have argued that the effects calculated in the last paragraph may be overestimates. Their reasoning can be outlined more simply in the case where subcentral tax payments cannot be deducted from taxable incomes for central income tax purposes. In this case, they argue that it is inappropriate to estimate the offset by applying the marginal propensity to consume to the change in disposable incomes. It would be right to do that only if the extra charge and tax revenue had been added to subcentral balances, because in that case citizens would have suffered a decline in possible living standards and would be likely to react by reducing their saving levels. However, the revenue is used to finance extra public spending, and there is no necessary decline in the overall level of living standards. The budget changes could simply be a result of a desire by electorates to reduce private consumption in favour of public consumption, and there is no reason why electorates should wish to react by reducing their saving levels. If this line of reasoning were correct, then the budget change under consideration would have no effect on aggregate demand at all. This line of reasoning must be qualified in a number of ways. First, some citizens are likely to be better off overall as a result of the budget increase and others worse off. The former group may contain many poor people with low marginal propensities to save, and the latter group may contain many rich people with high ones. Accordingly, aggregate saving could still fall even though average living standards may stay constant. Secondly, some of the rise in taxes may be exported, particularly perhaps if the taxes are levied on businesses, and there could be effects on the saving of their owners, customers and employees who finally bear the burden. Jackman (1982, p. 84) has suggested reducing the need for this second qualification by levying taxes on businesses at nationally uniform rates, although that would mean that business taxes would not vary in line with the various levels of the services they enjoy in different areas. Thirdly, there could be a fall in saving if the increased taxes were used to finance capital projects.

Even if this line of reasoning is only partially correct, or, for that matter, false, the effects on aggregate demand of likely aggregate changes in tax- and charge-financed expenditure would still seem to be very small unless subcentral tax payments could be deducted from taxable incomes for the central government's personal income tax. The concern felt by any government will, of course, depend on the precise size of effect likely to be encountered in its own country, and the degree to which it feels its policies are sensitive to small changes. The widespread disaffection for fine-tuning that occurred in the 1970s suggests that small changes might be more willingly tolerated than they used to be. Of course, it should be noted that if the expenditure changes in the example had been financed by loans or by reductions in balances then there would have been no direct

offset at all, and demand would have risen by some 1.25 per cent. This would have been of understandable concern to the central government.

Interest in the behaviour of subcentral budgets during the period of the trade cycle was stimulated by the observation in the United States that state and local budgets moved from an aggregate deficit in 1929 to an aggregate surplus in 1934, despite the tremendous contraction in the economy. This led in the 1940s to the formulation by Hansen and Perloff (1944, p. 199) of the 'perversity hypothesis' stating that subcentral governments tended to increase rather than reduce fluctuations in the economy. While this may have been true in the United States in the 1930s, it may not be true for all countries at all times. In a comparative study of six countries, Snyder (1973, p. 209) found that in the period 1955–65 the budgets of subcentral authorities in France, Italy, West Germany, the United Kingdom and the United States itself all had stabilising effects on gross national products. Moreover, subcentral authorities in the United States were found to be no less stabilising than those in the other countries. By way of contrast, the subcentral authorities in Belgium did have a destabilising effect.

Rafuse (1965, p. 66) has observed that subcentral fiscal behaviour will be perverse if subcentral revenues are less elastic in respect to changes in output and incomes than subcentral expenditures. The elasticity of the revenues depends on the taxes used and the prevailing circumstances. Suppose that real incomes fluctuate while prices are stable; in this case, the application of constant or near-constant tax rates would very likely mean that the revenues from taxes on income and expenditure would fluctuate more than the revenues from property taxes, so that property taxes would be the most perverse. Suppose, instead, that real incomes fluctuate during a period of fairly stable inflation; in this case, the application of constant or near-constant tax rates would mean that the real yields of taxes on income and expenditure would fluctuate in the same direction (if not always in the same proportions) as changes in incomes, thereby discouraging perversity, while the real yield of property taxes might fall fairly steadily (unless the base is frequently revalued), thereby aiding perversity in booms but opposing it in depressions. Of course, the effects of fluctuations on tax yields will be largely neutralised by corresponding fluctuations in tax rates or expenditures unless it is somehow arranged that subcentral authorities rely more on other sources of income, in particular grants and loans, during periods of low output than in periods of high output.

It is now appropriate to consider how much concern governments of monetarist and Keynesian persuasions might have with regard to subcentral budgets. For simplicity, the analysis will focus on the case where there is an increase in aggregate subcentral spending financed by higher borrowing. Very similar, although slightly smaller, effects on the level of aggregate demand and, in turn, employment would be expected if higher borrowing were matched by lower taxes and charges. Much smaller effects would be expected if increased spending were matched by higher taxes or charges; indeed, it has been noted that higher taxes or charges could cause changes in consumption that would wholly offset the increase in subcentral spending.

The effects of an increase in subcentral spending financed by loans vary according to whether the economy is closed or open, and with an open economy they depend on whether the exchange rate is floating or fixed. These varying circumstances will be taken in turn. In a closed economy, a government of monetarist persuasion will expect the increase in spending to have very little net effect on demand, for it will expect any tendency for it to cause aggregate demand to rise to be largely offset by falls in investment brought about by induced rises in interest rates.[8] A government of Keynesian persuasion, by contrast, may believe that the increase in subcentral spending will cause very little rise in interest rates, and so induce very little offsetting reduction in investment; hence it may expect an appreciable effect on demand and, in turn, on output.[9]

In an open economy with a floating exchange rate, a government with monetarist views will still expect an increase in subcentral spending to have little effect on demand, though for rather different reasons. It will again expect the increase to put upwards pressure on interest rates, but this should now induce an inflow of capital from abroad, which will cause the exchange rate to rise; in turn, the demand for tradeable goods – that is, exports and import substitutes – will fall, and this deflationary tendency will ameliorate the rise in interest rates and any fall in investment. In these circumstances, the increase in subcentral spending could still be largely offset by falls in other spending, but these falls may take place more in tradeable goods than in investment; indeed, if there were perfect mobility of capital between countries, then interest rates and so investment would end up at their original levels. A government with Keynesian views will still expect the increase in subcentral spending to raise output, but it will expect a larger increase than in the closed economy case. This is because the rise in output previously anticipated will be expected to lead to a rise in imports, thereby causing the exchange rate to fall and, in turn, the demand for tradeable goods to rise; this rise should stimulate a further rise in output.

The effects with fixed exchange rates are very different, partly because the absence of moves in the exchange rate prevent the changes in the demand for tradeable goods discussed in the last paragraph. As before, a government with monetarist views will expect an increase in subcentral spending to put upwards pressure on interest rates and to lead to an inflow of foreign capital, but this should now cause a balance of payments surplus and a rise in the money supply; consequently, interest rates might end up little higher than they were initially, and investment little lower, so there would be little to offset the increase in subcentral spending; hence demand could rise appreciably. A government with Keynesian views, on the other hand, will still expect the increase in subcentral spending to lead to a rise in imports, but this should now cause a balance of payments deficit and a fall in the money supply; this might produce a slight rise in interest rates and a slight fall in investment, and hence a rather smaller effect on output than anticipated in the closed economy case. In view of the prevalence of floating exchange rates at the present time, they will be presumed in the rest of this chapter to be in operation, and the results obtained in this paragraph will be ignored.

The impact of subcentral authorities on inflation, the balance of payments and growth

The previous section considered the effects of subcentral authority budgets on output levels. Attention now focuses on the effects of subcentral authority budgets on other matters of macroeconomic concern. Governments of a Keynesian persuasion, for example, believing that a change in subcentral spending may have an appreciable effect on output (unless they are financed by changes in taxes or charges, when the effect will be largely neutralised), may expect the induced change in output to affect in turn the rate of inflation.[10] It has been seen that they will also expect the change in output to affect the level of imports and so the balance of payments.

Monetarist governments, in contrast, would not expect changes in subcentral spending to have much effect on output and would not be much concerned by these possibilities. They would, however, have other worries. They would expect any rise in subcentral spending to disturb the balance of payments by causing a rise in capital flows from abroad that would be offset by a larger deficit (or a smaller surplus) on the current account. They might also fear that the rise in subcentral spending could raise interest rates a little, thereby causing falls in investment and the growth rate. Incidentally, it is arguable that the real cost of the higher subcentral spending would be felt in years to come either if it led to higher borrowing from abroad, since the loans would eventually have to be repaid, or if it led to lower investment, since this would mean a lower capital stock in the future. Keynesian governments would not generally share these concerns, but they might be worried about the possibility that increased subcentral spending would lead to a 'crowding out' of investment if the increase took place at a time of full employment.

It should be emphasised that subcentral authorities can affect macroeconomic variables in ways unrelated to their effects on output or interest rates. For example, an increase in borrowing from overseas would have a direct effect on the balance of payments. Also, governments influenced by 'supply-side' economics may be committed to the view that high levels of public spending, including high levels of subcentral authority spending, lead to high tax rates, which can act as a disincentive in respect of the supply of labour, capital and enterprise, and thereby have adverse effects on both the level of output and its rate of growth.

Another effect, much discussed in recent years,[11] is that rises in subcentral taxes could affect the level of wage increases demanded by trade unions. In particular, attention has focused on the effects that increases in the rates of taxes on property levied by subcentral authorities might have on published price indices. Cost-of-living indices typically reflect increases in the rates of taxes on property, and indeed all taxes on expenditure, and are often referred to in wage negotiations. Thus central governments may be particularly concerned to contain increases in subcentral authority property tax rates in inflationary times in order to help moderate rises in the cost-of-living index, and so, in turn, to moderate wage demands. Of course, increases in property tax rates will be needed

in inflationary times merely to maintain the real yield of the tax, so central governments may concentrate their efforts in seeking to discourage any further increases desired in order to raise service levels. At the same time, however, central governments may console themselves by considering that these further tax increases will cut consumers' real disposable incomes and hence reduce their demands for other goods and services; as a result, the prices for these items may fall, or at least rise more slowly than they would otherwise, and this could neutralise much of the rise in the price index caused by the tax rises.

In any case, it is unsatisfactory that those people who are responsible for calculating the published indices do not discriminate between property tax rate rises needed to finance a given level of services when costs rise, and rises needed to finance a higher level of services, for it is arguable that they should take into account only the first sort of increase.[12] It is, after all, only the first sort of increase that reflects an increase in the unit prices citizens must pay for goods and services, in this case, those supplied by subcentral authorities; the latter sort of increase reflects, in essence, a decision to purchase a greater quantity of goods and services. If the index omitted the second sort of rise, then those involved in collective bargaining would have a truer idea of the rate at which the cost of living was rising, and central governments would be unconcerned (in this context) about property tax increases used to pay for higher levels of subcentral services. However, it would not be easy to distinguish clearly between the two sorts of rise, particularly if subcentral authorities levied more than one tax; and the distinction would somehow have to be made in respect of all central taxes on expenditure also.

It should be noted that increases in charges by subcentral authorities will also be included in the price indices, but there will be limited concern over charge levels. If charges for subcentral services rise more (or less) than the amounts needed to cover any increases in unit costs, then their effect on the index will appear detrimental (or beneficial) to the central government. However, this rise will be offset by a reduction (or increase) in any subsidies to the services from tax revenue and the subsequent fall (or rise) in tax rates will largely offset any detrimental (or beneficial) effect of charges on the index.

So far, the concern with subcentral spending has generally related to its level, and changes in that level, although it has been noted that there may be less interest in purchases of existing assets than in other spending, and perhaps a shade less concern with subsidies and transfer payments than with purchases of new goods and services. However, it has been suggested (see CIPFA Study Group, 1977, p. 14) that, at times, more detailed changes in the composition of spending may have macroeconomic implications. For instance, some spending (such as housing subsidies) might affect the recorded rate of inflation more directly than other spending. The rate of growth might be helped more by spending on certain services, education perhaps, than others, or by a switch from current spending to capital spending. Furthermore, the level of imports may be raised more by increases in spending on some services, perhaps through induced imports of foreign vehicles, than by increases in spending on others. However, it would be hard to quantify the importance of examples such as these.

Regulating borrowing by subcentral authorities

The previous two sections have indicated a number of ways in which subcentral authorities can have an impact on the central government's macroeconomic policies. Mostly they are able to have such an impact by causing changes in the level of output or interest rates, and this they can do most effectively by matched changes in their borrowing and spending levels, or by changes in the mix of taxes and charges on the one hand and loans (and reductions of balances) on the other within a given revenue total. They can, perhaps, also cause changes to a limited degree by matched changes in spending and taxes or charges. It may be assumed that the broad pattern of spending is of little interest, for the effects of spending on subsidies or transfer payments will not be very different from those of spending on goods and services, and spending on purchases of existing assets, although it may have different effects, is typically a very small percentage of total spending.

It follows that the central government's primary concern with what subcentral authorities are likely to do is over their total level of borrowing. Changes in this level could have a significant macroeconomic impact, whether they were accompanied by matched changes in spending or offsetting changes in revenue from taxes or charges. It may be added that, in order to encourage subcentral authorities to act in a stabilising way during the course of trade cycles, they should be encouraged to borrow more in periods of depression and less in periods of boom.

It was suggested by the Layfield Committee (HMSO, 1976, p. 250) that one way of controlling aggregate subcentral borrowing would be to require specified proportions of capital spending to be borne from a surplus of current revenue over current spending and from loans; they noted that such an arrangement operated in Denmark. There are perhaps two principal objections to this proposal. First, it was argued in Chapter 8 that capital spending should be financed wholly from loans in order to secure intra- and inter-generational equity and to promote optimal spending levels. Secondly, the level of capital spending by subcentral authorities might fluctuate quite appreciably from year to year, and so too would the level of loans if they were fixed in relation to capital spending; thus this method might not help the central government to make the good forecasts of the aggregate level of subcentral borrowing that it would need if it wanted to take appropriate compensating expansionary or contractionary measures when that level seemed likely to be lower or greater than desired. From the central government's point of view, it might be more satisfactory for subcentral authorities to be required to make their aggregate levels of capital spending and loans equal to some specified fraction of their current spending, because fluctuations in the level of current spending are typically small.

In practice, various devices have been used to try to constrain subcentral borrowing. In the United Kingdom, for example, local authority borrowing from 1971 to 1981 was controlled by requiring authorities to obtain loan sanctions before they could borrow.[13] This arrangement was originally intended as a safeguard against impropriety (see the Layfield Committee,

HMSO, 1976, p. 242), but was later used to try to control the total level of borrowing and, indeed, capital expenditure. Loan sanctions were given for individual projects and for groups of projects, although the trend was to make sanctions wider in order to give local authorities more flexibility. From a macroeconomic point of view, this system had two drawbacks. First, even where permitted projects were undertaken, the relevant borrowing and expenditure were often spread over several years, so the actual borrowing and spending in any one year were not subject to close control. In an attempt to solve this problem, the United Kingdom government substantially modified the scheme in 1981 so that each authority is now given a total capital expenditure cash limit each year, which also serves as a total borrowing sanction; there is limited scope for capital spending permitted in one year to be deferred to the next, although spending entitlements can be transferred between individual authorities. Secondly, the arrangements fixed upper limits on borrowing until 1981, and they now fix upper limits on capital expenditure, but they do not ensure that the actual levels will be at, or even near, those limits. Indeed, in 1982, the central government appeared concerned that local authorities were spending about 24 per cent less than they were permitted to do.

In federal countries, central governments tend not to control state borrowing directly (since this would ordinarily be unconstitutional), and any control of subcentral borrowing tends to be a matter for the states. The states, in this context, are not so much interested in macroeconomic matters as with prudence. In the United States, for example, Maxwell and Aronson (1977, pp. 207–8) have explained that most states limit the power of the state legislature to borrow in order to guard against misdirected overborrowing, and states limit the power of local authorities to borrow, often by imposing debt ceilings related to the property tax base, in order to protect the solvency of local governments and to protect bondholders.

There are, however, central arrangements in some federal countries to try to control subcentral borrowing. In West Germany, a law was passed in 1972 relating to economic stability and growth, which permits the federal government to limit the total amount of borrowing by the *Länder* and local authorities.[14] In practice, there are several reasons for regarding this power as rather less potent than it might appear. First, the federal government needs the consent of the upper house, the Bundesrat, which comprises representatives of the *Länder*. Secondly, such borrowing limits apply only for a year at a time. Thirdly, the *Länder* have the power to relax the limitations on local authorities by an aggregate not exceeding 30 per cent.

The Australian arrangements might appear to give the federal government less power than its counterpart in West Germany, but in practice it almost certainly has more power. In theory, the power to regulate borrowing by the central government, the states and local authorities rests with the Australian Loan Council, which was established formally in 1927.[15] This council consists of the six state premiers and the commonwealth prime minister. Each state premier has one vote while the commonwealth premier has two and a casting vote. It might seem, therefore, that the central government is in a weak position because it could

be outvoted by the states, but in general it is in a strong position because, as Wiltshire (1980, p. 82) has explained, it needs the support of only two states, and there have been very few occasions when there were not at least two states governed by the same party as the one in control in Canberra. This factor does not guarantee that the central government will have its way completely for while it needs only a majority vote to endorse its preferred figure for total public borrowing, it needs a unanimous vote to secure its preferred allocation of that total between the three tiers of government; failing unanimity, the Constitution specifies a fallback allocation, but in fact it has never been necessary to apply that. Indeed, in practice, Mathews (1982a, p. 155) has claimed that the central government is now able to make unilateral decisions about all aspects of public borrowing, and can simply announce these to the Loan Council, which has ceased to be a deliberative body. There are a number of reasons for this central dominance. They include the facts that states rely heavily on grants from the central government and, more particularly, that they have had one-third of their approved capital spending financed by such grants since the early 1970s.

Quite apart from concern over the total level of loans, it was noted in the previous section that the central government would be concerned with the level of loans made from overseas to subcentral authorities. There is clearly a case for controlling such loans. The central government could fix a desired total of foreign loans and allow each authority a quota, defined in relation to a variable such as its total capital spending, total borrowing or merely population. In the United Kingdom, for example, individual local authorities must ask for central government permission whenever they wish to borrow abroad. In Australia, borrowing abroad is generally available only to the central government, although the states have some power to raise foreign loans for infrastructure improvements.

Regulating subcentral taxation and expenditure

Even if the levels of domestic and foreign subcentral borrowing were fixed, the central government might have some concern over changes in subcentral spending matched by changes in subcentral taxes and charges, although, as discussed earlier, such matched changes might have negligible effects on the level of output or the rate of interest unless subcentral tax payments could be deducted from taxable incomes for the central government's personal income tax. The concern over such changes might be increased if it were thought that tax rises increased inflationary pressures or caused disincentives. It is perhaps understandable that the most common action suggested for central governments in this context is to restrain the rate at which tax rates can rise, for instance by imposing a mandatory limit such as 3 per cent on the extent to which tax-financed expenditure can rise from year to year (King, 1973, p. 58), or by permitting tax revenues to increase above some such norm only on condition that a specified part of any extra tax revenue resulting from a tax rate in excess of that warranted by the norm should be surrendered to the central government

(the Layfield Committee, HMSO, 1976, pp. 248–9). It can be argued that such norms should be as generous as possible for, in the end, it is the central government's responsibility to manage the economy with the level of subcentral spending desired by subcentral electorates.

There remains the possibility that subcentral authorities may not plan any changes, yet the central government may wish to impose some. It may, for instance, wish to reduce aggregate demand and feel that all components of aggregate demand should be reduced, including subcentral spending. Alternatively, it may wish to reduce some parts of aggregate demand, such as consumers' spending and subcentral government spending, so that resources can be released to permit increases in others, such as private investment or exports. As Jackman (1982, p. 81) has observed, many monetarists have for a long time argued that governments should not attempt to tune the level of aggregate demand, but they would accept attempts to change the composition of demand. In his view, the attempts to restrain local government spending in the United Kingdom in the early 1980s, made by a central government sympathetic to monetarist tenets, resulted more from a concern with the share of resources consumed by local authorities than with any worries about demand management. This hypothesis is consistent with that government's concern that rising property taxes on businesses would raise their costs and reduce their output and employment levels, even though rising local budgets (as explained earlier) could slightly increase overall output and employment levels.

In most countries, a sizeable reduction in subcentral spending can be encouraged by a reduction in revenue-sharing grants.[16] It is true that the fall in spending might not take place for some time if subcentral authorities maintained expenditure levels for a while by reducing balances. Further, even when the fall does take place, it might be less than the fall in grants since subcentral authorities could soften the impact on services by raising taxes or charges. Indeed, they may feel obliged to meet much of the grant cut in this way owing to the serious problems incurred when trying to make sharp reductions in service levels in a short time period. However, this possibility may not matter as far as demand management is concerned, as the rises in taxes or charges will lead to cuts in consumption. Of course, the result of limited falls in subcentral authorities' spending accompanied by rises in their taxes may be a pattern of aggregate demand in which, in the cental government's view, there is too much subcentral spending in relation to consumers' spending and, perhaps, other components of aggregate demand. It may not be unduly worried, however, for if electors do not share its view then it will not be criticised by them, and, if they do, then they are likely to pressurise subcentral governments to reduce their spending.

It has been implied by Foster *et al.* (1980, p. 373) that general cuts in grants are an indiscriminate way of cutting subcentral spending because they penalise low-spending authorities as much as high-spending ones. It is not clear, however, that grants should be cut only for high-spending areas; there seems little case for extracting the whole sacrifices demanded by the central government's macroeconomic policy from those citizens who innocently prefer a relatively high subcentral spending to consumption

ratio. The central government might feel that a greater drawback of cutting grants is that they could make it unpopular if they led to rises in subcentral tax rates, and indeed this might well happen. It might be equally unpopular, however, if its grant cuts led to falls in service levels. It might also be worried that induced subcentral tax rate increases would aggravate wage demands, particularly if they raised published price indices. However, as noted earlier, it may console itself by considering that the tax rises will reduce consumers' spending and lower their demands for other items, thereby perhaps resulting in lower prices (or lower price rises) for those items and having an offsetting improvement in the price indices. In any case, it is also possible that cuts in services could aggravate wage demands as people sought the extra income needed to pay extra subcentral taxes with the aim of restoring services to their former levels.

In countries where there are no revenue-sharing grants, the central government may have to control subcentral spending indirectly by reducing the disposable income of consumers, probably chiefly by tax increases, and allow them subsequently to pressurise subcentral authorities into reducing their tax rates and spending. In this system, any deflation would be achieved primarily by the central government tax rises. It has been suggested (CIPFA Study Group, 1977, p. 11) that this system should be used even where revenue-sharing grants are in operation. However, it has the slightly curious result that central government tax rates will rise while subcentral tax rates will fall, and it may be felt preferable to moderate the rises in the former at the expense of some reduction in grants, thereby inducing some rises in the latter. In other words, it might be desired to use the mixture of central government tax rate rises and cuts in grants that resulted in subcentral taxes rising in about the same proportions as central ones.

Another possibility is that a central government, especially one that adopted supply-side economics tenets, might chiefly seek a cut in subcentral tax rates. It could achieve this, perhaps, by raising grants to subcentral authorities, but any benefits might then be largely or wholly offset by the rises in its own tax rates needed to finance the extra grants. Of course, it could finance the grants by borrowing or encourage subcentral authorities to increase their loans and set lower taxes, but it might not want the increases in aggregate demand that would follow. In such circumstances, it migh desire powers to enforce matched cuts in subcentral spending and taxation. These powers could be justified only if it were thought that the gains from the increased incentives outweighed the losses from preventing subcentral authorities from providing the service levels desired by their electorates.

Finally, the central government may wish to alter the pattern of subcentral spending on goods and services. The most likely situation is probably one in which it wished to promote economic growth by encouraging subcentral spending on any particular service that was believed to help growth (such as education, perhaps), or by encouraging a general move from current spending to capital spending. Particular services can be most effectively stimulated by open matching specific grants (as explained in Chapter 3). A switch from current spending to capital

spending might be achieved by reducing revenue-sharing grant payments and easing borrowing controls if authorities were required to use the grants for current spending and required to finance capital spending from loans. A switch would be harder to achieve, however, if authorities could finance some capital spending from a current account surplus; in that case, a reduction in grants and an easing of borrowing controls might reduce the current surplus and raise the level of loans, but leave the levels of current and capital spending substantially unaltered.

Summary

Central government economic policies can be broadly divided into allocation, distribution and stabilisation policies. The reasons for concern over subcentral economic behaviour are rather different in each case.

With resource allocation, it will presumably be expected that lower-tier authorities will be actively involved in formulating and implementing policies in their own areas. The main central concern will be to see that the system of subcentral government is so devised that each authority tends to produce its services at reasonable costs, and at efficient levels, without stimulating migration of labour, trade, capital or households except where households vote with their feet for preferred levels of services and taxes. It is suggested that these concerns can be greatly eased by determining the authority sizes, the list of devolved taxes and, also, grant policies in the light of the discussions in earlier chapters.

Subcentral authorities may be allowed some explicit role in redistribution through taxes and transfer payments or through merit good provision. Even if they had no such role, they would still affect the post-budget distribution of incomes with their allocation policies. As different levels of redistribution in different areas could lead to substantial migration, the central government is likely to have prime control over redistribution levels. It may then desire controls over lower-tier authorities whose own policies prevent it from achieving its desired redistributionary impact; it may also desire controls to ensure that a few individual authorities are not able to have a very regressive (or unprogressive) impact that others then have to follow.

As far as stabilisation is concerned, responsibility will probably lie entirely with the central government, except perhaps over regional development, although even here it may not permit complete subcentral autonomy. In general, the main concern is over the destabilising effects that changes in aggregate subcentral spending could have on the economy. Keynesian governments will expect the effects to be felt chiefly on aggregate demand, and in turn output, and so on the level of employment, the rate of inflation and spending on imports. Monetarist governments will expect them to be felt initially on interest rates, and so on investment and the rate of growth, but more importantly on capital flows from abroad and, in turn, the exchange rate and the output of exports and import substitutes. The effects will be small, and conceivably non-existent, however, if the spending changes are matched by changes in charges or taxes (other than

taxes that can be offset against liability for central taxes), but will be much more significant if matched by changes in borrowing levels. Accordingly, there is an argument here for controls over subcentral borrowing, though it may not be easy to devise satisfactory controls. For example, controls that seek to put a ceiling on subcentral borrowing levels are not sufficient to ensure stable or predictable levels since the actual level may often be lower than the ceiling.

There are a number of other macroeconomic reasons why a central government might be concerned with lower-tier authorities. For example, it could be worried about the effects if those authorities borrowed from abroad and so disturbed the balance of payments; such borrowing is often regulated. It might fear that rises in some subcentral taxes (such as the property tax) would raise cost-of-living indices and exacerbate wage demands, although the tax rises should reduce consumer spending and ease inflation elsewhere, thereby perhaps having a trivial net effect on the indices. It might seek to reduce subcentral spending in order to cut aggregate demand or to release resources for investment; cuts in revenue-sharing grants might have the desired effects on the level of subcentral spending, but if this level were sustained by higher subcentral taxes, thereby causing consumption rather than subcentral spending to fall, then the central government might not be unduly concerned. However, if the central government cut revenue-sharing grants in order to reduce subcentral spending either for ideological reasons or in the belief that a smaller public sector with lower taxes would increase incentives, then it would be concerned about grantees that sustained their spending levels with higher taxes; the central government might then seek further controls on subcentral spending. An objection to such controls is that they would reduce the freedom of subcentral electorates to choose how to allocate the funds available to them between public and private spending.

Notes

1 For a study showing the sensitivity of those on low incomes to subcentral service levels, see King (1973, pp. 51–6). A similar sensitivity would be expected in all countries and is implied in the United States, for example, by the work of Gillespie (1965). However, King and Gillespie have broadly estimated the value of benefits from any service as total expenditure divided by the number of beneficiaries. The benefits to the poor might well have lower values if they were valued in accordance with recipients' marginal valuations. For a discussion of this method, see Aaron and McGuire (1970).

2 For a discussion of this issue, see King (1973, pp. 56–60) and Foster *et al.* (1980, pp. 369–81).

3 The monetarist and Keynesian (or fiscalist) positions are explained in many macroeconomic texts. See, for example, Branson and Litvack (1976, pp. 302–10).

4 For a discussion of this point, see CIPFA Study Group (1977, pp. 10 and 14).

5 Strictly speaking, the increases in spending discussed in this paragraph relate to how far the various types of subcentral spending shift the *IS* schedule to the right in an *IS–LM–BP* model of a four-sector economy. The final effects on demand depend principally on the interest-elasticity of the *IS, LM* and *BP* schedules. These effects are considered at the end of this section. For a more formal presentation of an *IS–LM–BP* model, see, for example, Dernburg and McDougall (1980, pp. 148–55 and 342–55).

6 Strictly speaking, the reductions in spending discussed in this paragraph relate to how far the various types of subcentral revenue shift the *IS* schedule to the left.

7 The example is this paragraph considers only the effects on demand of the changes in taxes, charges and spending. In effect, it considers only the induced shifts in the *IS* schedule.

8 The text supposes that a monetarist government would believe the *LM* schedule to be interest-inelastic in the relevant range, and to be less elastic than the *BP* schedule for an open economy. The final effect on output also depends on the price-elasticity of the aggregate supply schedule; for a discussion of this schedule, see, for example, Dernburg and McDougall (1980, Chapter 9). A monetarist government may balance this schedule to be inelastic, in which case any change there is in aggregate demand will affect the price level more than output. To present text suggests that a monetarist government would not expect changes in subcentral authorities' budgets to have much effect on aggregate demand unless those authorities belonged to an open economy with fixed exchange rates.

9 The text supposes that a Keynesian government would believe the *LM* schedule to be interest-elastic in the relevant range, and to be more elastic than the *BP* schedule for an open economy. It also supposes that a Keynerians government would believe the aggregate supply schedule to be sufficiently price elastic (at least if output is below the full employment level) for changes in aggregate demand to have significant effects on output.

10 Until the 1960s, there appeared in many countries to be a trade-off between the level of unemployment and the rate of wage inflation. Since then, it has been argued that changes in the level of unemployment affect the rate at which the rate of inflation changes; at the so-called 'natural' level of unemployment, the rate of inflation is expected to be stable. For a discussion of this issue, see, for example, Dernburg and McDougall (1980, pp. 281–6).

11 See, for example, CIPFA Study Group (1977, p. 10) and Foster *et al* (1980, p. 378).

12 It has been argued that more attention should be paid to other price indices (such as one relating to gross domestic product at factor cost) on which taxes on expenditure have no direct effect. See, for example, Jackman and Klappholz (1975, pp. 46–9).

13 For a discussion of the pre-1981 arrangements, see Hepworth (1980, pp. 137–40). For an outline of the new arrangements, see Watt (1982).

14 For a discussion of this law, see ACIR (1981b, pp. 54–6) and Hunter (1973, pp. 21–6).

15 For discussions of the Australian Loan Council, see Jay (in ACIR, 1981c, Chapter 6) and Wiltshire (1980, pp. 79–83).

16 Similarly, sizeable increases in subcentral spending can be encouraged by an increase in those (or, no doubt, any other) grants. The evidence for the 'flypaper' effect discussed in Chapter 3 suggests such grant rises may be very effective.

Bibliography

This bibliography lists all the works referred to in the text. The following abbreviations are used:

ACIR	The Advisory Commission on Intergovernmental Relations
CIPFA	The Chartered Institute of Public Finance and Accountancy
Cmnd	Command Paper Number
CRFFR	The Centre for Research on Federal Financial Relations
HMSO	Her Majesty's Stationery Office
OECD	The Organisation for Economic Co-operation and Development

Items marked with an asterisk (*) are reprinted in Grewal, Brennan and Mathews (1980).

ACIR (1980) *The State of State–Local Revenue Sharing* (Washington, DC).

ACIR (1981a) *Regional Growth: Interstate Tax Competition* (Washington DC).

ACIR (1981b) *Studies in Comparative Federalism: West Germany* (Washington, DC).

ACIR (1981c) *Studies in Comparative Federalism: Australia* (Washington, DC).

ACIR (1981d) *Studies in Comparative Federalism: Australia, Canada, the United States and West Germany* (Washington, DC).

Aaron, H. J. (1975) *Who Pays the Property Tax? A New View* (Washington, DC: Brookings Institution).

Aaron, H. J. and McGuire, M. C. (1970) 'Public goods and income distribution', *Econometrica*, Vol. 38, No. 6 (November), pp. 907–20.

Akin J. S. (1973) 'Fiscal capacity and the estimation method of the Advisory Commission on Intergovernmental Relations', *National Tax Journal*, Vol. 26, No. 2 (June), pp. 275–91.

Allan, C. M. (1971) *The Theory of Taxation* (London: Penguin Books).

Aronson, J. R. (1971) 'A comment on optimality in local debt limitation', *National Tax Journal*, Vol 24, No. 1 (March), pp. 107–8.

Aronson, J. R. (1974) 'Financing public goods and the distribution of population in metropolitan areas: an analysis of fiscal migration in the United States and England', in A. J. Culyer (ed.) *Economic Policies and Social Goals*, York Studies in Economics (London: Martin Robertson), pp. 313–41.

Atkinson, A. B. and Stiglitz, J. E. (1980) *Lectures on Public Economics* (London: McGraw-Hill).

Auld, D. A. L. and Eden, L. B. (1983) *Provincial–Municipal Fiscal Equalization and Revenue Sharing*, Department of Economics Discussion Paper 83–1 (Ontario: University of Guelph).

Barnett, R. and Topham, N. (1980) 'A critique of equalising grants to local governments', *Scottish Journal of Political Economy,* Vol. 27, No. 3 (November), pp. 235–49.

Barnett, R. and Topham, N. (1982) 'A critique of a defence of equalising grants to local authorities', *Scottish Journal of Political Economy,* Vol. 29, No. 1 (February), pp. 111–17.

Barr, J. L. and Davis, O. A. (1966) 'An elementary political and economic theory of the expenditures of local governments', *Southern Economic Journal,* Vol. 33, No. 2 (October), pp. 149–65.

Bennett, J. T. and Johnson, M. H. (1980) 'Tax reduction without sacrifice: private-sector production of public services', *Public Finance Quarterly,* Vol. 8, No. 4 (October), pp. 363–96.

Bennett, R. J. (1982) *Central Grants to Local Governments: The Political and Economic Impact of the Rate Support Grant in England and Wales* (Cambridge: Cambridge University Press).

Bergstrom, T. C. and Goodman, R. P. (1973) 'Private demands for public goods', *American Economic Review,* Vol. 63, No. 3 (June), pp. 280–96.

Bewley, T. F. (1981) 'A critique of Tiebout's theory of local public expenditures', *Econometrica,* Vol. 49, No. 3 (May), pp. 713–40.

Bieri, S. (1979) *Fiscal Federalism in Switzerland,* CRFFR Research Monograph 26 (Canberra: Australian National University).

Bird, R. M. (1976) *Charging for Public Services: A New Look at an Old Idea* (Toronto: Canadian Tax Foundation).

Blaug, M. (1967) 'Economic aspects of vouchers for education', in *Education: A Framework for Choice,* Readings in Political Economy 1 (London: Institute of Economic Affairs), pp. 21–47.

Boadway, R. W. (1979) *Public Sector Economics* (Cambridge, Mass.: Winthrop).

Boadway, R. W. and Flatters, F. (1982) *Equalization in a Federal State: An Economic Analysis,* Economic Council of Canada (Ottawa: Canadian Government Publishing Centre).

Borcherding, T. E. (1977) 'The sources of growth of public expenditures in the United States 1902–1970', in T. E. Borcherding (ed.) *Budgets and Bureaucrats: The Sources of Government Growth* (Durham, North Carolina: Duke University Press), pp. 45–70.

Borcherding, T. E. and Deacon, R. T. (1972) 'The demand for the services of non-federal governments', *American Economic Review,* Vol. 62, No. 5 (December), pp. 891–901.

Bradford, D. F. and Oates, W. E. (1971a) 'Towards a predictive theory of intergovernmental grants', *American Economic Review,* Vol. 61, No. 2 (May), pp. 440–8.

Bradford, D. F. and Oates, W. E. (1971b) 'The analysis of revenue sharing in a new approach to collective fiscal decisions', *Quarterly Journal of Economics,* Vol. 85, No. 3 (August), pp. 416–39.

*Brainard, W. C. and Trenery Dolbear, F. (1967) 'The possibility of oversupply of local "public" goods: a critical note', *Journal of Political Economy,* Vol. 75, No. 1 (February), pp. 86–90.

Branson, W. H. and Litvack, J. M. (1976) *Macroeconomics* (New York: Harper and Row).

Brazer, H. E. (1981) 'On tax limitation', in N. Walzer and D. L. Chicoine (eds) *Financing State and Local Governments in the 1980s: Issues and Trends* (Cambridge, Mass.: Oelgeschlager, Gunn & Hain), pp. 19–34.

Break, G. F. (1980) *Financing Government in a Federal System,* Studies of Government Finance (Washington, DC: Brookings Institution).

Brennan, H. G. (1977) 'Criteria for state and local taxes', in R. L. Mathews (ed.) *State and Local Taxation* (Canberra: Australian National University Press and CRFFR), pp. 1–9.

Brennan, H. G. and Buchanan, J. M. (1979) 'The logic of tax limits: alternative constitutional constraints on the power to tax', *National Tax Journal*, Vol 32, No. 2 Supplement (June), pp. 11–22.

Brennan, H. G. and McGuire, T. (1975) 'Optimal public choice under uncertainty', *Journal of Public Economics*, Vol. 4, No. 2 (February), pp. 205–9.

*Breton, A. (1977) 'The theory of local government finance and the debt regulation of local governments', *Public Finance*, Vol. 32, No. 1, pp. 16–28.

Breton, A. and Scott, A. D. (1978) *The Economic Constitution of Federal States* (Canberra: Australian National University Press).

Breton, A. and Scott, A. D. (1980) *The Design of Federations* (Montreal: Institute for Research on Public Policy).

Breton, A. and Wintrobe, R. (1975) 'The equilibrium size of a budget-maximizing bureau: a note on Niskanen's theory of bureaucracy', *Journal of Political Economy*, Vol. 83, No. 1 (February), pp. 195–207.

Brown, C. V. and Jackson, P. M. (1982) *Public Sector Economics*, 2nd edn (Oxford: Martin Robertson).

Browning, E. K. and Browning, J. M. (1979) *Public Finance and the Price System* (New York: Macmillan).

Brownlee, O. H. (1961) 'User prices *vs.* taxes', in *Public Finances: Needs, Sources and Utilization*. National Bureau of Economic Research (Princeton, NJ: Princeton University Press), pp. 421–32.

*Buchanan, J. M. (1950) 'Federalism and fiscal equity', *American Economic Review*, Vol. 40, No. 4 (September), pp. 583–99.

Buchanan J. M. (1952a) 'Federal grants and resource allocation', *Journal of Political Economy*, Vol. 60, No. 3 (June), pp. 208–17.

Buchanan, J. M. (1952b) 'A reply', *Journal of Political Economy*, Vol. 60, No. 6 (December), pp. 536–8.

Buchanan, J. M. (1961) 'Comments', in *Public Finances: Needs, Sources and Utilization*, National Bureau of Economic Research (Princeton, NJ: Princeton University Press), pp. 122–9.

Buchanan, J. M. (1963) 'The economics of earmarked taxes', *Journal of Political Economy*, Vol. 71, No. 5 (October), pp. 457–69.

Buchanan, J. M. (1965) 'An economic theory of clubs', *Economica*, Vol. 32, No. 125 (February), pp. 1–14.

Buchanan, J. M. (1967) *Public Finance in Democratic Process* (Chapel Hill, North Carolina: University of North Carolina Press).

*Buchanan, J. M. and Goetz, C. J. (1972) 'Efficiency limits of fiscal mobility: an assessment of the Tiebout model', *Journal of Public Economics*, Vol. 1. No. 1 (April), pp. 25–43.

Buchanan, J. M. and Kafoglis, M. Z. (1963) 'A note on public goods supply', *American Economic Review*, Vol. 53, No. 3 (June), pp. 403–14.

*Buchanan, J. M. and Wagner, R. (1970) 'An efficiency basis for federal fiscal equalization', in J. Margolis (ed.) *Analysis of Public Output*, National Bureau of Economic Research (New York: Columbia University Press), pp. 139–58.

Burns, R. M. (1977) *Intergovernmental Fiscal Transfers: Canadian and Australian Experiences*, CRFFR Research Monograph 22 (Canberra: Australian National University).

CIPFA (1982) *Finance and General Statistics 1982–83* (London).

CIPFA Scottish Branch (1982) *Rating Review – 37th Year of Issue* (Hamilton, Glasgow).

CIPFA Study Group (1977) *Local Government Finance and Macro-Economic Policy* (London).

Castles, A. C. (1974) 'Responsibility sharing in a federal system: political and legal issues', in R. L. Mathews (ed.) *Fiscal Federalism: Retrospect and Prospect*, CRFFR Research Monograph 7 (Canberra: Australian National University), pp. 83–103.

Cebula, R. J. (1979) 'A survey of the literature on the migration–impact of state and local government policies', *Public Finance*, Vol. 34, No. 1, pp. 69–84.

Coase, R. H. (1960) 'The problem of social cost', *Journal of Law and Economics*, Vol. 3, No. 1 (April), pp. 1–44.

Collins, R. A. (1977) 'Optimality in municipal debt, *Public Finance Quarterly*, Vol. 5, No. 1 (January), pp. 117–26.

Collins, R. A. (1979) 'Optimality in municipal debt: a reply', *Public Finance Quarterly*, Vol. 7, No. 3 (July), pp. 364–7.

Commission of the European Communities (1977) *Report of the Study Group on the Role of Public Finance in European Integration*, Volume II (Brussels).

Commonwealth Grants Commission (1982) *Report on State Tax Sharing and Health Grants 1982: Volume I – Main Report* (Canberra: Australian Government Publishing Service).

Courant, P. N., Gramlich, E. M. and Rubinfeld, D. C. (1979) 'The stimulative effects of intergovernmental grants or why money sticks where it hits', in P. Mieszkowski and W. H. Oakland (eds) *Fiscal Federalism and Grants-in-Aid*, Coupe Papers on Public Economies (Washington, DC: Urban Institute), pp. 5–21.

Cripps, F. and Godley, W. (1976) *Local Government Finance and its Reform* (Cambridge: University of Cambridge Department of Applied Economics).

Crispin, A. (1980) 'The new block grant and education – central control or national direction?' *Local Government Studies*, Vol. 6, No. 6 (November/December), pp. 25 – 37.

Dafflon, B. (1977) *Federal Finance in Theory and Practice with Special Reference to Switzerland* (Berne: Paul Haupt).

Daly, G. C. (1969) 'The burden of the debt and future generations in local finance', *Southern Economic Journal*, Vol. 36, No. 1 (July), pp. 44–51.

Davis, A. and Lucke, R. (1982) 'The rich-state – poor-state problem in a federal system', *National Tax Journal*, Vol. 35, No. 3 (September), pp. 337–63.

Davis, O. A. and Haines, G. H. (1966) 'A political approach to a theory of public expenditure: the case of municipalities', *National Tax Journal*, Vol. 19, No. 3 (September), pp. 259–75.

Davis, O. A. and Winston, A. (1962) 'Externalities, welfare and the theory of games', *Journal of Political Economy*, Vol. 70, No. 3 (June), pp. 241–62.

Dernburg, T. F. and McDougall, D. M. (1980) *Macroeconomics*, 6th edn (New York: McGraw-Hill).

Due, J. K. (1961) 'Studies of state–local tax influences on location of industry', *National Tax Journal*, Vol. 14, No. 2 (June), pp. 163–73.

Edel, M. and Sclar, E. (1974) 'Taxes, spending, and property values: supply adjustment in a Tiebout–Oates Model', *Journal of Political Economy*, Vol. 82, No. 5 (September/October), pp. 941–54.

Ellson, R. (1980) 'Fiscal impacts on intrametropolitan residential location: further insights on the Tiebout hypothesis', *Public Finance Quarterly*, Vol. 8, No. 2 (April), pp. 189–212.

Engerman, S. (1965) 'Regional aspects of stabilization policy' in R. A. Musgrave (ed.) *Essays in Fiscal Federalism* (Washington, DC: Brookings Institution), pp. 7–62.

Epple, D., Zelenitz, A. and Visscher, M. (1978) 'A search for testable implications of the Tiebout hypothesis', *Journal of Political Economy*, Vol. 86, No. 3 (June), pp. 405–25.

Feldman, S. L. (1972) 'Waste collection services: a survey of costs and pricing', in S. J. Mushkin (ed.) *Public Prices for Public Products* (Washington, DC: Urban Institute), pp. 217–41.

Fisher, R. C. (1979) 'A theoretical view of revenue sharing grants', *National Tax Journal*, Vol 32, No. 2 (June), pp. 173–84.

*Flatters, F., Henderson, V. and Mieszkowski, P. (1974) 'Public goods, efficiency, and regional fiscal equalization', *Journal of Public Economics*, Vol. 3, No. 2 (May), pp. 99–112.

Follain, J. R. (1979) 'Grant impacts on local fiscal behaviour: full-information maximum likelihood estimates', *Public Finance Quarterly*, Vol. 7, No. 4 (October), pp. 479–500.

Foster, C. D., Jackman, R. A. and Perlman, M. (1980) *Local Government Finance in a Unitary State* (London: Allen & Unwin).

Friedman, M. (1962) *Capitalism and Freedom*, Phoenix Books (Chicago: University of Chicago Press).

Gillespie, W. I. (1965) 'Effect of public expenditures on the distribution of income', in R. A. Musgrave (ed.) *Essays in Fiscal Federalism* (Washington, DC: Brookings Institution), pp. 122–86.

*Goetz, C. J. and McKnew, C. R. (1972) 'Paradoxical results in a public choice model of alternative government grant forms', in J. M. Buchanan and R. D. Tollison (eds) *The Theory of Public Choice: Political Applications of Economics* (Michigan, Ann Arbor: University of Michigan Press), pp. 224–35.

Gold, S. D. (1979) *Property Tax Relief* (Lexington, Mass.: Lexington Books).

Gramlich, E. M. (1977) 'Intergovernmental grants: a review of the empirical literature', in W. E. Oates (ed.) *The Political Economy of Fiscal Federalism* (Lexington, Mass.: Lexington Books), pp. 219–39.

Gray, H. (1968) *The Cost of Council Housing,* Research Monograph 18 (London: Institute of Economic Affairs).

Grewal, B. S. (1981a) 'Economic criteria for the assignment of functions in a federal system', in *Towards Adaptive Federalism*, Advisory Council for Intergovernmental Relations Information Paper 9 (Canberra: Australian Government Publishing Service), pp. 1–57.

Grewal, B. S. (1981b) *Equalisation Techniques for School Finance: An Evaluation,* CRFFR Occasional Paper 20 (Canberra: Australian National University).

Grewal, B. S., Brennan, H. G. and Mathews, R. L. (eds) (1980) *The Economics of Federalism* (Canberra: Australian National University Press).

Gewal, B. S. and Mathews, R. L. (1977) 'Intergovernmental tax competition and co-ordination', in R. L. Mathews (ed.) *State and Local Taxation* (Canberra: Australian National University Press and CRFFR), pp. 83–93.

HMSO (1964) *Road Pricing: The Economic and Technical Possibilities*, Report of a Panel set up by the Ministry of Transport (Chairman R. J. Smeed) (London).

HMSO (1969a) *Report*, Royal Commission on Local Government in England 1966–69 (Chairman Lord Redcliffe-Maud), Volume I, Cmnd 4040 (London).

HMSO (1969b) *Report*, Royal Commission on Local Government in Scotland 1966–69 (Chairman Lord Wheatley), Cmnd 4150 (London).

HMSO (1971a) *Local Government in England: Government Proposals for Reorganisation*, Cmnd 4584 (London).
HMSO (1971b) *The Future Shape of Local Government Finance*, Cmnd 4741 (London).
HMSO (1973a) *Report*, Royal Commission on the Constitution 1969–73 (Chairman Lord Kilbrandon), Volume I, Cmnd 5460 (London).
HMSO (1973b) *Memorandum of Dissent by Lord Crowther-Hunt and Professor A. T. Peacock*, Royal Commission on the Constitution 1969–73 (Chairman Lord Kilbrandon), Volume II, Cmnd 5460–I (London).
HMSO (1975) *Our Changing Democracy: Devolution to Scotland and Wales*, Cmnd 6348 (London).
HMSO (1976) *Local Government Finance*, Report of the Committee of Enquiry (Chairman F. Layfield), Cmnd 6453 (London).
HMSO (1981a) *Local Government Financial Statistics: England and Wales 1979–80* (London).
HMSO (1981b) *Alternatives to Domestic Rates*, Cmnd 8449 (London).
HMSO (1982a) *Local Authority Vital Statistics 1980*, Office of Population Censuses and Surveys (London).
HMSO (1982b) *National Income and Expenditure: 1982 Edition*, Central Statistical Office (London).
HMSO (1983) *Local Government Financial Statistics: England and Wales 1981–82* (London).
Hamilton, B. W. (1976) 'Capitalization of intrajurisdictional differences in local tax prices', *American Economic Review*, Vol. 66, No. 5 (December), pp. 743–53.
Hamovitch, W. (1966) 'Sales taxation: an analysis of the effects of rate increases in two contrasting cases'. *National Tax Journal*, Vol. 19, No. 4 (December), pp. 411–20.
Hand, J. H. and Mitchell, W. E. (1971) 'Optimality in local debt limitation: comment', *National Tax Journal*, Vol. 24, No. 1 (March), pp. 101–6.
Hansen, A. H. and Perloff, H. S. (1944) *State and Local Finance in the National Economy* (New York: W. W. Norton).
Harris, R. and Seldon, A. (1976) *Pricing or Taxing?* Hobart Paper 71 (London: Institute of Economic Affairs).
Hayes, J. A. (1977) 'Provincial and local taxation in Canada', in R. L. Mathews (ed.) *State and Local Taxation* (Canberra: Australian National University Press and CRFFR), pp. 145–74.
Head, J. G. (1974) *Public Goods and Public Welfare* (Durham, North Carolina: Duke University Press).
Hepworth, N. P. (1980) *The Finance of Local Government*, 6th edn (London: Allen & Unwin).
Herber, B. P. (1975) *Fiscal Federalism in the United States of America*, CRFFR Research Monograph 10 (Canberra: Australian National University).
Hicks, U. K. (1968) 'Autonomous revenue for local government', *Western Economic Journal*, Vol. 6, No. 3 (June), pp. 177–94.
Hicks, U. K. (1978) *Federalism: Failure and Success: A Comparative Study* (London: Macmillan).
Hildersley, S. H. H. and Nottage, R. (1968) *Sources of Local Revenue* (London: Royal Institute of Public Administration).
Hochman, H. M. and Rodgers, J. D. (1969) 'Pareto optimal redistribution', *American Economic Review*, Vol. 59, No. 4 Part 1 (September), pp. 542–57.
Hogan, T. D. and Shelton, R. B. (1973) 'Interstate tax exportation and states' fiscal structures', *National Tax Journal*, Vol. 26, No. 4 (December), pp. 553–64.
Hughes, G. and McCormick, B. (1981) *Do Council Housing Policies Reduce*

Migration Between Regions? Department of Economics, Discussion Paper 8104 (Southampton: University of Southampton).
Hunter, J. S. H. (1973) *Revenue Sharing in the Federal Republic of Germany,* CRFFR Research Monograph 2 (Canberra: Australian National University).

Inman, R. P. (1977) 'Macro–fiscal planning in the regional economy: a general equilibrium approach', *Journal of Public Economics,* Vol. 7, No. 2 (April), pp. 237–60.

Jackman, R. A. (1981) 'The block grant in England', in *Public Expenditure Policy and the Rate Support Grant,* Centre for Urban and Regional Research Discussion Paper 2 (Glasgow: University of Glasgow), pp. 25–36.
Jackman, R. A. (1982) 'Does central government need to control the total of local government spending?' *Local Government Studies,* Vol. 8, No. 3 (May/June), pp. 75–90.
Jackman, R. A. and Klappholz, K. (1975) *Taming the Tiger,* Hobart Paper 63 (London: Institute of Economic Affairs).
James, L. J. (1973) 'The stimulation and substitution effects of grants-in-aid: a general equilibrium analysis', *National Tax Journal,* Vol. 26, No. 2 (June), pp. 251–65.
Jones, G. W. and Stewart, J. D. (1982) 'Policy-making in central and local government compared', *Local Government Studies,* Vol. 8, No. 1 (January/February), pp. 73–9.
Jurion, B. J. (1979) 'Matching grants and unconditional grants: the case with *n* goods', *Public Finance,* Vol. 34, No. 2, pp. 234–44.

Kahn, C. H. (1961) 'Comments', in *Public Finances: Needs, Sources and Utilization,* National Bureau of Economic Research (Princeton, NJ: Princeton University Press), pp. 436–7.
Katzman, M. T. (1972) 'Pricing primary and secondary education', in S. J. Mushkin (ed.) *Public Prices for Public Products* (Washington, DC: Urban Institute), pp. 371–93.
King, D. N. (1971) 'The distribution of taxable resources among regions and local authorities', D. Phil. thesis (York: University of York).
King, D. N. (1973) *Financial and Economic Aspects of Regionalism and Separatism,* Commission on the Constitution Research Papers 10 (London: HMSO).
King, D. N. (1976) *The Fiscal Implications of Devolution,* Lecture Series 5 (London: Institute for Fiscal Studies).
King, D. N. (1980a) *Models of Devolution,* Discussion Papers in Economics, Finance and Investment 83 (Stirling: University of Stirling).
King, D. N. (1980b) *The Theory of Equalisation Grants,* Working Paper 8 (London: Institute for Fiscal Studies).
King, D. N. (1981) *Stabilization and Subcentral Government,* Discussion Papers in Economics, Finance and Investment 89 (Stirling: University of Stirling).
King, D. N. (1982) 'A defence of equalising grants to local authorities', *Scottish Journal of Political Economy,* Vol. 29, No. 1 (February), pp. 102–10.

Ladd, H.F. (1978) 'An economic evaluation of state limitations on local taxing and spending powers', *National Tax Journal,* Vol. 31, No. 1 (March), pp. 1–18.
Lane, S. R. (1974) 'Direct taxes in relation to the division of fiscal powers', in R. L. Mathews (ed.) *Intergovernmental Relations in Australia* (Sydney: Angus & Robertson), pp. 132–58.
Lawless, P. (1980) 'New approaches to local authority economic intervention', *Local Government Studies,* Vol. 6, No. 1 (January/February), pp. 17–32.

Le Grand, J. (1975) 'Fiscal equity and central government grants to local authorities', *Economic Journal*, Vol. 85, No. 339 (September), pp. 531–47.

Macdonald, H. I. (1982) 'Federal–provincial fiscal issues in Canada' in R. L. Mathews (ed.) *Public Policies in Two Federal Countries: Canada and Australia,* CRFFR (Canberra: Australian National University), pp. 145–52.

McGuire, M. C. (1973) 'Notes on grants-in-aid and economic interactions among governments', *Canadian Journal of Economics*, Vol. 6, No. 2 (May), pp. 207–21.

McLure, C. E. (1967) 'The interstate exporting of state and local taxes: estimates for 1962', *National Tax Journal*, Vol. 20, No. 1 (March), pp. 49–77.

McLure, C. E. (1969) 'The inter-regional incidence of general regional taxes', *Public Finance*, Vol. 24, No. 3, pp. 457–83.

McMillan, M. L., Wilson, W. R. and Arthur, L.M. (1981)'The publicness of local public goods: evidence from Ontario municipalities' *Canadian Journal of Economics*, Vol. 14, No. 4 (November), pp. 596–608.

Manvel, A. D. (1971) 'Differences in fiscal capacity and effort: their significance for a federal revenue-sharing system', *National Tax Journal*, Vol. 24, No. 2 (June), pp. 193–204.

*Mathews, R. L. (1977) 'Mechanisms for fiscal equalisation in an integrating economic community', in *Report of the Study Group on the Role of Public Finance in European Integration*, Volume II (Brussels: Commission of the European Communities), pp. 401–32.

Mathews, R. L. (1980a) *Revenue Sharing in Federal Systems*, CRFFR Research Monograph 31 (Canberra: Australian National University).

Mathews, R. L. (1980b) *Australian Federalism 1979*, CRFFR (Canberra: Australian National University).

Mathews, R. L. (1982a) 'Intergovernmental financial arrangements and taxation', in R. L. Mathews (ed.) *Public Policies in Two Federal Countries: Canada and Australia,* CRFFR (Canberra: Australian National University), pp. 153–81.

Mathews, R. L. (1982b) *Federalism in Retreat: The Abandonment of Tax Sharing and Fiscal Equalisation*, CRFFR Reprint Series 50 (Canberra: Australian National University).

Maxwell, J. A. and Aronson, J. R. (1977) *Financing State and Local Governments*, 3rd edn (Washington, DC: Brookings Institution).

Maynard, A. K. (1975) *Experiment with Choice in Education*, Hobart Paper 64 (London: Institute of Economic Affairs).

Maynard, A. K. and King, D. N. (1972) *Rates or Prices?* Hobart Paper 54 (London: Institute of Economic Affairs).

Miller, F. C. (1979) *The Feasibility of Regionally Differentiated Fiscal Policies*, Department of Economics Discussion Paper 1979–6 (Ontario: University of Guelph).

Ministry of Supply and Services (1981) *Canada Year Book 1980–81* (Ottawa).

Mueller, D. C. (1979) *Public Choice* (Cambridge: Cambridge University Press).

Musgrave, R. A. (1959) *The Theory of Public Finance* (New York: McGraw-Hill).

*Musgrave, R. A. (1961a) 'Approaches to a fiscal theory of political federalism', in *Public Finances: Needs, Sources and Utilization*, National Bureau of Economic Research (Princeton, NJ: Princeton University Press), pp. 97–122.

Musgrave, R. A. (1961b) 'Reply' in *Public Finances: Needs, Sources and Utilization,* National Bureau of Economic Research (Princeton, NJ: Princeton University Press), pp. 132–3.

Musgrave, R. A. (1970) 'Pareto optimal redistribution: comment', *American Economic Review*, Vol. 60, No. 5 (December), pp. 991–3.

Musgrave, R. A. and Musgrave, P. B. (1976) *Public Finance in Theory and Practice,* 2nd edn (Tokyo: McGraw-Hill).

Mushkin, S.J. (ed.) (1972) *Public Prices for Public Products* (Washington, DC: Urban Institute).

Nathan, R. P., Manvel, A. D., Calkins, S. E. and associates (1975) *Monitoring Revenue Sharing* (Washington, DC: Brookings Institution).
Niskanen, W. A. (1968) 'Nonmarket decision making: the peculiar economics of bureaucracy', *American Economic Review*, Vol. 58, No. 2 (May), pp. 293–305.
Niskanen, W. A. (1973) *Bureaucracy: Servant or Master?* Hobart Paperback 5 (London: Institute of Economic Affairs).

OECD (1983a) *Revenue Statistics of O.E.C.D. Member Countries 1965–82* (Paris).
OECD (1983b) *Taxes on Immovable Property* (Paris).
Oates, W. E. (1972) *Fiscal Federalism* (New York: Harcourt Brace Jovanovich).
Oates, W. E. (1975) ' "Automatic" increases in tax revenues – the effect on the size of the public budget', in W. E. Oates (ed.) *Financing the New Federalism* (Baltimore, Maryland: John Hopkins University Press), pp. 139–60.
Oates, W. E. (1977) 'Fiscal federalism in theory and practice: applications to the European Community', in *Report of the Study Group on the Role of Public Finance in European Integration*, Volume II (Brussels: Commission of the European Communities), pp. 279–320.
Oates, W. E. (1979a) *Intergovernmental Grants and Revenue Sharing: Theory and Evidence*, Public Sector Economic Research Centre Discussion Paper 79–07 (Leicester: University of Leicester).
Oates, W. E. (1979b) 'Lump–sum intergovernmental grants have price effects', in P. Mieszkowski and W. H. Oakland (eds) *Fiscal Federalism and Grants-in-Aid*, Coupe Papers on Public Economics (Washington, DC: Urban Institute), pp. 23–30.
O'Higgins, M. and Ruggles, P. (1981) 'The distribution of public expenditures and taxes among households in the United Kingdom', *Review of Income and Wealth*, Series 27, No. 3 (September), pp. 298–326.
*Olson, M. (1969) 'The principle of "fiscal equivalence": the division of responsibilities among different levels of government', *American Economic Review* (Proceedings), Vol. 49 (May), pp, 479–87.

Pack, H. and Pack, J. R. (1978) 'Metropolitan fragmentation and local public expenditures', *National Tax Journal*, Vol. 31, No. 4 (December), pp. 349–62.
*Pauly, M. V. (1970) 'Optimality, "public" goods and local governments: a general theoretical analysis', *Journal of Political Economy*, Vol. 78, No. 3 (May/June), pp. 572–85.
Pauly, M. V. (1973) 'Income redistribution as a local public good', *Journal of Public Economics*, Vol. 2, No. 1 (February), pp. 35–58.
Peacock, A. T. (1977) 'The political economy of devolution: the British case', in W. E. Oates (ed.) *The Political Economy of Fiscal Federalism* (Lexington, Mass.: Lexington Books), pp. 49–63.
Peacock, A. T. (1979) *The Economic Analysis of Government and Related Themes* (London: Martin Robertson).
Peacock, A. T. and Wiseman, J. (1964) *Education for Democrats*, Hobart Paper 25 (London: Institute of Economic Affairs).
Pestieau, P. (1977) 'The optimality limits of the Tiebout model' in W. E. Oates (ed.) *The Political Economy of Fiscal Federalism* (Lexington, Mass.: Lexington Books), pp. 173–86.
Peston. M. (1972) *Public Goods and the Public Sector* (London: Macmillan).
Pigou, A. C. (1932) *The Economics of Welfare*, 4th edn (London: Macmillan).

Pogue, T. F. (1970) 'The effect of debt limits: some new evidence', *National Tax Journal*, Vol. 23, No. 1 (March), pp. 36–49.

Pommerehne, W. W. (1978) 'Institutional approaches to public expenditure: empirical evidence from Swiss municipalities' *Journal of Public Economics*, Vol. 9, No. 2 (April), pp. 255–80.

Prest, A. R. (1982) 'On charging for local government services', *Three Banks Review*, No. 133 (March), pp. 3–23.

Prest, W. (1977) 'State and local taxes in Australia: a general perspective', in R. L. Mathews (ed.) *State and Local Taxation* (Canberra: Australian National University Press and CRFFR), pp. 245–58.

Puryear, D. L., Ross, J. and Shapiro, P. (eds) (1979) 'Proceedings of a conference on tax and expenditure limitations', *National Tax Journal*, Vol. 32, No. 2 Supplement (June).

Rafuse, R. W. (1965) 'Cyclical behavior of state–local finances', in R. A. Musgrave (ed.) *Essays in Fiscal Federalism* (Washington, DC: Brookings Institution), pp. 63–121.

Reagan, M. D. and Sanzone, J. G. (1981) *The New Federalism*, 2nd edn (New York: Oxford University Press).

Reischauer, R. and Hartman, R. W. (1973) *Reforming School Finance* (Washington DC: Brookings Institution).

Robinson, R. (1979) *Housing Economics and Public Policy* (London: Macmillan).

Romer, T. and Rosenthal, H. (1980) 'An institutional theory of the effect of intergovernmental grants', *National Tax Journal*, Vol. 33, No. 4 (December), pp. 451–8.

Roth, G. J. (1965) *A Self-Financing Road System*, Research Monograph 3 (London: Institute of Economic Affairs).

Royal Institute of Public Administration Study Group (1956) *New Sources of Local Revenue* (London: Allen & Unwin).

Samuelson, P. A. (1954) 'The pure theory of public expenditure', *Review of Economics and Statistics*, Vol. 36, No. 4 (November), pp. 387–9.

Samuelson, P. A. (1958) 'Aspects of public expenditure theories', *Review of Economics and Statistics*, Vol. 40, No. 4 (November), pp. 332–8.

Sandford, C. T., Godwin, M. R., Hardwick, P. J. W. and Butterworth, M. I. (1981) *Costs and Benefits of VAT* (London: Heinemann).

Sandler, T. M. and Shelton, R. B. (1972) 'Fiscal federalism, spillovers and the export of taxes', *Kyklos*, Vol. 25, No. 4, pp. 736–53.

*Scott, A. D. (1950) 'A note on grants in federal countries', *Economica*, Vol. 17, No. 68 (November), pp. 416–22.

Scott, A. D. (1952) 'Federal grants and resource allocation', *Journal of Political Economy*, Vol. 60, No. 6 (December), pp. 534–6.

Scott, A. D. (1964) 'The economic goals of federal finance', *Public Finance*, Vol. 19, No. 3, pp. 241–88.

Seldon, A. (1977) *Charge* (London: Temple Smith).

Snyder, W. W. (1973) 'Are the budgets of state and local governments destabilizing? A six country comparison', *European Economic Review*, Vol. 4, No. 2 (June), pp. 197–213.

Spahn, P. B. (1977) 'The pattern of state and local taxation in the Federal Republic of Germany', in R. L. Mathews (ed.) *State and Local Taxation* (Canberra: Australian National University Press and CRFFR), pp. 190–241.

Sproule-Jones, M. H. (1975) *Public Choice and Federalism in Australia and Canada*, CRFFR Research Monograph 11 (Canberra: Australian National University).

Stafford, D. C. (1978) *The Economics of Housing Policy* (London: Croom Helm).
Sullivan, P. J. (1979) 'Optimality in municipal debt: a comment and respecification', *Public Finance Quarterly*, Vol. 7, No. 3 (July), pp. 352–63.

Thurow, L. C. (1966) 'The theory of grants-in-aid', *National Tax Journal*, Vol. 19, No. 4 (December), pp. 373–7.
Thurow, L. C. (1969) 'The optimum lifetime distribution of consumption expenditures', *American Economic Review*, Vol. 59, No. 3 (June), pp. 324–30.
*Tiebout, C. M. (1956) 'A pure theory of local expenditures', *Journal of Political Economy*, Vol. 64, No. 5 (October), pp. 416–24.
Tiebout, C. M. (1961) 'An economic theory of fiscal decentralization', in *Public Finances: Needs, Sources and Utilization*, National Bureau of Economic Research (Princeton, NJ: Princeton University Press), pp. 79–96.
Traber, T. (1980) 'Receipts and burden sharing, the taxation system and fiscal equalisation in the Federal Republic of Germany', in R. L. Mathews (ed.) *Federalism in Australia and the Federal Republic of Germany: A Comparative Study* (Canberra: Australian National University Press and CRFFR), pp. 261–80.
*Tulloch, G. (1969) 'Federalism: problems of scale', *Public Choice*, Vol. 6 (Spring), pp. 19–29.

United States Department of Commerce, Bureau of the Census (1981) *Statistical Abstract of the United States: 1981* (Washington).

Vickrey, W. W. (1963) 'General and specific financing of urban services', in H. G. Schaller (ed.) *Public Expenditure Decisions in the Urban Community* (Washington, DC: Resources for the Future), pp. 62–90.

*Wagner, R. E. (1970) 'Optimality in local debt limitation', *National Tax Journal*, Vol. 23, No. 3 (September), pp. 297–305.
Wagner, R. E. (1971) 'Optimality in local debt limitation: reply', *National Tax Journal*, Vol. 24, No. 1 (March), pp. 109–11.
Waldauer, C. (1973) 'Grant structures and their effects on aided government expenditures: an indifference curve analysis', *Public Finance*, Vol. 28, No. 2, pp. 212–26.
Walsh, C. (1974) 'Economic aspects of multi-level government and the distribution of fiscal responsibilities' in R. L. Mathews (ed.) *Responsibility Sharing in a Federal System*, CRFFR Research Monograph 8 (Canberra: Australian National University), pp. 1–40.
Watt, P. A. (1982) 'The control of local authority capital expenditure', *Local Government Studies*, Vol. 8, No. 3 (May/June), pp. 91–5.
Webb, M. G. (1976) *Pricing Policies for Public Enterprises* (London: Macmillan),
Wheare, K. C. (1946) *Federal Government*, 1st edn (Oxford: Oxford University Press).
Wilde, J. A. (1968) 'The expenditure effects of grant-in-aid programs', *National Tax Journal*, Vol. 21, No. 3 (September), pp. 340–8.
Wilde, J. A. (1971) 'Grant-in-aid: the analytics of design and response', *National Tax Journal*, Vol. 24, No. 2 (June), pp. 143–55.
*Williams, A. H. (1966) 'The optimal provision of public goods in a system of local government', *Journal of Political Economy*, Vol. 74, No. 1 (February), pp. 18–33.
Wiltshire, K. (1980) 'Australian state participation in federal decisions', in R. L. Mathews (ed.) *Federalism in Australia and the Federal Republic of Germany: A Comparative Study* (Canberra: Australian National University Press and CRFFR), pp. 65–135.

Wiseman, J. (1981) 'Revenue sharing between central and local governments in a developing country with a unitary system of government', in Halil Nadaroglu (ed.) *Vergi Reformlari [Tax Reforms]* (Istanbul: Istanbul Iktisadive ve Ticari Ilimler Akademisi), pp. 367–77.

Wohl, M. (1972) 'Congestion toll pricing for public transport facilities', in S. J. Mushkin (ed.) *Public Prices for Public Products* (Washington DC: Urban Institute), pp. 234–66.

Yonehara, J. (1981) *Local Public Finance in Japan*, CRFFR Research Monograph 36 (Canberra: Australian National University).

Zilderman, A. (1980) 'The economics of educational policy', in R. M. Grant and G. K. Shaw (eds) *Current Issues in Economic Policy*, 2nd edn (Oxford: Philip Allan), pp. 143–63.

Name Index

Aaron, H. J. 228, 234, 242, 303
Akin, J. S. 175
Allan, C. M. R. 9
Aronson, J. R. 30, 44, 198, 211, 242–3, 282, 298
Arthur, L. M. 20, 55
Atkinson, A. B. 48, 73–4, 85
Auld, D. A. L. 197

Barnett, R. 146, 163
Barr, J. L. 16
Bennett, J. T. 68
Bennett, R. J. 182, 194, 198
Bergstrom, T. C. 19–20
Bewley, T. F. 48
Bieri, S. 198
Bird, R. M. 255, 256, 262–9
Blaug, M. 260, 261
Boadway, R. W. 49, 89, 163, 197, 242, 269
Borcherding, T. E. 19, 68
Bradford, D. F. 120
Brainard, W. C. 163
Branson, W. H. 303
Brazer, H. E. 26
Break, G. F. 45, 113, 120, 135, 163, 191, 198
Brennan, H. G. 26, 143, 217, 242
Breton, A. 59, 62, 71, 85, 103, 107, 120, 284
Brown, C. V. 243
Browning, E. K. 49, 243
Browning, J. M. 49, 243
Brownlee, O. H. 269
Buchanan, J. M. 26, 48, 73, 143, 147–51, 155, 157–8, 163, 267
Burns, R. M. 136, 197
Butterworth, M. I. 242

Calkins, S. E. 176, 197–8
Castles, A. C. 48
Cebula, R. J. 30
Coase, R. H. 85, 126
Collins, R. A. 285
Courant, P. N. 120
Cripps, F. 197
Crispin, A. 198
Dafflon, B. 33
Daly, G. C. 274–7
Davis, A. 190
Davis, O. A. 16, 163
Deacon, R. T. 19
Dernburg, T. F. 303–4
Due, J. F. 49

Edel, M. 48
Eden, L. B. 197
Ellson, R. 30
Engerman, S. 49
Epple, D. 48

Feldman, S. L. 266
Fisher, R. C. 179, 181
Flatters, F. 48, 197
Follain, J. R. 92
Foster, C. D. 3, 5, 20, 22, 35, 48–9, 54, 77, 160, 176, 208, 220, 242–3, 250, 256, 259, 264–7, 292, 300, 303–4
Friedman, M. 263

Gillespie, W. I. 49, 303
Godley, W. 197
Godwin, M. R. 242
Goetz, C. J. 48, 120
Gold, S. D. 26, 49
Goodman, R. P. 19–20
Gramlich, E. M. 102, 113–14, 120
Gray, H. 263
Grewal, B. S. 33–4, 36, 38, 48–9, 85, 87, 120, 131, 143, 170, 230

Haines, G. H. 16
Hamilton, B. W. 147
Hamovitch, W. 241
Hand, J. H. 284
Hansen, A. H. 293
Hardwick, P. J. W. 242
Harris, C. P. 85
Harris, R. 256, 269
Hartman, R. W. 197
Hayes, J. A. 160, 243
Head, J. G. 49
Henderson, V. 48
Hepworth, N. P. 304
Herber, B. P. 44, 135
Hicks, U. K. 3, 242
Hildersley, S. H. H. 242
Hochman, H. M. 49
Hogan, T. D. 225
Hughes, G. 30
Hunter, J. S. H. 197, 242, 304

Inman, R. P. 120

Jackman, R. A. 3, 5, 20, 22, 35, 48–9, 54, 77, 85, 120, 160, 176, 195, 198, 208, 220, 242–3, 250, 256, 259, 264–7, 292, 300, 303–4
Jackson, P. M. 243
James, L. J. 120
Jay, W. R. C. 304
Johnson, M. H. 68
Jones, G. W. 22
Jurion, B. J. 120

Kafoglis, M. Z. 163
Kahn, C. H. 266
Katzman, M. T. 261
King, D. N. 49, 85, 163, 171, 178, 180, 197, 210, 211, 242, 286, 288, 299, 303
Klappholz, K. 304

Ladd, H. F. 48
Lane, S. R. 209
Lawless, P. 49
Layfield, F. 132–3, 213, 216, 229, 234–5, 242–3, 297, 300
Le Grand, J. 171
Litvack, J. M. 303
Lucke, R. 190

McCormick, B. 30
Macdonald, H. I. 136
McDougall, D. M. 303–4
McGuire, M. C. 303
McGuire, T. 120, 217
McKnew, C. R. 120
McLure, C. E. 24, 225, 227, 229, 243
McMillan, M. L. 20, 55
Manvel, A. D. 175–6, 197–8
Mathews, R. L. 5, 49, 86, 88, 129, 131, 136, 139, 143, 163, 182, 185, 188, 189, 197–8, 230, 299
Maxwell, J. A. 44, 198, 211, 242–3, 298
Maynard, A. K. 210, 242, 269
Mieszkowski, P. 48
Miller, F. C. 49
Mitchell, W. E. 284
Mueller, D. C. 15, 29, 48
Musgrave, P. B. 9, 15, 33–4, 36, 73, 85, 88, 127, 129, 131–2, 163, 242, 243, 250, 269, 273–4
Musgrave, R. A. 9, 15, 33–4, 36, 49, 73, 85, 88, 127, 129, 131–2, 147–51, 163, 168, 170, 197, 242, 243, 250, 269, 273–4
Mushkin, S. J. 269

Nathan, R. P. 176, 197–8
Niskanen, W. A. 103–7, 120, 224
Nottage, R. 242

O'Higgins, M. 243
Oates, W. E. 3, 21, 23, 30–1, 33, 48, 49, 72, 85, 89, 110–13, 120, 126, 133, 135, 143, 155, 162–3, 170, 208, 215, 224, 241, 242–3, 275, 284
Olsen, M. 85

Pack, H. 30
Pack, J. R. 30
Pauly, M. V. 35, 49, 163
Peacock, A. T. 22, 26, 107, 255, 260–1, 269
Perlman, M. 3, 5, 20, 22, 35, 48–9, 54, 77, 85, 160, 176, 208, 220, 242, 243, 250, 256, 259, 264–7, 292, 300, 303–4
Perloff, H. S. 293
Pestieau, P. 48
Peston, M. 269
Pigou, A. C. 126
Pogue, T. F. 277
Pommerehne, W. W. 16
Prest, A. R. 256, 269
Prest, W. 243
Puryear, D. L. 48

Rafuse, R. W. 293
Reagan, M. D. 3
Reischauer, R. 197
Robinson, R. 254, 269
Rodgers, J. D. 49
Romer, T. 89, 107–10, 224
Rosenthal, H. 89, 107–10, 224
Ross, J. 48
Roth, G. J. 262–3, 269
Rubinfeld, D. C. 120
Ruggles, P. 243

Samuelson, P. A. 15, 17, 29, 48
Sandford, C. T. 242
Sandler, T. M. 243
Sanzone, J. G. 3
Sclar, E. 48
Scott, A. D. 59, 62, 71, 85, 143, 156–8
Seldon, A. 255–6, 260–1, 263, 265, 269
Shapiro, P. 48
Shelton, R. B. 225, 243
Smeed, R. J. 262
Snyder, W. W. 293
Spahn, P. B. 243
Sproule-Jones, M. H. 23, 48
Stafford, D. C. 254, 269
Stewart, J. D. 22
Stiglitz, J. E. 48, 73–4, 85
Sullivan, P. J. 285

Thurow, L. C. 120, 284
Tiebout, C. M. 28–30, 48, 65, 85
Topham, N. 146, 163
Traber, T. 186, 197
Trenery Dolbear, F. 163
Tulloch, G. 66, 85

Vickrey, W. W. 269
Visscher, M. 48

Wagner, R. E. 48, 158, 277, 281–4
Waldauer, C. 120
Walsh, C. 42
Watt, P. A. 304
Webb, M. G. 269
Wheare, K. C. 2–3
Wilde, J. A. 89, 92, 120
Williams, A. H. 163
Wilson, W. R. 20, 55
Wiltshire, K. 299, 304
Winston, A. 163
Wintrobe, R. 103, 107, 120
Wiseman, J. 131, 255, 260–1, 269
Wohl, M. 269

Yonehara, J. 237

Zelenitz, A. 48
Ziderman, A. 260–1

Subject Index

agency functions of subcentral government 4–5, 7, 36, 77, 133–4
alcohol tax:
  elasticity of yield 213
  excess burden 218, 220
  geographical distribution of yield 212
  incidence 229
  perceptibility 230
  spread of incidence 231
  tax exporting 24, 229
  see also taxes on expenditure
Australia:
  Commonwealth Grants Commission 188
  elasticity of tax yields 214
  employment-creating policies 45
  equalisation 136, 182, 188
  loan controls for local authorities 299
  Loan Council 45, 298
  local authority populations 82
  need for equalisation grants 182–3
  need for revenue-sharing grants 184–6
  Premiers' Conferences 45
  revenue-sharing 136, 182, 188
  special educational projects grants 131
  state populations 80–1
  subcentral taxes 4, 236–7
  use of specific grants 136
Austria, subcentral taxes 236–7

balance of payments 288, 294–6
Belgium, de-stabilising effect of local budgets 293
benefit taxation 14, 28, 32, 147–54, 267
block grants 87, 88, 120
  see also general grants (lump-sum) *and* specific grants (lump-sum) borrowing by subcentral authorities:
  controls over use 270, 276–7, 283, 297–9
  inter-generational equity aspects 272–7, 283
  intra-generational equity aspects 276–83
  possibility of excess debt 25, 277, 281–3
boundaries of subcentral authorities:
  at local level 81–2
  at state level 80–1
  geographical factors 79
  influence of history 79
  see also optimal size of subcentral authorities
buoyancy of taxes,
  see taxes (elasticity of yields)
bureaucracy:
  and case for several tiers of government 22–3
  and consumer choice 255
  and excessive budgets 16, 25, 102–7, 215
  and production costs 68, 105–7
  and tax competition 40
bureaucratic rule of two 68

Canada:
  elasticity of tax yields 214
  equalisation 136, 160, 182, 186–8
  Established Programs Financing 136
  need for equalisation grants 183
  need for revenue-sharing grants 184–6
  non-renewable resource revenues 187
  provincial jurisdiction 43
  provincial populations 80–1
  publicness of fire service 20
  regional development 45
  revenue-sharing 136, 182, 186–8
  subcentral taxes 236–7
  use of specific grants 136
capitalisation 144, 146–7, 203, 205, 242, 274–5, 281–3
cash transfer payments 36–7, 250–8
categorical grants 88, 134
  see also specific grants
charges:
  and revenue effort 179
  arguments for increased use 258–9
  as alternative to revenue-sharing 138, 259
  case for experiments 267
  for art galleries 266
  for education 259
  for five services 265
  for housing 263–4
  for merit goods 250–8
  for museums 266
  for parks 266
  for police services 264–5
  for refuse services 265
  for roads 261–3
  for services to businesses 225
  for water supply 266
  for welfare services 266
China, grants paid by provinces to central government 86
closed-ended grants,
  see specific grants (closed matching)
clubs, economic theory of 73
co-operative federalism 7, 23, 45, 132
conditional grants 87
  see also specific grants
control of subcentral authorities 139, 184, 224, 297–302

see also grants to enforce grantor preferences *and* borrowing by subcentral authorities (controls over use)
corporation tax:
allocation of yield 211
effect on household location 223
elasticity of yield 213
excess burden 218, 219
geographical distribution of yield 212
incidence 227
macroeconomic implications 232–3
perceptibility 230
spread of incidence 231
tax exporting 24, 227
use in tax competitions 230
cost index,
see expenditure requirements (unit cost index)

defence 10, 11, 17, 18
deflective effect 92, 96
demand management 288–93
see also stabilisation
demerit goods 9, 76–7, 249
Denmark:
loan controls for local authorities 297
subcentral taxes 236–7
discretionary purchasing by subcentral authorities 27, 39, 288
distribution policy 9, 32–7, 44, 76–7, 133, 207, 249–58, 287
see also merit goods
district power equalisation 168
see also equalisation grants (scheme to equalise fiscal potential)

earmarked revenues 267
education 8, 9, 26, 131, 259–61
effort-related grants,
see general grants (effort-related) *and* equalisation grants (proposed new scheme *and* scheme to equalise fiscal potential)
equalisation grants:
administration costs 27, 79
as alternative to benefit taxation 147–54
as alternative to capitalisation 146–7
as alternative to geographical variations in central tax rates 155
difficulties with very uneven bases 211–12
effect of ignoring variations in unit cost indices 145, 166, 168, 174
effect of ignoring variations in unit need indices 166, 168
effect on effort-related schemes of having few authorities 180–1
effects on development 159–62
effects on locational efficiency 155–8
efficiency aspects 32, 144, 170
equity aspects 62, 79, 140–3, 170, 175
need in practice 182–3
negative grant entitlements 165, 166, 170–1, 173, 211
proposed new scheme 171–4, 182, 194
scheme to equalise fiscal capacity 168–71, 182, 186–9
scheme to equalise fiscal potential 164–8, 170–1, 182, 188, 287
to reduce problems of unharmonised tax structure 204–7
to reduce problems of varying subcentral distribution policies 34, 207
see also expenditure requirements, resource measurement, revenue effort measurement *and* service level index
equity:
inter-generational 272–7
intra-generational 277–83, 297
exchange rate 294
excise taxes,
see alcohol tax, motor fuel tax *and* tobacco tax
exclusion costs 11–13, 44, 249, 259
expenditure control techniques 27, 54, 55
expenditure requirements:
measurement 174–5
unit cost index 145, 165
unit need index 142–3, 145, 165, 197
externalities:
and fire services 265
and refuse services 265
arising with private goods 7–9, 43, 247–9
internalised by subcentral authorities 52–7
possible resolution by bargaining 126
produced by subcentral authorities 17, 18, 27, 39, 41–2, 62–4, 132–3, 225
see also grants to internalise externalities

federal countries, nature and existence of 2–4
Finland, subcentral taxes 236–7
fire services 17, 18, 19–20, 76, 265
fiscal illusion 110–13, 214
fiscal imbalance:
horizontal 137
see also equalisation grants
vertical 137–40
see also revenue-sharing grants
fiscal residuums 143, 150, 152–5
flood protection 17, 72–4, 132
fly repellent effect 118
flypaper effect:
bureaucracy model 103–7
evidence 102
fiscal illusion model 110–13
implications 121
referendum model 109–110
subcentral tax constraint model 114–19
France, stabilising effect of local budgets 293

garbage services,
  see refuse services
general grants:
  definition 87–8
  effect of constraint on grantee tax rates 114–19
  effect of grantor taxes 99
  effort-related 24, 87, 88, 100, 129, 134, 138, 166, 171, 179–81, 224, 287, 289
  lump-sum 87, 88, 91–2, 93–4, 96, 97, 100, 129
    see also flypaper effect *and* fly repellent effect
  used for equalisation 146, 159
    see also equalisation grants
  used for revenue-sharing 140
    see also revenue-sharing grants
general sales tax:
  definition of base in equalisation schemes 177
  elasticity of yield 213
  excess burden 218, 220
  geographical distribution of yield 212
  incidence 228–9
  lumpy yield 201
  perceptibility 230
  spread of incidence 231
  tax exporting 228–9
  unpredictable yield 201
    see also taxes on expenditure
grants to enforce grantor preferences:
  for macroeconomic reasons 133–4
  merit good cases 129–32
  minimum standards cases 132–3, 173
grants to internalise externalities:
  as alternative to tax exporting 226
  design 123–6
  effect on output of subcentral services 127–9, 163
  rationale 18, 27, 122–3
growth rate 288, 295–6

health services 26
housing 9, 26, 250–8, 263–4

income tax,
  see corporation tax, payroll tax *and* personal income tax
India, pilot agricultural projects grants 131
inflation 288, 290, 295–6
Italy, stabilising effect of local budgets 293

Japan, subcentral taxes 236–7

Keynesianism 289, 293–4

land tax,
  see property tax
land use controls 8
legislation by subcentral government 7, 43, 76–7, 258
libraries 132
loans to subcentral authorities
  see borrowing by subcentral authorities
local monopolies 7–8, 245–8
  see also trading activities of subcentral authorities
local public goods:
  cases with benefit areas fixed by technological factors 72–4
  economies of scale 20, 27, 52–8, 68–70
  efficiency in production 23, 26–7
  evidence on publicness 19
  nature 17–20
  optimal provision 18, 21–9, 122–4
  provision in practice 44
  subcentral provision 20–32
  technological aspects 17, 18
    see also public goods
lump-sum grants,
  see general grants (lump-sum) *and* specific grants (lump-sum)

matching grants,
  see specific grants (closed matching, conditional matching *and* open matching)
median voter principle 16, 89, 120, 231, 279–83
merit goods 9, 36–7, 44, 76–7, 249–59, 271
migration:
  of capital 78, 206, 209, 219
    see also tax competition
  of households 29–35, 38, 65–6, 73–4, 85, 132–3, 144–8, 152–4, 203–9, 220–3, 271, 274, 279–83
    see also voting-with-the-feet
  of labour 38, 78, 156–8, 206, 209, 264, 277
  of trade 78, 204, 209, 220, 277
minimum standards requirements 32
  see also grants enforce grantor preferences (minimum standards cases)
monetarism 289, 293–4
monopolies
  see local monopolies *and* trading activities of subcentral authorities
motor fuel tax:
  elasticity of yield 213
  excess burden 218, 220
  geographical distribution of yield 212
  incidence 229
  perceptibility 230
  spread of incidence 231
  tax exporting 24, 229
  use for road charging 262
    see also taxes on expenditure
motor vehicle licence duties 210
museums 132

natural resource revenues:
elasticity of yield 213
excess burden 218, 220
geographical distribution of yield 212
incidence 229
perceptibility 230
spread of incidence 231
tax exporting 229
see also taxes on expenditure
needs index,
see expenditure requirements (unit need index)
Norway, subcentral taxes 236–7

open-ended grants,
see specific grants (open matching)
optimal size of subcentral authorities:
effect of economies of scale 20, 52–7, 81, 83
effect of internalising externalities 52–7
effect of many services 70–2
effect of migration by households 65–6
effect of purchasing services 66–70
effect of service levels demaded 55–8
effect of tax-raising 77–9
effect of variations in population density 79
effect of varying incomes and tastes 58–62
factors arising with impure public goods 74–6
organisation costs:
administration 27, 62–4, 71, 79
ascertaining preferences 22, 62–4
co-ordination 62–4, 71
signalling 63, 71–2
tax collection 78

parks 15, 19
payroll tax:
allocation of yield 211
elasticity of yield 213
definition of base in equalisation schemes 177
effect on household location 223
excess burden 218, 219
geographical distribution of yield 212
incidence 227
macroeconomic implications 232–3
perceptibility 230
spread of incidence 231
tax exporting 227
use in tax competitions 230
percentage grants 88
see also specific grants (closed matching, conditional matching *and* open matching)
personal income tax:
allocation of yield 210
and distribution policy 36
definition of base in equalisation schemes 177
effect on household location 222
elasticity of yield 213
excess burden 217, 220
geographical distribution of yield 212
incidence 226
lumpy yield 201
macroeconomic implications 232–3
perceptibility 230
spread of incidence 231
tax exporting 24, 226
unpredictable yield 201
use in tax competitions 231
perversity hypothesis 293
police services 19, 68, 76, 264–5
poll tax:
allocation of yield 210
circumstances when best for poor 147
effects on household location 28, 222
elasticity of yield 213
excess burden 219, 220
geographical distribution of yield 212
incidence 229
macroeconomic implications 232–3
perceptibility 229
spread of incidence 231
tax exporting 229
use in tax competitions 231
price indices 295–6, 304
project appraisal 27, 54, 55
property tax:
allocation of yield 210
definition of base in equalisation schemes 176–7
effect on household location 49, 222–3
elasticity of yield 213
equity aspects of tax on domestic property 233–5
excess burden of tax on buildings 219, 220
excess burden of tax on land 219, 220, 242
geographical distribution of yield 212
incidence of tax on buildings 228
incidence of tax on domestic property 115, 234–5, 243
incidence of tax on land 227–8
macroeconomic implications 232–3
perceptibility 230
spread of incidence 231
tax exporting 24, 227–8, 242
tax limitation proposals 26
use in tax competitions 230
use for servicing of debt 274, 275, 277
public goods:
nature 10–11
optimal provision 14–15, 249
pure and impure 15
see also local public goods
public transport 8

purchasing of public services:
advantages 66–8
disadvantages 68–70

referenda 22, 25–6, 107–110
refuse services 8, 9, 68, 265
regional development 24, 40–3, 45–6, 76–7, 134, 288
resource measurement:
case for using modified tax bases 176–8
income base versus tax base 176
revenue deficiency 137
see also revenue-sharing grants
revenue effort measurement:
inclusion of charge revenue 179
multiple tax case 178, 197
possibility of including loan revenue 179
single tax case 178
revenue-sharing grants:
administration costs 27, 79
and efficient service levels 224
choice of grant 140
combined with equalisation grants 140, 165
definition 138
determination of aggregate grant payment 179–81, 289–90
need in practice 183–6
use to control subcentral spending 300
see also equalisation grants (for various schemes)
road congestion charges 262–3
roads 15, 17, 68, 76, 261–3

sales tax,
see general sales tax
seed-money function of grants 131
selective grants 87
see also specific grants
service level index 165–71
signalling 22, 59
see also organisation costs (signalling)
size of authorities,
see optimal size of subcentral authorities
specific grants:
closed matching 87–8, 94–7, 129
conditional matching 98
definition 87–8
effect of constraint on grantee tax rates 114–19, 129
effect of grantor taxes 99, 127–30
lump-sum 87, 90–2, 96, 97, 126, 133
narrow 88
open matching 87–8, 93–4, 96, 110, 113, 126, 129, 301
use in practice 134–7
used as means of grantor control 184
used for equalisation 131, 135–6
used for revenue-sharing 131, 135–6, 140, 184
wide 88
stabilisation 37–40, 44, 133
see also demand management
street lighting 17, 18, 76
Sweden, subcentral taxes 236–7
Switzerland:
agency functions of cantons 5
Constitution and federalism 4
evidence for median voter hypothesis 16
revenue-sharing grants 189
subcentral taxes 236–7

taste variations between areas 20–2, 35, 37, 61–2, 79, 81, 288
tax competition 24, 40, 230–1
tax exporting 24, 79, 224–9, 243, 271, 292
tax harmonisation 32
tax-sharing 138, 139, 237–40
taxes:
administration and compliance costs 215–16
elasticity of yields 212–15
equity aspects 233–5, 242
excess burdens 216–19
geographically discriminating rates 22
incidence 226–9
perceptibility 25, 229–30, 243
progressive and repressive 235, 243
spread of incidence 25, 231–2
taxes for subcentral authorities:
allocation of yields 210
case for a harmonised structure 201–9
case for devolved taxes 138–9
controls over tax rate changes 299
effect on household location 31, 220–3
excess burdens when levied at uniform rates 216–19
excess burdens when levied at varying rates 219–23
geographical distribution of yields 211–12
macroeconomic implications 232–3
offset against central tax liability 24, 34, 40, 225, 290, 291
use for capital projects 275–7
use in tax competitions 230–1
taxes on expenditure:
allocation of yield 210
effect on household location 223
macroeconomic implications 232–3
tax exporting 24
use in tax competitions 230
see also alcohol tax, general sales tax, motor fuel tax, natural resource revenues, tobacco tax *and* value added tax
Tiebout hypothesis 28
see also voting-with-the feet
tiers of subcentral authorities 71–2, 82
tobacco tax:
elasticity of yield 213
excess burden 218, 220
geographical distribution of yield 212

incidence 229
perceptibility 230
spread of incidence 231
tax exporting 24, 229
see also taxes on expenditure
trade cycle 293
trading activities of subcentral authorities:
and optimal size of authority 76–7
as alternative to regulation of local monopolies 8
average cost pricing 246
marginal cost pricing 245–8, 258, 266, 269
transfer payments,
see cash transfer payments

unconditional grants 87
see also general grants
unitary countries, nature of 2
United Kingdom:
agency functions of local authorities 4, 137
charges by local authorities 244
controls over local authorities 184
council housing 30, 46
devolution debate 22–3, 43
elasticity of tax yields 214
equalisation 182, 191–5
growing use of general grants 131
lack of buoyancy of property tax 214
loan controls for local authorities 297, 299
local authority populations 81, 82
local authority services 44
local authority spending restrained 300
local government reform 82–5
London rate equalisation scheme 86
migration of households 30
multiplier effects of subcentral authorities 49
municipal enterprises 43
need for equalisation grants 182
need for revenue-sharing grants 184
possible introduction of poll tax 210
property tax 216, 225, 234–5, 236
rate rebates 235
rate support grant 180, 191–5
regional development 46
revenue effort measurement 178
revenue-sharing 182, 191–5
stabilising effect of local budgets 293
use of specific grants 136
value added tax compliance costs 216
United States:
allotted formula categorical grants 134
block grants 134
changing nature of federalism 3
charges by subcentral authorities 244
circuit breakers 44, 235
Constitution and federalism 4, 43
distribution policy 44
elasticity of tax yields 214
equalisation 135
fall in number of local authorities 4
loan controls for subcentral authorities 298
local authority populations 81–2
measurement of resources 175
migration of households 30
migration of trade with sales tax 241
need for equalisation grants 183
need for revenue-sharing grants 184–6
perversity hypothesis 293
project categorical grants 134
Proposition 13 (California) 26
Proposition 2½ (Massachusetts) 26
publicness of local goods 19
relationship of state tax structure and expenditure growth 215
revenue-sharing 131, 135, 180, 181, 182, 189–91, 197
school provision 23
state income tax administration costs 216
state income tax structures 44
state populations 80–1
state sales tax administration costs 216
state tax-sharing 139
subcentral taxes 210, 236–7
tax competition 40, 45
tax exporting by states 24, 225, 227–9
tax limitation proposals 25–6
use of specific grants 134–5, 184
zoning laws 30

value added tax:
compliance costs 216
definition of base in equalisation schemes 177
elasticity of yield 213
excess burden 218, 220
geographical distribution of yield 212
incidence 228–9
perceptibility 230
spread of incidence 231
tax exporting 228–9
see also taxes on expenditure
voting-with-the-feet:
and allocation policies 20, 28–32, 255
and distribution policies 34
and optimal size of authority 65–6, 81
not always substitute for minimum standards 132
with unharmonised tax structures 208
vouchers 37, 250–8, 259, 260–1, 263, 266

water supply 12, 68, 70
wealth tax 210
West Germany:
Bundesrat 45
elasticity of tax yields 214
equalisation arrangements 182, 186
fall in number of local authorities 4
Land populations 80–1
loan controls for subcentral authorities 298

local authority populations 81–2
merit good provision 44
need for equalisation grants 183
need for revenue-sharing grants 184–6
regional development 45
stabilising effect of local budgets 293
subcentral taxes 236–7
tax-sharing 139, 182, 200, 236, 237–9