An Essay on
URBAN ECONOMIC THEORY

Advances in Urban and Regional Economics

VOLUME 1

Series Editor

Richard Arnott, *Boston College, U.S.A.*

An Essay on
URBAN ECONOMIC THEORY

by

YORGOS Y. PAPAGEORGIOU
McMaster University

and

DAVID PINES
Tel Aviv University

SPRINGER SCIENCE+BUSINESS MEDIA, LLC

Library of Congress Cataloging-in-Publication Data

Papageorgiou, Y. Y.
An essay on urban economic theory / by Yorgos Y. Papageorgiou and David Pines.
p. cm. — (Advances in urban and regional economics ; v. 1)
Includes bibliographical references (p.) and index.
ISBN 978-1-4613-7245-5 ISBN 978-1-4615-4947-5 (eBook)
DOI 10.1007/978-1-4615-4947-5
1. Urban economics. I. Pines, David. II. Title III. Series.
HT321.P34 1998
330.9173'2—dc21 98-44115
CIP

Originally published by Kluwer Academic Publishers in 1999
Softcover reprint of the hardcover 1st edition 1999

Printed on acid-free paper.

To Tamar and to Maria

Contents

List of Figures

List of Tables

Preface

This book launches a new series, Advances in Urban and Regional Economics. The series aims to provide an outlet for longer scholarly works dealing with topics in urban and regional economics. For over a decade, there has been no active book series by a leading academic publisher devoted to these fields. As a result, many longer studies have been unhappily adapted for journal publication. As well, the incentive has been to work on projects conducive to journal rather than book publication. This book series will hopefully encourage researchers to undertake broader and deeper projects, leading to enduring works of scholarship.

The scope of the series is intended to be broad, in terms of both approach and subject. The only criterion for selecting a book is that it makes a significant contribution to scholarship in urban and regional economics. Thus, a history of urbanization in Libya, for example, would be welcome if it made methodological innovations or provided state–of–the–art analysis; so too, subject to the same provisos, would be a study of regional transportation policy in Norway, a time series analysis of residential real estate cycles in Vancouver, a theoretical analysis of the economic effects of urbanization on kinship structures, an examination of alternative methods of local school finance in Michigan, or a review of the use of geographical information systems in studying the spatial patterns of economic growth. Conference volumes and volumes of collected papers will be considered too, based on the same criteria.

Over the past thirty years, urban economic theory has been one of the most active areas of urban and regional economic research. Just as static general equilibrium theory is at the core of modern microeconomics, so is the topic of this book—the static allocation of resources within a city and between cities—at the core of urban economic theory. The centerpiece of urban economic theory is the monocentric (city) model. It was the first spatial general equilibrium model of the city, and provided a way of conceptually integrating what had previously been disparate strands of literature on urban location and land use, urban transportation, housing, and local public finance. The monocentric model has also provided the foundation on which spatial models of polycentric cities, system of cities, and even more generally an urbanized economy, have been built. While the monocentric model has been exhaustively examined, many of its recent extensions are in the process of becoming, and are accordingly tentative, less general, and less elegant.

This book by David Pines and Yorgos Papageorgiou, two of the most distinguished and venerable urban economic theorists, well reflects the state of the field. Part I provides an elegant, coherent, and rigorous presentation of several variants of the monocentric model, treating equilibrium, optimum, and comparative statics. Then, building on the results of Part I, Part II explores less familiar and even some uncharted territory. The monocentric model looks at a single city in isolation, taking as given a central business district surrounded by residences. Part II, in contrast, makes the intra–urban location of residential and non–residential activity the outcome of the fundamental trade-off between the propensity to interact and the aversion to crowding; the resulting pattern of agglomeration may be polycentric. Part II also develops models of an urbanized economy with trade between specialized cities and examines how the market–determined size distribution of cities differs from the optimum.

I am proud to have An Essay on Urban Economic Theory as the lead book in the series. Not only does it set a high standard for the series, in terms of both intellectual and presentational quality, but it also conveys the fascination of urban economics and the dynamism of current research in the field.

Richard Arnott
Series Editor

Acknowledgments

First we would like to acknowledge the encouragement of Richard Arnott, without whom this project could not have been accomplished—certainly not in its present form. We also would like to acknowledge helpful suggestions and comments from colleagues with whom we discussed various issues of this book. We received very helpful suggestions and comments on specific issues from Yossi Hadar, Elhanan Helpman, Oded Hochman and Efraim Sadka. We discussed specific issues with Alex Anas, Tomoya Mori and Myrna Wooders, and we have benefited considerably from these discussions. Yossi Hadar and Loay Alemi helped us with the numerical calculations. We were advised and helped in locating references by Tomoya Mori, Tamar Pines and Jacques–François Thisse. The figures were produced by Richard Hamilton. Finally, in preparing the camera–ready copy, we received extensive technical assistance from the people of TCI Software Research, especially George Pearson.

1
Introduction

1.1 Scope of this Book

One can choose almost any topic of economics (say, according to the classification of the *Journal of Economic Literature*) and add the adjective 'urban' to define a legitimate subject of urban economics in the broad sense. But the reader should know in advance that the scope of this book is much more modest. We only focus on studies pertaining to issues of location and land use which are peculiar to cities, and we discuss both positive and normative aspects of these issues. Many other urban issues, some of which are extremely important like urban finance, urban income distribution, urban poverty, urban discrimination, and urban growth, are not discussed in this book.

The literature of urban economics pertaining to issues of location and land use can be classified according to three main topics. The first, which we refer to as *urban land use and rent theory*, deals with the qualitative and quantitative aspects of urban land use and with the rôle of the centrifugal and centripetal market forces that generate them at the intra–city level. This topic of urban economics became an important focus of interest following the ambitious and comprehensive studies on transportation and land use undertaken in the 1950s and 1960s.

The second topic is mainly concerned with the *size and functional distribution of cities* as an outcome of the interplay between the centripetal and centrifugal market forces that generate them. This topic of urban economics also includes local public good theory, where the provision of a public good (rather than the agglomeration of market–oriented activities), becomes the source of the required centripetal force. In many of these studies, however, the intra–city resource allocation is completely suppressed and cities are treated as spaceless entities. At best, the urban space is reduced to a fixed amount of land which implies land scarcity and, therefore, generates the required centrifugal force. The importance of this topic has grown fast since the early 1990s. In the most recent studies it is combined with the first, closely related topic to create a unified field of study.

The third topic of urban economics, which is perhaps more ambitious than the first two, can be referred to as the *economic geography of cities* and deals with the function, size and distance between cities. Although the formal study of this topic preceded the other two in the context of *central place theory*, it remains the least developed one because it is the most complicated. It is only in recent years that some ambitious efforts are made to re–establish the conjectures of central place theory on solid microeconomic foundations.

We aim to provide the reader with a systematic exposition of the above three topics, and we make every effort to recognise all the breakthrough contributions we are familiar with in an appropriate manner. But since these contributions are viewed from the authors' viewpoint, the presentation cannot remain unbiased. We summarise the important findings in each chapter as 'results'. The list of results pertains not only to the surveyed studies published elsewhere through the years, but also to some new ones which represent original contributions of this book and which have not been published elsewhere.

1.2 Fundamental Determinants of Spatial Structure

The propensity of economic agents to interact with each other lies behind the very existence of cities—hence behind all three of the topics we presented in section 1.1. By interaction we mean any exchange of goods, services and ideas among agents, as well as any common action of agents to promote their interests. Since spatial interaction requires either transportation or communication or both, and since the amount of resources spent for interaction increases with distance between those who interact, saving on such costs implies the agglomeration of agents. Agglomeration, in turn, runs counter to a preference of individuals for non–congested sites. We thus arrive at a fundamental trade–off between the *propensity to interact* and the *aversion to crowding*, which determines the spatial structure of human settlement. On the one hand agglomeration increases agents' welfare because it saves on the *cost of interaction*. On the other hand it decreases agents' welfare because it reduces the space available to them.

Different combinations of the above three elements determine different spatial distributions of agents. If there is no propensity to interact, or if there is no cost of interaction, the aversion to crowding will cause an even distribution of agents over the landscape as in figure 1.1(1). If crowding is immaterial and if interaction is costly then the propensity to interact will dictate a perfectly concentrated spatial pattern as in figure 1.1(2). Finally, if all three factors act to produce spatial structure then the fundamental trade–off between costly interaction and the aversion to crowding will be expressed as in figure 1.1(3).

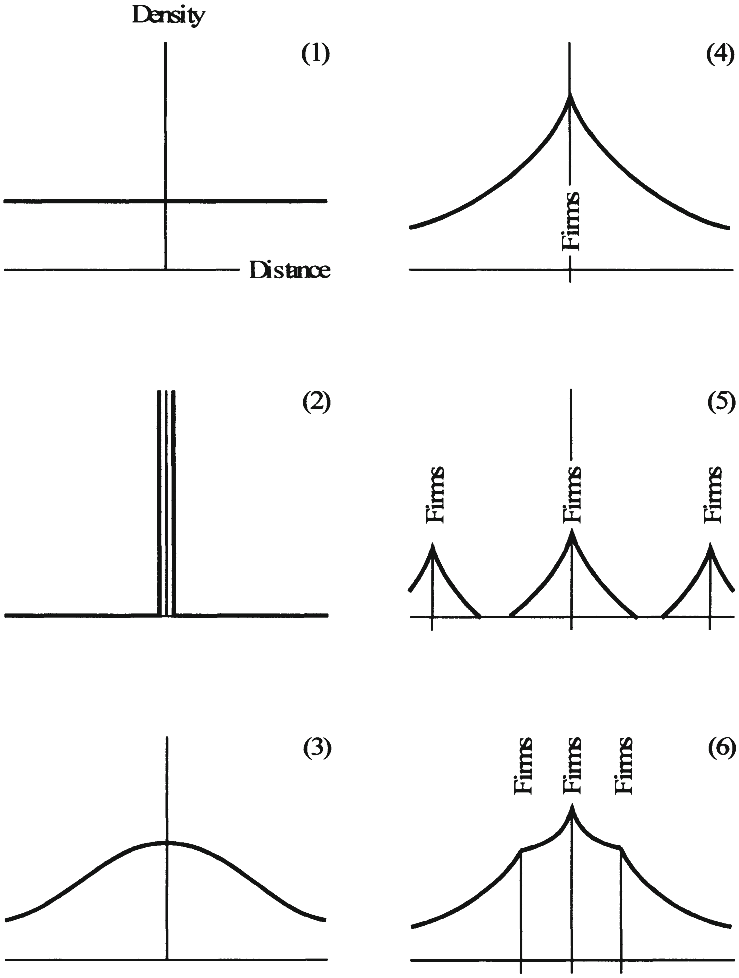

FIGURE 1.1. Alternative Spatial Arrangements.

New possibilities emerge if we distinguish between firms and individuals (or between production and consumption) as two distinct types of agent. If there is no congestion cost for firms and if interaction between firms is both beneficial and costly then firms will concentrate either on a single location to create a single *centre* as in figure 1.1(4), or on several locations to create several centres as in figure 1.1(6). The choice between a monocentric or polycentric configuration will depend on the relative importance and the relative cost of interactions to the parties involved. The first case provides the theme for the first part of our book while the second appears in the second part. Consider now the same two distinct types, but impose the additional feature that agents can be partitioned into groups, and that within–group interactions (between individuals, between individuals and firms, and between firms) are both beneficial and costly while interactions between agents belonging to distinct groups are not beneficial. Under these circumstances we obtain a number of cities as in figure 1.1.(5)—the simplest example of a city system.

Once again, we can use figure 1 to illustrate the classification of the literature presented in section 1.1. The first topic, urban land use and rent theory, is reflected in figures 1.1(3), 1.1(4) and 1.1(6). Of these, figures 1.1(3) and 1.1(4) refer to the monocentric city while figure 1.1(6) refers to the polycentric city. Figure 1.1(5) and the discussion thereof is associated with the second topic, that is, the size and functional distribution of cities. In terms of figure 1.1(5), the economic geography of cities relates to both the size and the distance between them.

1.3 Historical Sketch

1.3.1 Urban Land Use and Rent Theory

The Monocentric City

At the core of urban land use and rent theory lies the concept of land rent: the use of land goes to the agent who is prepared to offer the highest amount per unit of land. Early rent theorists, such as David Ricardo (1821) and Johann H. von Thünen (1826), wrote about agricultural land rent. On this subject, Ricardo emphasised fertility differences while von Thünen concentrated on location differences. For the latter, the sum of the market rent plus the cost of transporting the output of a given crop raised on a unit of land is fixed and equal to the price the crop can fetch in the market. This implies that, given any two plots used to raise the same crop, the rent differentials reflect the transportation cost differentials (per unit of crop per land coefficient) or, equivalently, the net productivity differentials if we consider the transportation cost as any other input required to produce the output. Under this interpretation, excepting the source of productivity differentials, the fundamental concepts of rent in Ricardo and von Thünen are identical.

Obviously, von Thünen's contributions about the effect of location on prices in general and on land rent in particular, about the effect of urban demand fluctuations on the corresponding agricultural supply area, about interactions between the city and its hinterland, and so on, exerted a stronger direct influence on spatial analysis than those of Ricardo. Indeed, the foundations of modern urban economics lie on von Thünen's theory of agricultural land use. Nevertheless, the indirect influence of Ricardo is also important—albeit negative. What Ricardo essentially did was to eliminate space by subsuming transportation costs along with other costs of production and, through the influence of his writings, to eliminate spatial analysis almost completely and for generations to come from the mainstream of English-speaking economics.[1] This is a reason why, since the

[1] The same point is made both in Ponsard (1958) and in Ekelund and Hubert (1993). Our historical sketch draws on both sources. Ricardo's reversal appears even stronger if one

time of Ricardo and until August Lösch's (1940) English translation in 1954, most developments in spatial analysis came from Germany.

During the rest of the nineteenth century, various individuals extended and sharpened the ideas of their predecessors about pricing policy of firms in a spatial economy, about their location policy, and about the corresponding optimal shape and extent of their market areas. Of particular importance are two further contributions made toward the end of the nineteenth century because they have become relevant to modern urban economics. The first one refers to Alfred Marshall (1890) who, following the tradition of the English school, kept spatial issues on the margin of his analysis. Marshall, nevertheless, devoted an entire chapter to urban land rent in which he took explicitly into account the relationship between transportation cost and distance from the market. He partitioned urban land rent, the 'site value' according to his terminology, into the 'situation value' (which arises from the specific urban location advantages of the site) and agricultural rent. The second contribution was made by Bleicher (1892), and it refers to his data collection and analysis of the population distribution of Frankfurt am Main. This can be considered as the origin of the extensive empirical work on urban population densities and land values.

We now turn to the modern urban land use and rent theory. As with von Thünen (1826), modern urban economics is based on the observation that urban location differences, characterised by differences in transportation cost, imply compensating differences in the value of urban land: for central locations, where transportation cost is low, land rent is high while for peripheral locations, where transportation cost is high, land rent is low. This idea, which was elucidated as a 'complementarity hypothesis' by Robert Haig (1926), can lead to the first clear statement of spatial equilibrium within an urban context. Haig maintained that urban rent "...appears as the charge which the owner of a relatively accessible site can impose because of the saving in transport costs which the use of this site makes possible."[2] This very concept of spatial equilibrium has been used in some early modern works on urban economics, such as Herbert Mohring (1961) and Lowdon Wingo (1961).

The period just after the second world war marked the beginning of a strong interest in empirical work about the spatial distribution of urban population. As we mentioned earlier, the origins can be traced back to Bleicher (1892). In modern times Stewart (1947) was the first one who observed that the exponential

takes into account that a tradition in spatial analysis already existed before Adam Smith in eighteenth century England. For example, Sir James Steuart (1767) predated von Thünen in many ways including the idea that agricultural land uses are arranged in concentric zones around a market.

[2](Haig (1926, p. 421).) As Alonso (1964, p.6) points out, there are close similarities in the treatment of urban rent between Haig and his predecessors, namely, Marshall (1890) and Hurd (1903). According to the later, since "...value depends on economic rent, and rent on location, and location on convenience, and convenience on nearness, we may eliminate the intermediate steps and say that value depends on nearness." (Hurd (p. 13).) Haig's argument is justified in the special case where the indifference curves are Leontieff type (zero elasticity of substitution between land and other commodities).

function represented a good model for describing the dependence of population density on distance from the city centre.[3] But the driving force in this area of urban research was provided by the seminal paper of Colin Clark (1951), who proposed for the first time the well–known negative exponential formula, now under his name, and proceeded to test it in a systematic fashion. His empirical analysis created the preconditions for the emergence of an important literature concerning urban population densities, land values and the reasons behind the strong regularities observed with regard to their spatial distributions.

Returning once again to theoretical issues, the far–reaching importance of Haig's complementarity hypothesis remained unexplored in its verbal state until the availability of a well–developed microeconomic theory set the preconditions for the growth of modern urban economics.[4] The first landmark in this direction was laid down by Martin Beckmann (1957), who spoke about the determination of equilibrium residential land rents and quantities in a *monocentric city*, where all employment and services are concentrated in a central business district surrounded by a residential area. He proposed an equilibrium concept for an income distribution of households, and he concluded that higher–income households locate further away from the centre in equilibrium.[5] Although Beckmann's seminar was undoubtedly prophetic on all these counts, it seems in retrospect that its scope was prematurely wide. More urgently needed during this early stage in the development of modern urban economics was a deeper analysis of narrower problems. This began with William Alonso (1960) and Richard Muth (1961), and culminated with the publication of their classic books in 1964 and 1969 respectively.

Alonso's (1964) greatest contribution can perhaps be found in his proposal for the matching between spatial analysis and microeconomic theory which, as we mention above, was necessary for the growth of modern urban economics. This matching was established through a careful extension of the microeconomic model of individual choice behaviour into an explicitly spatial context, and it gave urban economics a strong behavioural foundation at the micro–level while, at the same time, it proved convenient for aggregating individual decisions into urban structure at the macro–level. Since Alonso confined his urban residential analysis to land, rather than housing, his results were directly comparable to those of earlier land–rent studies. We already know how traditional land economists, drawing from the agricultural rent theory of von Thünen (1826), believed in the complementarity between land rents and transportation costs. Alonso's framework established that this is not generally true once the elastic-

[3]Drawing on a study which used 1940 census tract data, Stewart (p. 180) wrote that "[f]rom edge to centre, the density tends to increase exponentially with the distance, reaching a peak density in some inner census tract which is usually adjacent to others having densities nearly as great." The information is taken from McDonald (1989), a source we use repeatedly in this chapter.

[4]This point is made by Mills (1987) in his joint introductory chapter to urban economics.

[5]As observed later on by Muth (1969), this conclusion holds when the income elasticity of housing is larger than the income elasticity of the transportation rate.

ity of substitution between land and other commodities is not zero, and that the relationship between land rents and transportation costs is complex as it involves substitution and variations in tastes, income and consumption. Although Alonso extended his theory to commercial land uses as well, and although he proposed a market equilibrium for all urban land and its surrounding countryside, his great influence on modern urban economics comes from the residential part of his theory.

The difference between Alonso (1964) and Muth (1969) is threefold. Firstly, whereas Alonso is concerned about urban land, Muth is primarily concerned about urban housing. Secondly, in order to account for an heterogeneous space and the inconvenience of commuting, Alonso includes location as an argument of the utility function, while Muth does not. Thirdly, whereas the work of Alonso has a distinct theoretical flavour, Muth's work involves both empirical and theoretical aspects. The alternative utility specifications adopted by Alonso and Muth, which allow for the substitution between housing and other commodities, imply that location costs (the sum of housing and transport costs) are no longer constant as in Haig (1926) and Mohring (1961). In particular, the simplified version of Muth's utility, which became almost a standard specification, implies the useful 'principle of zero marginal location costs'. According to this principle, location cost, the sum of housing and transportation costs, is constant for marginal change in equilibrium—rather than globally constant as Haig would maintain.[6] In addition to his analysis of residential land markets, Muth investigated the formation of urban housing prices in considerable depth, and he wrote about the determinants of housing quality in inner–city neighbourhoods, including slums, conditions for their creation and policies aimed at improving them. His empirical work concerning urban population includes studies about the factors that determine net and gross densities, and studies about spatial population distribution. The latter led him to conclude that the simple model he developed in his 1961 paper to explain the Clark formula was basically supported by observation.

Any account of the *first–generation* literature on spatial population distribution, which begins after the second world war, is incomplete without reference to the work of Edwin S. Mills (1972) who devised a method for studying the temporal behaviour of population densities with a minimum of urban area and population data. In particular, assuming a negative exponential density function, Mills was "...able to obtain a perfect fit for the function yielding the population density at the CBD and the density gradient from information on the population of central city, the population of the suburbs and the distance of the central city boundary from the CBD." (Mieszkowski and Mills (1993, p. 139).) Using this method, Mills and other empirical analysts have extended and confirmed Clark's findings about suburbanisation trends around the world, that is, about a flattening of the population density gradient through time.

---

[6] In Muth's model location costs are increasing with distance from the centre.

The *second generation* of intra–urban modelling is connected with the calculus of variations—a commonly used technique in growth theory. By replacing time with distance from the centre as the running variable, more sophisticated problems became analytically manageable. In particular the issues of traffic congestion and the allocation of land to roads under first– and second–best optimality criteria reached the top of the research agenda. These issues were first raised by Robert H. Strotz (1965) and Mills (1967). Strotz's contribution represents one of the best–written articles in urban transportation theory. Building on Walters (1961) and on Mohring and Harwitz (1962), Strotz presents us with a gallery of urban transportation parables ranked according to a roughly increasing degree of complexity and developed around a small number of themes: from the basic model of a single, spaceless road under fixed traffic flow, Strotz generalises to a number of roads, to a road with variable traffic flow, to both pleasure and work trips, to issues of urban residential and urban transportation land use, and to external economies in the geographical concentration of urban economic activities. His treatment remains prophetic in the sense that it includes most of the fundamental intuitions now available on optimal transportation policy. At about the same time Strotz published his study, Mills introduced an endogenous transportation system in a peculiar manner. Individuals in his system purchase their exclusive right–of–way by paying its full market price as reflected in the land rent, which implies crowded transportation with an internalised congestion effect. However, the individual's local demand for right–of–way remains completely inelastic with respect to its price. This limitation was dealt with in the subsequent flow of publications which started with Mills and de Ferranti (1971) and Solow and Vickrey (1971). The transportation literature in urban economics achieved its peak of sophistication and policy content with the second–best analysis of allocating land to roads, which originated with Solow (1973*b*) and achieved its perfection in Kanemoto (1977) and Arnott (1979*b*).

Unlike the first generation, the second generation's perspective was mainly normative. Here the contribution of James Mirrlees (1972) represents the first, and definitive, analysis of the optimal spatial distribution of welfare within a city. His main results include a proof that the optimal city can be decentralised under an appropriate income redistribution, and that the optimal utility level increases away from the centre. Both results apply to the basic, monocentric city model populated by identical individuals. It is therefore not surprising that his second result, which is known as 'unequal treatment of equals', has been characterised by Mills and MacKinnon (1973) as one of the most intriguing in urban economic theory. According to this result, it is optimal to discriminate among otherwise identical individuals on the basis of their residential location within the city. The main reasons are that location is an indivisible resource and that, roughly speaking, locations further away from the centre make the same individual a better utility–generating machine. It was subsequently established that the results of Mirrlees apply to any distributive criterion that can be expressed as a degree of aversion to inequality, except to the criterion of maximising the utility of the least advantaged proposed by Rawls (1971), where optimality yields equal treatment of equals. Taking into account that if we abstract location, optimality yields equal treatment of equals for any distributive criterion which can be expressed as a degree of aversion to inequality, unequal treatment of equals provides a striking justification of the belief long held by spatial analysts that geography is not neutral, in the sense that introducing location can affect the standard solutions of spaceless economics.

The Polycentric City

In contrast to the rich tradition of studies about the monocentric city, relatively little has been written about the polycentric city even though urban land–use theorists were uncomfortable with the limitations of monocentricity.[7] The first model that can handle multiple centres was produced by Herbert and Stevens (1960) in the context of urban planning. Perhaps because of its discrete–space structure and its emphasis on numerical analysis, the relevance of this model to urban theory construction remained unnoticed more than fifteen years after its publication.[8] A different approach was adopted by Papageorgiou (1971) who treats the urban area as a dense concentration of various–order centres that form a nested hierarchy, with the single highest–order centre corresponding to the CBD of the polycentric city. Individuals visit several centres in order to obtain the different types of goods available in the polycentric city. This model can be based on a natural extension of Muth's principle of zero marginal location costs, so that it becomes possible to obtain a generalised Clark (1951) formula and to build urban rent and density profiles which are consistent with empirical evidence. It is also possible to provide a justification of the monocentric city model by explaining how the Clark formula may describe the general trend of the corresponding polycentric profile. However, in contrast to Herbert and Stevens, centres and their locations continue to be exogenous as in the monocentric city model; and it is not possible to investigate conditions under which urban agglomerations become monocentric or polycentric.

The question about monocentricity versus polycentricity can only be answered using models that allow for endogenous agglomerations. In the context of urban land use and rent theory, the first such model has been developed by Beckmann (1976) who considers the simplest case of a single type of agents interacting over a linear, bounded landscape without a predetermined centre. Agents in this model derive utility from their interaction and they dislike crowd-

[7] For example, Muth (1969) suggested a model in which monocentricity was no longer absolute in the sense that some of the workplaces were scattered over the residential rings. His contribution was followed by a few other papers within the mainstream of urban economics, which justify the exogenous centre as an export node and distribute firms that serve the needs of the urban population endogenously over parts of the residential rings (see Solow (1973*a*) and White (1976, 1978)).

[8] A very insightful interpretation of the Herbert–Stevens model and its comparison to the continuous model of Alonso (1964) was provided by Britton Harris (1966) and by Harris, Nathanson and Rosenberg (1966). Only later the potential applications of this model became recognised in the published literature of urban economics (see Wheaton (1974) and Fujita (1989)).

ing. If we start with a uniform distribution of agents over the land, everyone will suffer the same level of crowding but those in central locations will enjoy a higher degree of accessibility. Competition for land will eliminate this advantage through agglomeration. In this manner Beckmann derives a bell–shaped population density profile as an outcome of the fundamental trade–off between the propensity to interact and the aversion to crowding presented in section 1.2. However, Beckmann requires that the demand of individuals for interaction be infinitely inelastic with respect to its price. This limitation, together with the single–agent scope of his model, exclude the polycentric city as a possible outcome.

In two path–breaking papers, Ogawa and Fujita (1980) and Fujita and Ogawa (1982) advance significantly our understanding of urban monocentricity versus polycentricity. Their works can be thought of as an extension of Solow and Vickrey (1971) and Beckmann (1976) to more than a single type of interaction.[9] This allows them to connect the question of monocentricity versus polycentricity with the relative importance of costs and benefits that correspond to different types of interaction. They find that a monocentric city cannot be sustained for sufficiently high commuting cost—a result that agrees with our intuition. More subtle is the finding in Fujita and Ogawa (1982) that the monocentric city cannot be sustained *both* when the decay rate in the benefit of interaction between firms is sufficiently low *or* sufficiently high. Therefore a monocentric city emerges only at intermediate values of that decay rate.

We close this section with a model of the polycentric city developed along the local public good and club theories of Charles M. Tiebout (1956) and James Buchanan (1965).[10] Hochman, Pines and Thisse (1995) imagine a system of spatial clubs that correspond to a polycentric city, each supplying a single collective good to a group of identical individuals who pay a user charge to partly finance its cost. Individuals belong simultaneously to several clubs that differ from one–another with respect to the collective good they supply and, therefore, to the group size they serve. The spatial structure of this polycentric city can remind us in some respects of the corresponding hierarchical structure in Papageorgiou (1971), although in the first case the composition of clubs and their locations are endogenous while in the second case it is not. Hochman, Pines and Thisse show that an optimal club complex of this kind must be compact, and that every sub–area of the urban territory at the optimum is involved in extensive spatial interaction with other sub–areas. A significant implication of their model is the policy proposal for city–wide governments that runs con-

---

[9] They take into account interactions between employers and employees, which require commuting of each employee to a single workplace, as well as interactions between any pair of firms.

[10] The difference between the two theories is vague since, in both, a collective good is efficiently provided when its use is subject to an appropriate user charge. Sometimes, however, a distinction is made according to whether more than one collective good is simultaneously used by the same group and/or whether the group is engaged not only with a common use of the collective good but also with production. As noted by Berglas and Pines (1981), this distinction is somewhat artificial.

trary to the basic policy proposal of fiscal federalism. First elaborated by Olson (1969), fiscal federalism calls for the decentralised provision of collective goods as a basis for reforming the present system of local governments. However, as Hochman, Pines and Thisse point out, the global justification of this policy crucially depends on the non–spatial nature of fiscal federalism. When geographical space is introduced, decentralised provision is no longer possible in the case of spatially overlapping jurisdictions.

Finally, it is significant to notice that this model can yield the non–monotonic effect of the basic centrifugal force (interaction among firms) on monocentricity versus polycentricity, which is similar to the one by Fujita and Ogawa (1982), although due to completely different reasons.

1.3.2 Size and Functional Distribution of Cities

Optimal City Size Under Perfect Replicability and Divisibility

The literature on this topic can be traced back to Lösch (1940) who maintains that there exists an optimal city size as the outcome of two opposing scale effects on feasible utility: a positive scale effect that stems from the agglomeration of economic activities and a negative scale effect that stems from crowding. The formal description of the trade–off between these two scale effects, and the consequent $\cap$–shape of the utility level as a function of city size, was illustrated as an equilibrium only in the 1970s by Dixit (1973) and Henderson (1974). In both models scale economies are represented by increasing returns in the production of a composite good, where the scale effect is external to the individual firm in order to allow for a competitive allocation. For scale diseconomies, Dixit uses the standard case of a monocentric city with internalised traffic congestion, while Henderson, who suppresses the spatial aspect of the city, uses land scarcity and its effect on the production of the composite good and housing.[11] Another source for the $\cap$–shape utility is provided within the tradition of local public good (LPG hereafter) and club theories. The common positive scale effect here is represented by the declining burden of sharing the cost of the LPG as the population increases. By contrast scale diseconomies vary among authors, from land scarcity and its effect on the production of the composite good in the non–spatial model of Stiglitz (1977), to residential crowding in Arnott (1979*a*), to congestion from using the collective good in the standard club formulation of Berglas and Pines (1981) and Scotchmer and Wooders (1987). Whatever the source of scale economies and diseconomies, an optimal population size does emerge provided that scale economies dominate for sufficiently small population size and scale diseconomies for sufficiently large.

A significant outcome which is common to all the above diverse models is the vindication of Henry George (1896) and his single–tax proposal. Influenced by Thomas Malthus (1798) and David Ricardo (1821), George believed that land owners can take all the excess value derived by economic development

[11] Henderson (1986) elaborates on the primitive spatial foundations of this reduced model.

through land rent. He also contended, along with other nineteenth century rent theorists, that it is wrong for land owners to gain profit by renting their land since land itself is not a product of labour. Consequently George proposed that all land rents be taxed away and be used to finance government expenditures. At the time it was estimated that the single tax would be sufficient to support all levels of government in the United States. The relevance of the single–tax proposal to urban economics stems from a diversified set of modern results, including Flatters, Henderson and Mieszkowski (1974) and Arnott and Stiglitz (1979), which connect the aggregate 'situation value' of Marshall (1890) with the cost of providing an urban infrastructure in what has become known as the 'Henry George theorem'. More precisely the single–tax proposal is justified because, under marginal cost pricing, the aggregate urban land rent net of its opportunity cost plus the sum of user charge receipts precisely equals the corresponding aggregate cost for any level of provision at the optimal city size.

The relevance of optimal city size to the issue of city size distribution becomes transparent if cities are perfectly replicable and divisible, that is, if the total population can always be allocated to cities of optimal population size.[12] In that case, according to the Henry George theorem, land rent combined with the appropriate user charge in each city would not only suffice to finance its LPG, but it would also provide the incentive necessary to ensure its optimal provision. Suppose that there is only one tradeable private good. If the population is homogeneous then the optimal urban system is a set of identical optimal–size cities; and if the population is heterogeneous then the optimal urban system has a number of subsystems, one for each population type, each containing a set of identical optimal–size cities. But once there are more than one tradeable goods, a rich gamut of urban systems can emerge with intra–city trade. The reasons for specialisation and trade in those systems, as well as the associated city size distribution, vary with the underlying agglomeration advantages and disadvantages. In the standard tradition of urban economics, where the source of agglomeration advantage is increasing returns in production, it is more efficient to establish single–product cities. In the LPG tradition, where the source of agglomeration advantage is the declining cost of sharing the LPG as the population increases, specialisation and trade become desirable *even if the source of the agglomeration disadvantage is decreasing returns to scale in production.* The first results in this context belong to Wilson (1987), who found that the optimal urban structure for identical individuals is not necessarily a system of identical autarchies, but rather diversified city types differing from each other by size, production mix and consumption. A generalised version of his original result shows that specialisation admits gains from trade at an urban population size other than the optimal size of an autarchy. Since utility decreases away from the optimal city size, the choice between a uniform or diversified urban

[12] By replicability we mean that the number of potential cities is not constrained. By divisibility we mean that the total population can be divided into an integer number of optimal–size cities.

system hinges upon whether or not the losses from inefficient population size are greater than the corresponding gains from trade.

Resource Misallocation Under Externalities and Nonconvexities

There are two main causes of market failure associated with city size distributions. The first one is externalities coupled with insufficient policy instruments at the disposal of city governments. The second is either non–replicability or indivisibility or both, which we call in general *nonconvexity*. Discussions in the literature about what causes such failure are often confusing.

A well–known debate about whether or not external economies and diseconomies induce excessively big cities can be illustrated using the contributions of Tolley and Grihfield (1987) on the one hand and Mills and Hamilton (1984) on the other. Tolley and Grihfield defend the view that unpriced pollution and traffic congestion induce excessively big cities. They argue that, because of external diseconomies, there exists a gap between (1) the marginal social cost of providing an additional inhabitant with the prevailing utility and (2) the minimum market value of the consumption bundle required for such utility; and that this gap *increases* with city size. Mills and Hamilton challenge this conclusion on the ground that external scale economies also exist, which imply that there is a second gap with the opposite sign so that the sum of the two gaps can change in the opposite direction.[13] Hence it is not even clear whether the social cost of providing an additional inhabitant with the prevailing utility exceeds or is exceeded by the market cost. It appears that such a debate can be reduced to the empirical question about whether or not external scale diseconomies are stronger than the corresponding external economies, as indeed suggested by Mills and Hamilton. However, several issues in this debate need clarification before resorting to empirical analysis (see chapter twelve).

More recently, the same issue has been reconsidered in the context of the 'new economic geography' which we present in the next section. As we shall see, both scale and external economies in this literature are created by the production of differentiated goods. In all three models we discuss in this paragraph stable equilibria can produce an urban system with one small and one big city, and one of the issues addressed is whether the size of the big city is too large. In Helpman (1998) scale diseconomies are provided by residential crowding which is generated by a fixed amount of land, but there are no external diseconomies. Thus one tends to conclude that if there is market failure it must cause insufficient agglomeration as conjectured by Mills and Hamilton. This is precisely Helpman's conclusion. However, by just changing the utility specification in Helpman's model, Hadar (1997) produces a counter–example so that we are led to the paradoxical conclusion that, depending on the structure

[13] In both cases we use our own terms for compatibility. Observe that, in our terms, *scale* economies exert a *centripetal* effect but *external* economies exert a *centrifugal* effect (or they *reduce* the centripetal effect). Symmetrically, *scale* diseconomies exert a *centrifugal* effect but *external* diseconomies exert a *centripetal* effect (or they *increase* the centripetal effect). Beware that these terms are at odds with much of the literature.

of utility, external economies can be consistent with both insufficient and excessive agglomeration in big cities. A symmetrical conclusion is drawn by Tabuchi (forthcoming) who attributes his result about excessive agglomeration in big cities to the external diseconomies resulting from urban transportation. But a closer look at his model reveals that although a crowding effect does exist, which is associated with scale diseconomies alone, there is no traffic congestion of any sort—hence there are no external diseconomies. So, where is the culprit? It seems that the answer is *nonconvexity*, which is perhaps a more important source of market failure than external economies and diseconomies.

1.3.3 Economic Geography of Cities

Our interest about city size distributions, urban specialisation and the relationship between cities and their agricultural hinterland is closely related to central place theory. Its roots can be traced back to the times of Richard Cantillon (1755) and Adam Smith (1776), who wrote extensively about the rôle of cities in a national economy and about how important is transportation in shaping prices and market areas. But the fundamental concept of a central place system was first elaborated in a comprehensive manner by Walter Christaller (1933), whose explicit aim was to understand the laws that determine the *number*, *size*, *function* and *spacing* of settlements over an homogeneous area. It has often been said that Christaller's deductive structure is a theory about the location of tertiary activities, which stands alongside the work of von Thünen (1826) on the location of primary activities and that of Weber (1909) on the location of secondary activities. More importantly, it can be seen as the original integrative framework which introduced the basic concepts necessary for Lösch's theory of economic regions. Stepping on Christaller's shoulders, Lösch (1940) gave us in first approximation reasons why economic activities tend to agglomerate over an otherwise featureless plain. However, with the exception of Walter Isard (1956), his monumental work stood alone, probably because the research programme it suggested was too comprehensive—hence too difficult to undertake at the time.

Only recently have the fundamental aims of central place theory begun to emerge once again in the modern work on agglomeration, where aspects of location theory and urban economics integrate to produce models stressing more than anything else the extensive spatial interdependence of various economic activities over the landscape. If we recall the four basic aims in the research programme of Christaller, the first three concerning the *number, size* and *function* of settlements have already drawn a lot of interest as we have seen in section 1.3.2. But geographical configuration has been completely abstracted from these studies. More specifically *spacing* between cities has been ignored.

One of to-day's interesting problems, which is closely related to the spacing issue, is to strive for an *endogenous* central place theory in which central places will appear as a transition from a spatially uniform equilibrium to an agglomerated one. The objective then is to characterise the "succession of form", as Thom (1972) has named it in order to understand *how cities are born.*[14]

Papageorgiou and Smith (1983) have considered this problem in the case of a single type of identical agents who interact over an unbounded and homogeneous landscape. The spatially uniform distribution in this system is an equilibrium, and the question is under what conditions will it become locally unstable. Since any configuration other than the uniform, stable equilibrium implies some agglomeration, such an approach is sufficient. Papageorgiou and Smith found that the critical instability is characterised by an exact balance between the marginal disutility of local congestion and the marginal utility of global interaction. Agglomerations emerge when the positive interaction effect dominates over the negative congestion effect. In more recent studies by de Palma and Papageorgiou (1991*a,b)* it has been shown that the emerging regular pattern of potential settlements is degenerate in the case of a single type of interacting agents. Only when there is some specialisation, say, two types of agent, will the transition from a spatially uniform equilibrium produce a non–degenerate, regular settlement pattern. In such cases the *emergence* of a settlement pattern depends both on the preferences and on the characteristics of interaction among agents. By contrast, the *spacing* of settlements depends on the characteristics of spatial interaction alone.

The efforts to create an endogenous central place theory have been accelerated with the coming of the 'new economic geography'. The impetus in this area has been provided by Paul Krugman (1991) who applied a differentiated product approach within a monopolistic competition framework, developed by Dixit and Stiglitz (1977), in order to study how agglomeration shapes urban structure. In all the papers that follow his approach there is a unique, fundamental agglomeration advantage manifested in a self–enforcing, dynamic manner ('circular causation'). Namely, an increase of the population in a particular city implies an increased demand for brands which attracts new firms, each producing a new brand ('backward linkages'). It follows that the share of brands locally produced increases relative to brands imported from other cities, which reduces the average delivery price of differentiated products and increases real income—thus enhancing utility. Higher utility, in turn, encourages further immigration (forward linkages).[15]

There are several characteristics that generate this unique agglomeration advantage, which translates scale economies on the level of the individual firm to

[14] Notice that, although spacing has been a matter of concern for central place theorists, central places simply exist there and the problem is to account for their perfectly even arrangement on a Euclidean plane over which the exogenous spatial distribution of demand is perfectly even. In this respect classical central place theory is *exogenous.*

[15] The specification of the differentiated products in the utility formalises Jane Jacobs' (1969) idea that adding more works is more important for growth than expanding the volume of existing works (see for example Jacobs p.122). Thus, for a given increase of total output, further diversification is more important than larger volume of the same varieties.

scale economies on the aggregate level.[16] Firstly, due to scale economies in the production of a brand, each brand is produced by a single firm which is located in a single city. Therefore each city specialises in the production of a range of brands within the spectrum of differentiated products. Secondly, the utility specification implies that every individual consumes every brand wherever it is produced. Thirdly, the transportation cost is in the form of 'melting iceberg', that is, the amount shipped declines with distance from the origin. This specification implies that the mill price is the same, so that the city with the larger range of brands has an advantage on the other city which must import a larger proportion of the manufactured good.

In contrast to the unique agglomeration advantage they explore, the detailed reasons for dispersion underlying these models vary widely. With only few exceptions, the dispersion force is provided by some immobile resource located somewhere in the region. In the early studies the immobility of land was represented by immobile farmers. Only in a few studies the reasons for dispersion are to be found within the city, as in Helpman (1998) and Tabuchi (forthcoming) above. A significant characteristic in these models is that, as the reasons for dispersion vary, so do their conclusions. For example, consider the case where transportation cost affects agglomeration. In Krugman (1991), where the dispersion force is provided by an immobile agricultural population, low transportation cost leads to agglomeration and high transportation cost leads to dispersion. By contrast in Helpman (1998), where the dispersion force is provided by residential crowding, low transportation cost leads to dispersion and high transportation cost leads to agglomeration.

The reason why this class of models belongs to the new economic geography, rather than to the topic of city–size distributions, is the inter–city trade and its cost. Then and only then geography matters, in other words, distances between cities affect the equilibrium allocations. However, in the earlier models of the new economic geography that began with Krugman (1991), spacing is exogenous: although geography matters, it is not determined by the model itself.

A truly endogenous urban geography model that follows the tradition of differentiated products within a monopolistic competition framework has been proposed by Mori (1997) and Fujita and Mori (1997). Starting with Fujita and Krugman (1995), who define *sufficient* conditions for a single agglomeration over a bounded landscape, Fujita and Mori examine cases in which these conditions are violated so that either agglomeration or dispersion can prevail. As it turns out, a wide range of interesting possibilities do emerge. They find that a single agglomeration cannot be sustained if the transport cost of the differentiated good is *either* sufficiently large *or* sufficiently small relative to the transport cost of the agricultural good—which is reminiscent of Ogawa and Fujita (1980). The robustness of this result in such different contexts is worth

[16] Aggregate scale economies imply that the equilibrium utility level increases with population size.

mentioning, especially because the monopolistic competition framework based on Dixit and Stiglitz (1977) is *very* sensitive to specification as we have already seen. Suppose we start with a single agglomeration. On the one hand, raising sufficiently the transport cost of the differentiated good, Mori (1997) finds an equilibrium of two distinct cities where the spacing between the two is given endogenously as in de Palma and Papageorgiou (1991*b*). On the other hand, lowering sufficiently the transport cost of the differentiated good, he finds an analogous equilibrium such that the interval between the two cities is occupied by a continuous manufacturing belt where the density of workers is lower than the corresponding density in the two cities. This is the first model of megalopolis, first described by Gottman (1961) in an empirical context and elaborated by Doxiades (1968) in a planning context.

1.3.4 Some Final Comments

If we consider the developments that took place during the 1980s and the 1990s, two names emerge: Masahisa Fujita and Paul Krugman. These two, along with their students and other co–authors, have been more instrumental than anyone else in shaping the research agenda of spatial analysis to–day. A significant part of this research agenda rests on Krugman's (1995, pp.63–64) belief that his basic model, once extended to two dimensions will "... produce a lattice of central places with hexagonal market areas: Lösch vindicated."[17] With somewhat less confidence, he also expects to reproduce Christaller–type hierarchies and, probably, the rank–size rule of Zipf (1949). These current research priorities have imposed on him and his students a big challenge. Several related works have appeared, based on a dynamic framework in which the urban system is evolving through time as the population grows. This subject however is beyond the scope of our book.

Urban economics, regional science and geography have benefitted significantly from the recent crusade initiated by Krugman, a mainstream economist, set forth to convince the economics profession that the new economic geography should be accepted as a legitimate branch of economics with recognised achievements. Nevertheless, we must reject the unfounded criticism of what Krugman (1995, pp.57–58) refers to as 'new urban economics' which, in our terminology, represents the urban land use and rent theory as it developed after the second world war. We reject this criticism explicitly because we believe that one of the main weaknesses of the new economic geography is precisely its overlooking the robust implications of the monocentric and polycentric city models. Every model is a drastic simplification of the phenomenon it purports to illuminate, hence every model can be attacked for some of its unrealistic assumptions. But such criticism, although true, is also vacuous. For what matters is whether or not unrealistic assumptions simplify the problem in a way that allows for a clear, partial and undistorted glimpse of the object we are interested in learning

[17] The same expectation arises explicitly in de Palma and Papageorgiou (1991*b*).

about. Therefore blanket attacks based on the selective use of a field's unrealistic assumptions are not productive, especially if they are not balanced with corresponding fair criticisms about the field one wants to defend and promote. Assuming for example that every individual commutes to the centre—as necessarily implied by the monocentric city model, is no more unrealistic than assuming that every individual in New York City enjoys the services of every pub in New York State—as necessarily implied by the formulation of the new economic geography. Both assumptions are unrealistic, but both have produced some remarkable insights at both the intra– and the inter–city levels. More importantly, much is learned by the interaction between those different and partial approaches. For example, in an increasingly urban world, it makes sense to replace agriculture as the main source of dispersion with urban crowding, which is carefully modelled and documented in the standard literature of urban economics. Such cross–fertilisation, which is entirely possible as the studies of Hadar (1997) and Tabuchi (forthcoming) clearly suggest, will create a stronger new economic geography and, at the same time, will highlight the contributions of urban land use and rent theory.

1.4 Book Outline

The first part presents in a coherent way various aspects of the monocentric city model—the most developed theory in urban economics. Chapter two discusses the basic model, investigates the characteristics of equilibrium allocations and derives the standard negative exponential function which has been used extensively in the empirical literature for describing the spatial distribution of population density and land values. Chapter three examines how equilibrium allocations are affected by exogenous factors such as the transportation technology, the urban infrastructure and so on. We distinguish between open and closed cities, which correspond to early and late stages of urbanisation respectively, as well as among different patterns of urban property rights. In chapter four we introduce normative considerations and we find that the equilibrium allocation of chapter two is Pareto–efficient. By contrast, although the equilibrium allocation of chapter two yields the same utility for identical individuals over the entire urban area, the optimal allocation of chapter four yields different utility levels for identical individuals at different locations unless the criterion for optimality is to maximise minimum welfare. In chapter five we consider four direct extensions of the monocentric city model. Firstly, we examine the implications of a pure locational preference which can arise either when leisure is explicitly taken into account, or in the presence of heterogeneous locational attributes such as industrial pollution, scenic landscapes and so on. Secondly, we introduce production at the centre which provides employment to the urban labour force. This allows for an explanation of sudden urban growth which, without exception, has characterised the evolution of large cities around the world. Thirdly, we discuss the case of a city in which individuals may differ by income, we explain the locational principles of various income groups and how such a city adjusts as its population or income grows. Finally, we extend the standard monocentric model into a polycentric city using some key concepts of central place theory and we derive a polycentric counterpart of the negative exponential function. The next two chapters discuss two further extensions of the monocentric city model which are significantly more involved than the extensions of chapter five. Chapter six introduces a more detailed description of housing. Whereas in previous chapters housing consumption was treated as equivalent to current consumption of services rendered by land alone, housing here is a service created by land, capital and labour. Since housing is not perfectly malleable and its adjustment is costly, a clear distinction is made in this chapter between the housing stock and the flow of housing services within a dynamic analytical framework. In chapter seven we examine transportation infrastructure in some detail and we elaborate on specific reasons why it is publicly provided rather than supplied by competitive markets as, for example, housing. We find that public intervention becomes indispensable either because it is too costly to exclude or there are increasing returns to scale—rather than because of external effects as is commonly held in the literature. The first part of our book closes with chapter eight, which provides a bridge between the two parts. In this chapter we defend the utility of the monocentric paradigm and we give reasons why a robust polycentric model capable of yielding testable hypotheses is a most important research challenge to–day.

In contrast to the standard monocentric model, there is no coherent theory about the polycentric city and about city systems—only fragments of theories. Some of these are discussed in chapter nine, which displays a gallery of models organised around the concept of agglomeration as an equilibrium outcome of the fundamental trade–off between the propensity to interact and the aversion to crowding introduced in section 1.2. Polycentric agglomerations may represent a city comprising a number of centres within a relatively small area, or a system of monocentric cities, or an integrated regional system in which distinct urban production and market activities are no longer aggregated into single points. The next two chapters provide a systematic exploration of these spatial systems from the unified perspective of club and local public good theories. In chapter ten we introduce the idea of an urban centre as a collective facility used by the whole or part of the urban population. A single collective facility in this context provides the analog of a monocentric city, while a complex of facilities corresponds to a polycentric city. We characterise the optimal size of such a complex and we explain why decentralisation must be undertaken at the level of the urban territory itself, rather than at the level of individual facilities as suggested by fiscal federalism. We then study in detail a simplified version of this model in order to understand what characterises the difference between a monocentric and a polycentric urban structure, and what determines whether the location of a public facility will be central or peripheral. Chapter eleven investigates the reasons behind the formation of heterogeneous groups by identical individuals when there are more than a single private good—a structure that leads to specialisation and trade among cities. We use the reduced form of a model that sheds new light on the advantages and disadvantages of choosing an heterogeneous group of cities with specialisation and trade over a system of identical autarchies. Which of the two is preferable depends on the size of gains from trade in a diversified system relative to gains from an efficient city size in a system of identical autarchies. Finally, in chapter twelve, we discuss market failure at the intra–urban and the inter–urban levels due to resource misallocation arising from external economies and diseconomies or nonconvexity. Choosing the correct reason for market failure is important in order to propose the right policy prescription. We emphasise this issue throughout, and we compare the 'right' policy with laissez–faire in every case examined. We investigate resource misallocation in the context of unpriced intra–urban traffic congestion, as well as the rôle of nonconvexity in distorting the distribution of people between cities and their hinterlands.

1.5 Appendix: Notation

We use three types of brackets for three different purposes. Square brackets contain the domain of a function. For example, $f[x,y]$ denotes a function with domain (x,y). Angled brackets contain the specific value of the variable(s) in the domain at which a function is evaluated. For example, $f\langle x'\rangle$ denotes the same function at (x',y). Finally, curved brackets are used to collect terms. For example, $g[x+y]$ denotes a function with domain $(x+y)$, but $g(x+y)$ means that g is multiplied by $x+y$.

As a shortcut, we often use

$$\frac{\partial f^i}{\partial x} \equiv \frac{\partial f\langle x^i\rangle}{\partial x}$$

where the superscript $i = e, *, \circ$.

A basic distinction for all parameters and variables used in our analysis is whether or not they vary with distance x from the centre. Those that do not vary with distance from the centre are indicated with a bar. For example, $U \equiv U[x]$ but $\bar{U}$ does not vary with distance from the centre. In a few places where this distinction does not apply (because geographical space is absent) we state it explicitly. We denote $a \le x \le b$ as $x \in [a,b]$; $a < x \le b$ as $x \in (a,b]$; $a \le x < b$ as $x \in [a,b)$; and $a < x < b$ as $x \in (a,b)$.

1.5.1 *Latin Symbols*

- a Aggregate marginal resource cost of travel on a particular road at a particular distance from the centre.

  A Demolition age of a housing unit.

  AC Average cost.

  ADC Aggregate development cost.

  ADC Aggregate development cost.

ADT Aggregate distance travelled.

ALR Aggregate land rent.

ATC Aggregate transportation cost.

ATD Aggregate transportation demand.

- b Transportation infrastructure function.

  B Transportation infrastructure corresponding to a particular road.

- c Cost function for the provision of a local public good.

  C Cost of providing a local public good.

  CBD Central business district.

- d Differential.

  D Population density.

  DLR Aggregate differential land rent.

- e Minimum expenditure function.

  E Minimum expenditure level.

  $\mathcal{E}$ Economy.

  EDF Exponential density function.

- f Function.

  F Function.

  $\mathcal{F}$ Fundamentals of the economy.

- g Compensated demand function for the local public good.

  G The frequency of visits to a public facility by an individual. The quantity of a local public good consumed by an individual.

  $\mathbf{G}$ A vector of frequencies. A vector of quantities of local public goods.

  $\bar{G}$ The aggregate visits to a public facility. The total demand for all public facilities in a city. Crowding.

- h Compensated demand function for housing.

  H Amount of housing consumed by an individual.

  $\mathcal{H}$ Hamiltonian.

- i Index.

  I

- j Index.

  $\mathcal{J}$ Maximum revenue per housing unit over one building cycle.

- k Relative share of the urban population in the aggregate profit from urban land transactions.

  K Housing construction cost per unit of land.

- l Width of a road.

  ℓ Leisure.

  L Land.

  L_t Total width of land used in transportation.

  $\mathcal{L}$ Lagrangean function.

  LPG Local public good.

- m Expenditure on maintenance per housing unit.

  $\bar{m}$ Number of available brands. Number of firms in a city.

  M Number of commuters using a specific road.

  $\bar{M}$ The total number of customers serviced by a particular public facility. Urban population size.

  $\mathcal{M}$ Market Structure.

  MPC Marginal private cost.

  MRS$[x, y]$ Marginal rate of substitution of x for y.

  MSB Marginal social benefit.

  MSC Marginal social cost.

  MSP Marginal social product.

  MV Market value of a bundle.

- n First derivative of N with respect to distance.

  $\bar{n}$ Number of levels in a central place hierarchy. Number of types of collective goods.

  N Urban population living within a particular distance from the centre.

  $\bar{N}$ Total urban population.

  $\bar{N}^d$ Total demand for residential sites.

  $\bar{N}^s$ Total supply of residential sites.

  $\overline{\mathbb{N}}$ Total regional population.

  NTR Net toll revenue at a particular distance from the centre.

- o

  O

- p Rent per dwelling unit.

  p_t A function that determines the price of the right to use the road.

  P Price.

  $\bar{P}$ User charge per visit to a public facility. Price.

  $\bar{\mathbf{P}}$ A vector of prices.

  P_t Price of the right to use the road.

- q Quality of housing.

  Q Total trip demand at a particular distance from the centre.

- r Bid–rent function.

  $\bar{r}$ The interest rate.

  R Land rent.

  $\bar{R}$ Cost of developing a unit of land. It equals the agricultural land rent.

  $\mathbb{R}$ The set of real numbers.

- s Traffic speed.

  S Hicks neutral shift factor.

  $\mathfrak{S}$ Spacing of centres.

  SWF Social welfare function.

- t Transportation rate, defined as the first derivative of T with respect to distance.

  T Transportation cost function.

  $\mathbb{T}$ Congestion toll.

- u Utility function.

  U Utility level.

  UPF Utility possibility frontier.

- v Indirect utility function.

  V Maximum feasible utility level.

- w Weight assigned to the utility of an individual.

  W Social welfare function.

  $\bar{W}$ Aggregate level of social welfare.

- x Distance from the centre.

  $\mathbf{x}$ A vector of distances to various order centres.

  $\bar{x}$ Maximum distance of an urban location from the centre. This is the border between urban and agricultural land use.

$\hat{x}$ Average distance travelled.

X

$\mathcal{X}$ Urban area.

- y Quantity of a differentiated good manufactured by a particular firm which is consumed by an individual.

  Y Per capita income.

- z Compensated demand function for the composite good.

  z_t Amount of the composite good used in the construction of a road.

  Z Amount of composite good consumed by an individual.

  $\mathbf{Z}$ A vector of amounts of composite goods consumed by an individual.

  $\mathbb{Z}$ Per capita amount of a private good used for consumption and for the production of a public good.

  $\mathcal{Z}$ The quantity of composite good produced by a firm.

1.5.2 Greek Symbols

- α A shift parameter which represents the quality of transportation infrastructure. A parameter.

- β A shift parameter which represents the possibility of reclaiming unavailable land. A parameter.

- γ Amount of the public good consumed by an individual. A parameter.

  Γ

- δ The density gradient.

  Δ Difference.

- ε Degree of homogeneity.

  ϵ Infinitesimal distance.

- ζ Degree of aversion to inequality.

- η Elasticity.

- θ First derivative of Θ with respect to distance.

  Θ The available land function, which determines the total amount of available land within a particular distance from the centre.

- ι

- κ Number of clubs. Number of public facilities.

- λ Lagrangean multiplier.

  Λ

- μ Lagrangean multiplier.

- ν Lagrangean multiplier.

- ξ Discounting factor.

  Ξ

- π 3.14...

  ϖ

  Π Maximum profit.

- ρ Structural density of housing.

- σ The ratio θ/t.

  ς A shift parameter.

  Σ Sum.

- τ Time.

- υ Utility–size configuration.

  Υ The resources allocated to a public facility. The characteristics of a public facility. The total resources allocated to public facilities in the city.

- ϕ An angle.

  φ

  Φ Tax or subsidy.

- χ

- ψ A function that determines the marginal resource cost of travel.

  Ψ Marginal resource cost of travel.

- ω Wage rate.

  Ω Per capita initial endowment of composite good.

1.6 References

W. Alonso, 1960, A Theory of the Urban Land Market. *Papers and Proceedings of the Regional Science Association* 6, 149–158.

W. Alonso, 1964, *Location and Land Use.* Harvard University Press, Cambridge, Mass.

R. J. Arnott, 1979*a*, Optimal City Size in a Spatial Economy. *Journal of Urban Economics* 6, 65-89.

R. J. Arnott, 1979*b*, Unpriced Transportation Congestion. *Journal of Economic Theory* 21, 294–314.

R. J. Arnott and J. E. Stiglitz, 1979, Aggregate Land Rents, Expenditure on Public Goods, and Optimal City Size. *Quarterly Journal of Economics* 93, 471–500.

M. Beckmann, 1957, On the Distribution of Urban Rent and Residential Density in Cities. Paper presented at the interdepartmental seminar on Mathematical Applications in the Social Sciences, Yale University, New Haven.

M. Beckmann, 1976, Spatial Equilibrium in a Dispersed City. In Y. Y. Papageorgiou (ed.) *Mathematical Land Use Theory*, Lexington Books, Lexington Mass., 117–125.

E. Berglas and D. Pines, 1981, Clubs, Local Public Goods, and Transportation Models: A Synthesis. *Journal of Public Economics* 15, 141–162.

H. Bleicher, 1892, *Statistische Beschreibung Der Staadt Frankfurt Am Main Und Ihre Bevölkerung.* Frankfurt Am Main.

J. M. Buchanan, 1965, An Economic Theory of Clubs. *Economica* 32, 1-14.

R. Cantillon, 1755, *Essay on the Nature of Trade in General.* Reprinted by Augustus Kelley, Bookseller, New York, 1964.

W. Christaller, 1933, *Die Centralen Orte in Süddeutschland.* Translated in English as *Central Places in Southern Germany* by C. W. Baskin, Prentice–Hall, Englewood Cliffs, N.J., 1966.

C. Clark, 1951, Urban Population Densities. *Journal of the Royal Statistical Society* 114 (Series A), 490–496.

A. de Palma and Y. Y. Papageorgiou, 1991*a*, Agglomeration as Local Instability of Spatially Uniform Steady–States: A Note on Degeneracy. *Working Paper* 91–08, Department of Economics, McMaster University.

A. de Palma and Y. Y. Papageorgiou, 1991*b*, Settlement Systems as Spontaneous Order. McMaster University, (unpublished).

A. K. Dixit, 1973, "The Optimum Factory town", *Bell Journal of Economics and Management Science* 4, 637-654.

A. K. Dixit and J. E. Stiglitz, 1977, Monopolistic Competition and Optimum Product Diversity. *American Economic Review* 67, 297–308.

C. A. Doxiades, 1968, *Ekistics: An Introduction to the Science of Human Settlements.* Hutchinson, London.

R. B. Ekelund Jr. and R. F. Hubert, 1993, Cycles in the Development of Spatial Economics. In H. Ohta and J.– F. Thisse (eds.) *Does Economic Space Matter? Essays in Honor of Melvin L. Greenhut.* St. Martin's Press, New York, 21–37.

F. Flatters, J. V. Henderson, and P. Mieszkowski, 1974, Public Good, Efficiency, and Regional Fiscal Equalization. *Journal of Public Economics* 3, 99-112.

M. Fujita, 1989, *Urban Economic Theory.* Cambridge University Press, Cambridge.

M. Fujita and P. Krugman, 1995, When is the Economy Monocentric? von Thünen and Chamberlin Unified. *Regional Science and Urban Economics* 25, 525–528.

M. Fujita and T. Mori, 1997, Transport Development and the Evolution of Economic Geography. Kyoto University, (unpublished).

M. Fujita and H. Ogawa, 1982, Multiple Equilibria and Structural Transition of Non–Monocentric Urban Configurations. *Regional Science and Urban Economics* 12, 161–196.

H. George, 1896, *Progress and Poverty.* Reprinted by the Robert Schalkenbach Foundation, New York, 1954.

J. Gottman, 1961, *Megalopolis: The Urbanized Northeastern Seaboard of the United States.* MIT Press, Cambridge, Mass.

Y. Hadar, 1997, The Formation of Cities. Tel–Aviv University, (unpublished).

R. M. Haig, 1926, Toward an Understanding of the Metropolis. *Quarterly Journal of Economics* 40, 421–423.

B. Harris, 1966, Basic Assumptions for a Simulation of the Urban Residential Housing and Land Market. The University of Pennsylvania, (unpublished).

B. Harris, J. Nathanson and L. Rosenberg, 1966, Research on Equilibrium Model of Metropolitan Housing and Location Choice: Interim Report. The University of Pennsylvania, (unpublished).

E. Helpman, 1998, The Size of Regions. In D. Pines, E. Sadka, and I. Zilcha (eds.) *Topics in Public Economics.* Cambridge University Press, New York, 33–54.

J. V. Henderson, 1974, The Sizes and Types of Cities, *American Economic Review* 64, 640-656.

J. V. Henderson, 1986, Efficiency of Resources Usage and City Size. *Journal of Urban Economics* 19, 47–70.

J. D. Herbert and B. H. Stevens, 1960, A Model for the Distribution of Residential Activity in Urban Areas. *Journal of Regional Science* 2, 21–36.

O. Hochman, D. Pines and J.–F. Thisse, 1995, On the Optimal Structure of Local Governments. *The American Economic Review* 85, 1224–1240.

R. M. Hurd, 1903, *Principles of City Land Values.* The Record and Guide, New York.

W. Isard, 1956, *Location and Space Economy.* MIT Press, Cambridge, Mass.

J. Jacobs, 1969, *The Economy of Cities.* Random House, New York.

Y. Kanemoto, 1977, Cost–Benefit Analysis and the Second–Best Land Use in Transportation. *Journal of Urban Economics* 4, 483–507.

P. Krugman, 1991, Increasing Returns and Economic Geography. *Journal of Political Economy* 99, 483–499.

P. Krugman, 1995, *Development, Geography and Economic Theory.* Cambridge University Press, Cambridge.

A. Lösch, 1940, *Die Raumliche Ordnung der Wirtschaft.* Translated in English as *The Economics of Location* by W. H. Woglom and W. F. Stopler. Yale University Press, New Haven, 1954.

T. Malthus, 1798, *An Essay on the Principle of Population as it Affects the Future Improvement of Society.* Reprinted by Pelican, London, 1970.

A. Marshall, 1890, *Principles of Economics.* Reprinted by MacMillan, London, 1946.

J. F. McDonald, 1989, Econometric Studies of Urban Population Density: A Survey. *Journal of Urban Economics* 26, 361–385.

P. Mieszkowski and E. S. Mills, 1993, The Causes of Metropolitan Suburbanization. *Journal of Economic Perspectives* 7, 135–147.

E. S. Mills, 1967, An Aggregative Model of Resource Allocation in a Metropolitan Area. *American Economic Review* 57, 197–210.

E. S. Mills, 1972, *Studies in the Structure of the Urban Economy.* The Johns Hopkins University Press, Baltimore.

E. S. Mills (ed.), 1987, *Handbook of Regional and Urban Economics* II*: Urban Economics.* North–Holland, Amsterdam.

E. S. Mills and D. M. de Ferranti, 1971, Market Choices and Optimum City Size. *American Economic Review* 61, 340–345..

E. S. Mills and B. W. Hamilton, 1984, Urban Economics. Harper Collins College Publishers, New York.

E. S. Mills and J. MacKinnon, 1973, Notes on the New Urban Economics. *Bell Journal of Economics and Management Science* 4, 593–601.

J. A. Mirrlees, 1972, The Optimum Town. *Swedish Journal of Economics* 74, 114–135.

H. Mohring, 1961, Land Values and the Measurement of Highway Benefits. *Journal of Political Economy* 69, 236–249.

H. Mohring and M. Harwitz, 1962, *Highway Benefits.* Northwestern University Press, Evanston.

T. Mori, 1997, A Modelling of Megalopolis Formation: The Maturing of City Systems. *Journal of Urban Economics* 42, 133–157.

R. F. Muth, 1961, The Spatial Structure of the Housing Market, *Papers and Proceedings of the Regional Science Association* 7, 207–220.

R. F. Muth, 1969, *Cities and Housing.* The University of Chicago Press, Chicago.

H. Ogawa and M. Fujita, 1980, Equilibrium Land Use Pattern in a Non–Monocentric City. *Journal of Regional Science* 20, 455–475.

M. Olson Jr., 1969, The Principle of 'Fiscal Equivalence': The Division of Responsibility Among Different Levels of Government. *American Economic Review, Papers and Proceedings* 59, 479–487.

Y. Y. Papageorgiou, 1971, The Population Density and Rent Distribution Models Within a Multicentre Framework. *Environment and Planning* 3, 267–282.

Y. Y. Papageorgiou and T. R. Smith, 1983, Agglomeration as Local Instability of Spatially Uniform Steady–States. *Econometrica* 51, 1109–1119.

C. Ponsard, 1958, *Histoire des Théories Economiques Spatiales.* Translated in English as *History of Spatial Economic Theory* by B. H. Stevens. Springer–Verlag, Berlin, 1983.

J. Rawls, 1971, *A Theory of Justice.* Harvard University Press, Cambridge, Mass.

D. Ricardo, 1821, *Principles of Political Economy and Taxation.* Reprinted by George Bell & Sons Ltd, London, 1891.

S. A. Scotchmer and M. H. Wooders, 1987, Competitive Equilibrium and the Core in Club Economies with Anonymous Crowding. *Journal of Public Economics* 34, 159–174.

A. Smith, 1776, *The Wealth of Nations.* Reprinted by the Modern Library, New York, 1965.

R. M. Solow, 1973*a*, On Equilibrium Models of Urban Location. In M. Parkin and A. R. Nobay (eds) *Essays in Modern Economics*, Longman, London, 2–16.

R. M. Solow, 1973*b*, Congestion Cost and the Use of Land for Streets. *Bell Journal of Economics and Management Science* 4, 602–618.

R. M. Solow and W. S. Vickrey, 1971, Land Use in a Long Narrow City. *Journal of Economic Theory* 3, 430–447.

J. Steuart, 1767, *An Inquiry into the Principles of Political Oeconomy.* Reprinted by Oliver and Boyd, Edinburgh, 1966.

J. Stewart, 1947, Suggested Principles of Social Physics. *Science*, 179–180.

J. E. Stiglitz, 1977, The Theory of Local Public Goods. In M. S. Feldstein and R. P. Inman (eds) *The Economics of Public Services*, Macmillan, London, 274-333.

R. H. Strotz, 1965, Urban Transportation Parables. In J. Margolis (ed.) *The Public Economy of Urban Communities.* Resources for the Future, Washington D.C., 127–169.

T. Tabuchi, forthcoming, Urban Agglomeration and Dispersion: A Synthesis of Alonso and Krugman. *Journal of Urban Economics.*

R. Thom, 1972, *Stabilité Structurelle et Morphogénèse: Essai d' Une Théorie Générale des Modèles.* Translated in English as *Structural Stability and Morphogenesis.* W. A. Benjamin, Reading, Mass., 1975.

C. M. Tiebout, 1956, A Pure Theory of Local Public Goods. *Journal of Political Economy* 64, 416-24.

G. Tolley and J. Grihfield, 1987, City Size and Place as Policy Issue. In E. S. Mills (ed.) *Handbook of Regional and Urban Economics* II*: Urban Economics.* North–Holland, Amsterdam, 1285-1311.

J. H. von Thünen, 1826, *Der Isolierte Staadt in Beziehung auf Landwirtschaft und Nationalöconomie.* Translated in English as *von Thünen's Isolated State* by C. W. Wartenberg. Pergamon Press, Oxford, 1966.

A. A. Walters, 1961, The Theory and Measurement of Private and Social Cost of Highway Congestion. *Econometrica* 29, 676–699.

A. Weber, 1909, *Über den Standort der Industrien.* Translated in English as *Alfred Weber's Theory of Location of Industries* by C. J. Friedrich. University of Chicago Press, Chicago, 1957.

W. C. Wheaton, 1974, Linear Programming and Locational Equilibrium: the Herbert–Stevens Model Revisited. *Journal of Urban Economics* 1, 278–287.

M. J. White, 1976, Firm Suburbanization and Urban Subcenters. *Journal of Urban Economics* 3, 323–343.

M. J. White, 1978, Location Choice and Commuting in Cities with Decentralized Employment. *Journal of Urban Economics* 24, 129–152.

J. D. Wilson, 1987, Trade in a Tiebout Economy. *American Economic Review* 77, 431–441.

L. Wingo, Jr., 1961, *Transportation and Urban Land.* Resources for the Future, Washington D.C.

G. K. Zipf, 1949, *Human Behavior and the Principle of Least Effort.* Addison–Wesley, Reading, Mass.

Part I

The Monocentric City

2
Urban Equilibrium

We begin with a simple model which we use to discuss a competitive allocation within a city of identical individuals. In this chapter, decentralisation is complete: everyone, within the framework of existing property rights, maximises either utility or profit. The urban infrastructure, which includes transportation, improved land and a public good, is determined exogenously. We investigate the characteristics of the allocation. In chapter three, we examine how these characteristics are affected by exogenous factors. In later chapters, we extend the model in several directions and we revise some of the results presented here.

2.1 Conceptual Framework

2.1.1 Landscape and the Shape of the Urban Area

We assume an unbounded landscape which can be represented by $\mathbb{R}^2$. There are two categories of land, *available* and *unavailable*. Available land can be either agricultural or urban.

There is a single *centre* on the landscape which is adjacent to available land. Available land is distributed around the centre so that a straight line segment connecting the centre with any available location passes entirely through available land. To fix ideas, unavailable land can exist as an open sector relative to the centre, but not as a concentric zone. The accessibility of any available location relative to the centre is simply determined by the Euclidean distance between the two. Since there are no locational features other than accessibility to the centre, available locations at the same *distance* x from the centre are indistinguishable. This is a crucial simplification, allowing us to treat space as if it were single-dimensional. For this reason, x is used throughout to represent both distance from the centre and location.

Production and distribution of private urban goods and services occurs only at the centre.[1] There is a fixed *total population* $\bar{N}$ of individuals who are regularly associated with the activities at the centre—the urban population. Those individuals travel to the centre using a radial, dense, non-congestible transportation network. No other type of spatial interaction occurs. Urban residents occupy the *urban area* $\mathcal{X} \subset \mathbb{R}^2$ and the remaining available land is used for agriculture.

Our entire analysis will be confined to the city. The surrounding non-urban land is only relevant inasmuch as it represents a possibility for converting it to urban. The amount of *available land* within a particular distance from the centre is denoted by $\Theta\left[x, \bar{\beta}\right]$ where $\bar{\beta}$ is a shift parameter representing the possibility of converting (reclaiming) unavailable land to available. If we define $\theta \equiv \partial\Theta/\partial x$, θdx is the amount of available land on a ring of width dx at distance x from the centre. Since available land is connected, $0 < \theta \leq 2\pi x$. As long as there still remains some unavailable land, the available land increases with increasing $\bar{\beta}$.

> **Assumption 2.1:** The available land function $\Theta[x, \bar{\beta}]$ is differentiable, increasing with x and non-decreasing with $\bar{\beta}$; $\Theta\langle 0\rangle = 0$; if $\theta\langle x'\rangle \equiv \partial\Theta\langle x'\rangle/\partial x = 2\pi x'$ for $x' \in \mathcal{X}$ then $\partial\theta\langle x'\rangle/\partial\bar{\beta} = 0$; and if $\theta\langle x''\rangle < 2\pi x''$ for $x'' \in \mathcal{X}$ then $\partial\theta/\partial\bar{\beta} > 0$.

Accessibility in our model matters because an individual living at distance x from the centre spends on transportation costs for travelling to the centre where he or she interacts with other individuals. This includes working, shopping, recreation, and other social activities. Transportation costs are represented by a known function $T\left[x, \bar{\alpha}\right]$ which depends on x and the quality of transportation infrastructure captured by a shift parameter $\bar{\alpha}$. Improvements in the quality of transportation infrastructure lower both the cost of transportation and the *transportation rate* $t \equiv \partial T/\partial x$.

> **Assumption 2.2:** The transportation cost function $T[x, \bar{\alpha}]$ is differentiable, a strictly increasing and unbounded function of x, and a strictly decreasing function of $\bar{\alpha}$; $\partial t/\partial\bar{\alpha} \equiv \partial(\partial T/\partial x)/\partial\bar{\alpha} < 0$.

Having defined the available land and transportation cost, the *transportation cost shape of the city* $\tilde{\Theta}\left[T, \bar{\alpha}, \bar{\beta}\right]$ can also be derived.[2] This function specifies

[1] In the simplest case, which is discussed in this chapter, the centre is represented by a single point.

[2] Using $T = T[x, \bar{\alpha}]$, we may write $x = T^{-1}[T, \bar{\alpha}]$ because, given $\bar{\alpha}$, the relationship between distance from the centre and transportation costs is one-to-one. Thus $\Theta[x, \bar{\beta}] = \Theta[T^{-1}[T, \bar{\alpha}], \bar{\beta}] \equiv \tilde{\Theta}[T, \bar{\alpha}, \bar{\beta}]$.

the total amount of available land for which travel costs are less than or equal to T (see Arnott and Stiglitz (1981)). The transportation cost shape of the city determines some important characteristics of the equilibrium allocation and the comparative statics thereof. In particular, its derivative

$$\begin{aligned} \sigma\left[x,\bar{\alpha},\bar{\beta}\right] &\equiv \frac{\partial\tilde{\Theta}}{\partial T} \\ &= \frac{\theta}{t} \end{aligned} \tag{2.1}$$

and its elasticity

$$\eta_{\tilde{\Theta}:T} = \frac{\eta_{T:x}}{\eta_{\Theta:x}}\ , \tag{2.2}$$

where $\eta_{T:x}$ is the *elasticity of transportation cost with respect to distance* and $\eta_{\Theta:x}$ is the *elasticity of the urban area with respect to distance,* will be particularly useful in the ensuing analysis. We impose

Assumption 2.3: $\dfrac{\partial\sigma}{\partial x} = \dfrac{\partial}{\partial x}\left(\dfrac{\theta}{t}\right) > 0.$

Notice that this assumption is consistent with $\partial\theta/\partial x \geq 0$ and $\partial t/\partial x \leq 0$, where at least one holds with strict inequality. The former only requires that the amount of urban land available at a distance increases away from the centre. The latter will be justified later on by establishing that transport congestion declines as one moves away from the city centre. Since both seem reasonable, and since they are only sufficient for assumption 2.3, this assumption seems also plausible.

2.1.2 Preferences

Preferences are defined over the consumption of a *composite good* Z, *housing* H and a *public good* $\bar{\gamma}$. In the simplest case, housing is produced with land only. Since individuals are identical, their preferences are represented by the same *utility function* $u\left[Z,H,\bar{\gamma}\right]$. We denote by U the corresponding *utility level*, $U = u\left[Z,H,\bar{\gamma}\right]$.

Assumption 2.4: The utility function $u[Z,H,\bar{\gamma}]$ is differentiable, strictly increasing and strictly quasi–concave in its arguments. Commodities Z and H are normal, and their *marginal rate of substitution of Z for H* MRS$[Z,H]$ is such that $\lim_{H\to\infty}$ MRS$[Z,H] = 0$ and $\lim_{H\to 0}$ MRS$[Z,H] = \infty$.

The restrictions placed on MRS, the Inada (1963) conditions, ensure that a consumption bundle which includes positive amounts for both commodities is always preferable to a bundle which includes a finite positive quantity for only one commodity.

2.1.3 Prices

The composite good serves as the numéraire commodity. As mentioned in section 2.1.1, it is available only at the centre. The *price of land* $R[x]$ is defined for every location. Both are taken parametrically by the individuals, *i.e.* every individual believes that none of his or her actions can change the observed pattern of prices and the cost of transportation.

2.2 The Structure of Urban Equilibrium

2.2.1 Land Ownership, Property Rights, and the Disposal of Urban Rent

We impose three requirements on our model concerning land: (1) a structure of land markets which allows for complete decentralisation, that is, one where everyone, within the framework of existing property rights, is either a utility-maximising individual or a profit-maximising firm; (2) that the use of land at any location is determined by the highest bid there; and (3) that our model allows for a variety of possibilities for the disposal of urban rent does exist, ranging from complete redistribution within the city to complete transfer outside of the city. In general, whether these requirements are realised or not, crucially depends upon existing property rights which include, but may not be limited to, the original distribution of land ownership. In the literature, this dependence is not always sufficiently elaborated. It will prove useful to review some of these approaches before suggesting an alternative which satisfies all our requirements.

A standard approach, first proposed by Alonso (1964), calls for numerous absentee landlords, each controlling only a small parcel of land in a single locality. These landlords confront numerous farmers and urban residents who bid for land in different locations. Parcels of land are supplied to the highest bidder. An alternative is to propose that all the land is originally owned by a single absentee landlord who behaves like a monopoly. The landlord, facing as before numerous farmers and urban residents, behaves so as to maximise the total revenue from land (see Markusen and Scheffman (1978)). The element shared by the above two approaches is that all urban rents necessarily vanish from the city, that is, they violate our third requirement. In addition, the second approach admits the possibility that land be withheld from urban development, that is, it violates our second requirement.

At the other end of the spectrum, there is a group of models in which all urban rents stay within the city (see, for example, Oron, Pines and Sheshinski (1973), Solow (1973), Kanemoto (1980), and Pines and Sadka (1986)). These models violate once again our third requirement. Moreover, when there is a single agent who controls the supply of urban land, the structure of property rights is not consistent with full decentralisation. For example, in Kanemoto (1980), there is a city government which buys the land from the farmers at a competitive price. Although the government controls all the supply of urban land, the residential land-market is competitive. This claim contradicts our first requirement, that agents be either utility-maximising individuals or profit-maximising firms. For if someone controls all the land then, under full decentralisation, he or she should behave as in Markusen and Scheffman (1978) rather than competitively.

We can satisfy all three requirements using the following construct. Suppose that the land is originally held by numerous farmers whose property rights are limited to agricultural use. Also suppose that there are numerous development corporations, each having an exclusive right to develop a specific small area, say, a narrow ring at a given distance from the centre. Every corporation purchases all the land from the farmers in its ring at a competitive price. Since any particular corporation controls only a small portion of the land, it cannot exert market power in its transactions with prospective urban residents because arbitrarily close substitutes are available in adjacent rings. Under these circumstances, we can admit various ways of urban rent disposal by assuming various ownership patterns concerning the development corporations. On the one hand, if corporations are exclusively owned by absentee investors, we obtain the case of Alonso (1964). If, on the other hand, they are exclusively owned by the city residents, we obtain the case of Solow (1973) etc. In–between, the corporations are jointly owned by the city residents and the absentee investors at varying proportions.

2.2.2 Individual Decisions

The *income* $\bar{Y}$ of every urban resident consists of an *initial endowment* of composite good $\bar{\Omega}$ and the share in the profit from land transactions discussed in section 2.2.1. Given their income, individuals in the city maximise their welfare in two stages: firstly, they determine the maximum feasible utility level for every location; secondly, they compare all those localised maxima and choose a location for which feasible utility is a global maximum.

Consumption Bundle

An individual located at distance x from the centre selects a feasible consumption bundle $(Z[x], H[x])$ that maximises his or her utility level at x. Formally, the problem can be expressed as

$$V \equiv \max_{Z,H} u[Z, H, \bar{\gamma}] \text{ subject to } Z + RH \leq \bar{Y} - T, \qquad (2.3)$$

where V is the *maximum feasible utility level.* The necessary conditions for this maximisation problem are

$$\begin{aligned} R &= \text{MRS}\left[Z^i, H^i\right] \quad (a) \\ Z^i + RH^i &= \bar{Y} - T, \qquad\quad (b) \end{aligned} \tag{2.4}$$

where superscript i denotes a solution to the problem of the individual. These conditions are also sufficient because the Inada restrictions in assumption 2.4 preclude corner solutions.

An alternative representation of the system (2.3) and (2.4), which we shall use extensively from now on, can be obtained in two steps.[3] The first step involves using $u[Z, H, \bar{\gamma}]$ and the first condition in (2.4) to solve for Z^i and H^i in terms of R, U and $\bar{\gamma}$ as

$$Z^i = z[R, U, \bar{\gamma}] \text{ and } H^i = h[R, U, \bar{\gamma}], \tag{2.5}$$

where z and h are the *compensated demand functions.* Upon substitution of (2.5) into the second condition of (2.4) we obtain the *minimum expenditure function* equated to the disposable income:

$$\begin{aligned} e[R, U, \bar{\gamma}] &\equiv z[R, U, \bar{\gamma}] + Rh[R, U, \bar{\gamma}] \\ &= \bar{Y} - T. \end{aligned} \tag{2.6}$$

The second step involves solving (2.6) for the maximum feasible utility level. Thus (2.3) and (2.4) can be replaced by (2.5) and (2.6).

Location

The solution of (2.6) for U, given the parameters $R, \bar{\gamma}, \overline{Y}$, and T, is the maximum feasible utility $V = v\left[R, \bar{Y} - T, \bar{\gamma}\right]$ at any location, where $v[\cdot]$ denotes the *indirect utility function.* What remains is to compare all those localised maxima and choose a location which maximises V. By definition, the urban area $\mathcal{X}$ is composed of such locations. It follows that there exists a constant $\bar{U}$ such that

$$\max_x v\left[R, \bar{Y} - T, \bar{\gamma}\right] = \bar{U} \text{ for } x \in \mathcal{X}. \tag{2.7}$$

Substitution of (2.7) into (2.6) gives

$$e\left[R, \bar{U}, \bar{\gamma}\right] = \bar{Y} - T \text{ for } x \in \mathcal{X} \tag{2.8}$$

The implication that the attainable utility level is uniform across the urban area is central to the concept of spatial equilibrium.

[3] A comparison of alternative representations for the individual's choice problem appears in the appendix (section 2.4.1).

2.2.3 Clearing the Urban Land Market

For $x \in \mathcal{X}$, $N[x]$ denotes *the number of individuals living within distance x* from the centre. If we define $n \equiv dN/dx$, ndx denotes *the number of individuals located at distance x* from the centre within the ring $[x, x+dx]$. Thus, equality between the demand for urban land at distance x from the centre and the corresponding supply $d\Theta$ is given by

$$nH = \theta \text{ for } x \in \mathcal{X}. \tag{2.9}$$

On the aggregate level, all $\bar{N}$ individuals must occupy land in the city. This happens because, as we mentioned in section 2.1.2, someone who does not occupy land does not maximise utility subject to his or her income constraint. Therefore we have

$$N\langle\bar{x}\rangle = \int_{\mathcal{X}} n\, dx = \bar{N}. \tag{2.10}$$

where $\bar{x}$ is the *maximum distance* of an urban location from the centre. For the same reason, the urban area $\mathcal{X}$ cannot be empty.[4]

2.2.4 Clearing the Available Land Market

Following the discussion at the end of section 2.2.1, development corporations purchase land from farmers at a price equal to the marginal productivity of land in agriculture. This represents the cost of developing available land for urban use. We assume that the marginal productivity of land in agriculture is constant.

> **Assumption 2.5:** The *cost of developing a unit of land* is a constant $\bar{R}$ up to the supply constraint θ.

Under these conditions, the supply of urban land is represented by figure 2.1.

Assumption 2.5, together with the profit–maximising development corporations, ascertains that available land is developed at x only if $R[x] \geq \bar{R}$. Moreover, given the utility–maximising behaviour of individuals, the price of urban land at the border $\bar{x}$ must equal the cost of development. If not, $R\langle\bar{x}\rangle > \bar{R}$ and the continuity of the transportation cost function imply that an individual can achieve $\bar{U}$ at lower cost than $\bar{Y}$ if he or she locates at $\bar{x} + \epsilon$ where $\epsilon > 0$ is sufficiently small. In summary,

$$R \geq \bar{R} \text{ for } x \in \mathcal{X} \text{ and } R\langle\bar{x}\rangle = \bar{R}. \tag{2.11}$$

[4] The reader should notice that we treat individuals as a continuous mass distributed over a continuous space. This is essential for the use of calculus in our analysis, but it has raised a controversy which strikes at the roots of urban economics. For details on this controversy and for a justification of continuous modelling in urban economics see section 2.4.2 in the appendix.

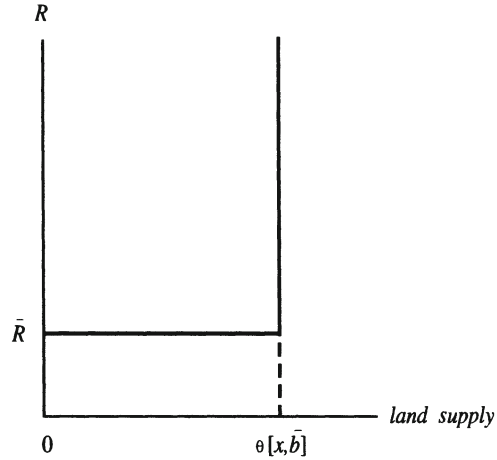

FIGURE 2.1. The Supply of Urban Land.

2.2.5 Rent Disposal and Income Determination

Let $\bar{k} \in [0,1]$ be *the relative share of the urban population in the aggregate profit from urban land transactions,* which corresponds to the relative share of the urban population in the ownership of the development corporations. As we have noted in section 2.2.2, the income $\bar{Y}$ of a representative urban resident arises from an initial endowment of composite good and the share in the profit from urban land transactions. Therefore

$$\begin{aligned} \bar{Y} &= \bar{\Omega} + \frac{\bar{k}}{\bar{N}}\overline{\text{DLR}} \qquad (a) \\ \overline{\text{DLR}} &\equiv \int_{\mathcal{X}} \theta(R-\bar{R})\,dx \qquad (b) \end{aligned} \tag{2.12}$$

where $\overline{\text{DLR}}$, the aggregate *differential land rent*, represents the total value of urban land net of its development cost.[5] When the entire $\overline{\text{DLR}}$ is taken out of the city $(\bar{k}=0)$ we have a *renter city* model; when only some of the $\overline{\text{DLR}}$ is taken out $(\bar{k} \in (0,1))$, we have a *mixed city* model; and when all $\overline{\text{DLR}}$ stays in the city $(\bar{k}=1)$, we have an *owner city* model.

[5] DLR arises from the difference between urban and agricultural rents at each location in the city. This difference is a manifestation of pure agglomeration advantages. The concept has a relatively long history. Marshall (1890), for example, "...emphasizes the importance of location within the city, and defines 'situation value' as the sum of money values of the situation advantages of a site. According to Marshall,'site value', the price a site would fetch if cleared of buildings and sold in the free market, is equal to situation value plus agricultural rent." (Alonso (1964, p.4). See also Book V Chapter XI.) Clearly, the 'situation value' gives rise to $\overline{\text{DLR}}$. In the modern literature, the concept has been given prominence by Arnott and Stiglitz (1981).

2.2.6 Definition of a Competitive Equilibrium

We are now ready to formulate the conditions for a competitive equilibrium in a concise way. Namely, for $x \in \mathcal{X}^e$, $\left(n^e, R^e, \bar{Y}^e, \bar{N}^e, \bar{U}^e, \bar{x}^e\right)$ is a *competitive equilibrium* if and only if

$$\begin{aligned}
e\left[R^e, \bar{U}^e, \bar{\gamma}\right] &\underset{(2.8)}{=} \bar{Y}^e - T && (a)\\
n^e h\left[R^e, \bar{U}^e, \bar{\gamma}\right] &\underset{(2.9)}{=} \theta && (b)\\
\int_{\mathcal{X}^e} n^e\, dx &\underset{(2.10)}{=} \bar{N}^e && (c)\\
R^e\left\langle \bar{x}^e \right\rangle &\underset{(2.11)}{=} \bar{R} && (d)\\
\bar{Y}^e &\underset{(2.12)}{=} \bar{\Omega} + \frac{\bar{k}}{\bar{N}} \int_{\mathcal{X}^e} \theta(R^e - \bar{R})\, dx. && (e)
\end{aligned} \qquad (2.13)$$

In this definition, z, h, T and θ are known functions, and $\bar{\Omega}$, $\bar{R}, \bar{k}$ and $\bar{\gamma}$ are given parameters.[6]

One may wonder whether the above definition of an equilibrium is consistent with the resource constraint. The answer is that this constraint is satisfied by the Walras law.[7]

---

[6] We omit the first condition in (2.11) because, as we establish in section 2.3.2, it is always satisfied in equilibrium.

[7] This can be established by using the income equation to derive

$$\begin{aligned}
\bar{N}\bar{\Omega} &= \bar{N}\bar{Y} - \bar{k}\overline{\text{DLR}}\\
&= \bar{N}\bar{Y} - \overline{\text{DLR}} + (1-\bar{k})\overline{\text{DLR}}\\
&\underset{(2.10)}{=} \int_{\mathcal{X}} n\bar{Y} dx - \int_{\mathcal{X}} \theta\left(R - \bar{R}\right) dx + (1-\bar{k})\overline{\text{DLR}}\\
&\underset{(2.10,11)}{=} \int_{\mathcal{X}} n\left(Z + RH + T\right) dx\\
&\quad - \int_{\mathcal{X}} \left(nRH - \theta\bar{R}\right) dx + (1-\bar{k})\overline{\text{DLR}}\\
&= \int_{\mathcal{X}} n\left(Z + T\right) dx + \Theta\left\langle \bar{x} \right\rangle \bar{R} + (1-\bar{k})\overline{\text{DLR}}
\end{aligned}$$

that is, total endowment in equilibrium must equal the total amount needed for the consumption of the composite good, transportation, the cost of urban development and the share of absentee developers in the profit from land transactions.

2.2.7 Closing the Model

Notice that the definition of an equilibrium (2.13) requires determining two spatial distributions, n^e and R^e, and four variables, $\bar{Y}^e$, $\bar{N}^e$, $\bar{U}^e$ and $\bar{x}^e$. However, conditions (2.13) involve two equations defined at each distance from the centre and only three single equations. The equilibrium is therefore underdetermined. We can either propose a new single equation, or reduce the number of variables by one. It turns out that the latter approach, first proposed by Wheaton (1974), can provide us with significant predictions about comparative statics of urban structure during the early and late stages of urban development.

Consider an isolated region of fixed total population $\overline{\mathbb{N}}$ partitioned between urban $(\bar{N})$ and agricultural $(\overline{\mathbb{N}} - \bar{N})$. The urban population is accommodated within a single city. If migration is costless, an equilibrium for the entire region implies that individuals in both sectors enjoy the same utility level. Now suppose that an external change alters the urban utility level. As long as the city population is sufficiently small relative to the total, any such urban utility change will be absorbed through migration between the two sectors. Since the city is insignificant relative to the region, changes in city size due to migration cannot have a measurable effect on the regional equilibrium utility level. Under these circumstances, which describe early stages of urbanisation, urban equilibrium utility remains constant $(\bar{U}^e = \bar{U})$ as urban equilibrium population varies. If we imagine that the city is the spearhead of innovation in the region, so that technological developments through time tend to favour the city, urban population will continue to grow. Eventually, its size will become significant relative to the region in the sense that external changes can affect both equilibrium population and utility. At the other extreme, during late stages of urbanisation, urban population in the region reaches an upper limit, $\bar{N}^e = \bar{N}$, which remains fixed as urban equilibrium utility varies.[8]

Following Wheaton (1974), let us now translate these observations in the context of section 2.2.6. During early stages of urbanisation, we can treat the equilibrium utility level as constant. Therefore $(n^e, R^e, \bar{Y}^e, \bar{N}^e, \bar{x}^e)$ is a competitive equilibrium if and only if it satisfies (2.13) for known parameters $\bar{\Omega}, \bar{R}, \bar{U}, \bar{k}, \bar{\alpha}, \bar{\beta}$ and $\bar{\gamma}$. This represents an *open city* model. On the other hand, during late stages of urbanisation, we can treat the equilibrium population size as constant. Therefore $(n^e, R^e, \bar{Y}^e, \bar{U}^e, \bar{x}^e)$ is a competitive equilibrium if and only if it satisfies (2.13) for known parameters $\bar{\Omega}, \bar{R}, \bar{N}, \bar{k}, \bar{\alpha}, \bar{\beta}$ and $\bar{\gamma}$. This represents an *closed city* model. Since in both the open and closed models of city we have reduced the set of unknowns by one variable, (2.13) determines equilibrium during the early and late stages of urbanisation.

---

[8] In technologically advanced countries, such as Canada, this limit represents about ninety-five percent of the total population. Evidently, migration cannot any longer play a significant rôle in affecting the equilibrium utility level. On the one hand, further migration toward the city is insignificant because of the small agricultural population size. On the other hand, migration away from the city is unlikely because, during late stages of urbanisation, there is no real alternative to urban life.

2.3 Properties of the Equilibrium Allocation

In sections 2.2.5 and 2.2.7, we have classified cities according to their patterns of urban land ownership and population closure respectively. Following this classification, there are in principle 3×2 city types ((renter, mixed, owner city)$\times$(open, closed city)). As we shall see in chapter three, different city types may respond differently to changes in external conditions. However, all city types have the same urban equilibrium structure which is described in this section.

2.3.1 The Shape of the Urban Area

The urban area in equilibrium is bounded because transportation cost is boundless by assumption 2.2 and total resources in the economy, $\bar{N}\bar{\Omega}$, are finite. We next show that the urban area is in one piece. If x' is a location that belongs to $\mathcal{X}^e$, we know from section 2.1.1 that a straight line segment connecting the centre with x' passes entirely through available land. We claim that, in addition, the line segment passes entirely through urban land, *i.e.* for all $x \in [0, x')$, $V[x] = \bar{U}^e$. Let E denote the *minimum expenditure level*, and let $E^e = e\left[R^e, \bar{U}^e, \bar{\gamma}\right]$. Suppose that the line segment under consideration does not pass entirely through urban land in equilibrium. Then there exists $x'' \in [0, x')$ such that $E^e \langle x'' \rangle + T\langle x'' \rangle > E^e \langle x' \rangle + T\langle x' \rangle$, that is, $\bar{U}^e$ is not feasible for some $x'' < x'$. Since, by assumption 2.2, $T\langle x'' \rangle < T\langle x' \rangle$, it follows that $E^e \langle x'' \rangle > E^e \langle x' \rangle$. Hence, by the first condition in (2.15), $R^e \langle x'' \rangle > R^e \langle x' \rangle$. But since $x' \in \mathcal{X}^e$ and, therefore, $R^e \langle x' \rangle \geq \bar{R}$ by (2.11), it must be true that $R^e \langle x'' \rangle > \bar{R}$. Under these circumstances, the development corporation at x'' would forego profits—a contradiction. In consequence, all land available between the centre and x' is developed. Letting $x' = \bar{x}^e$, all the land available between the centre and the border is developed in equilibrium, and this implies a connected urban area. We can therefore state

> **Result 2.1:** The urban area is bounded and connected at equilibrium.

Finally notice that, since all land available between the centre and any urban location is urban land, agricultural land cannot exist within the urban area in equilibrium.

2.3.2 The Principle of Zero Marginal Location Costs

In this section we discuss *location costs* in equilibrium, i.e. the sum of housing and transportation costs. We know that, in equilibrium, urban land values and transportation costs are defined over $[0, \bar{x}^e]$. Differentiability of the transportation cost function, together with the properties of an expenditure function, ensure that R^e is also differentiable. Hence, we can differentiate (2.8) in the interior of $\mathcal{X}^e$ to obtain

$$\frac{dE^e}{dx} + t = \frac{\partial e^e}{\partial R}\frac{dR^e}{dx} + t = 0. \tag{2.14}$$

Since R is a parameter in $e[\cdot]$, which minimises expenditure, we can apply the envelope theorem to obtain

$$\frac{\partial e}{\partial R} = h[R, U, \bar{\gamma}]. \tag{2.15}$$

Substituting (2.15) into (2.14) we get one of the best known and most useful results in urban economics. Namely, we have

Result 2.2 (Muth (1961)): $H^e \dfrac{dR^e}{dx} + t = 0.$

Since the first term in result 2.2 is the change in housing cost caused by a marginal move away from the centre, while the second term is the corresponding change in transportation cost, the condition of Muth implies that, in equilibrium, changes in housing cost evaluated at the optimal consumption and associated with short movements around a location in the interior of the urban area are balanced by the corresponding changes in transportation cost. Moreover, since transportation costs increase with distance by assumption 2.2, the transportation rate in result 2.2 represents marginal costs of moving further away from the centre. Then it must be that marginal benefits in result 2.2 are represented by the change in housing cost.

Using result 2.2, we have

$$\begin{aligned} \frac{dR^e}{dx} &= -\frac{t}{H^e} && (a) \\ \frac{d^2R^e}{dx^2} &= -\frac{1}{H^e}\left(\frac{\partial h^e}{\partial R}\left(\frac{dR^e}{dx}\right)^2 + \frac{\partial t}{\partial x}\right). && (b) \end{aligned} \tag{2.16}$$

Since $\partial h^e/\partial R$ is the substitution effect, it is negative. Thus the shape of the rent depends upon how the transportation rate varies with respect to distance. In the real world, where traffic density declines with distance from the centre, the transportation rate also declines.[9] In consequence, using assumption 2.2 and (2.16),

Result 2.3 (Muth (1961)): $\dfrac{dR^e}{dx} < 0$; if $\dfrac{\partial t}{\partial x} \leq 0$ then $\dfrac{d^2R^e}{dx^2} > 0.$

[9] This property is derived later on as result 7.1 of section 7.1, where the transportation cost function becomes endogenous.

The spatial distribution of equilibrium rent implied by result 2.3 is consistent with empirical analyses which fit urban land value data using specific functions, such as the negative exponential to be discussed in section 2.3.4 below (see, for example, Mills (1969)).

Differentiating (2.5) with respect to distance, and taking into account result 2.3, we have

$$\frac{dH^e}{dx} = \frac{\partial h^e}{\partial R}\frac{dR^e}{dx} > 0. \tag{2.17}$$

Furthermore, since

$$\frac{d\bar{U}^e}{dx} = \frac{\partial u^e}{\partial Z}\frac{dZ^e}{dx} + \frac{\partial u^e}{\partial H}\frac{dH^e}{dx} = 0, \tag{2.18}$$

we conclude that

$$\frac{dZ^e}{dx} < 0 \tag{2.19}$$

because of assumption 2.4 and (2.17). Finally, decreasing equilibrium consumption of the composite good away from the centre implies

Result 2.4: Location costs increase away from the centre at equilibrium.

This result contradicts the belief, long held by land economists and explicitly stated in Haig (1926, p. 421), that "rent appears as the charge which the owner of a relatively accessible site can impose because of the saving in transport costs which the use of his site makes possible." Early modern treatments on the subject, such as Mohring (1961) and Wingo (1961), adopted this traditional view which, further to location costs being constant for marginal change by result 2.2, it requires that they be globally constant.[10]

2.3.3 Population Density

By definition, population density D is the inverse of land per capita. Therefore, using (2.17), we have

$$\frac{dD^e}{dx} = -\frac{1}{H^{e2}}\frac{\partial h^e}{\partial R}\frac{dR^e}{dx}. \tag{2.20}$$

Recalling that the substitution effect is negative, and that the same is true for the change in equilibrium rents over distance, we obtain

Result 2.5 (Muth (1961)): $\dfrac{dD^e}{dx} < 0.$

[10]This requirement can be satisfied in the limiting case where the elasticity of substitution between housing and the composite good is zero. For an example, see section 2.3.5. This case is not included in result 2.4 because, by assumption 2.4, the utility function is differentiable—which implies nonzero price elasticity of the compensated demand.

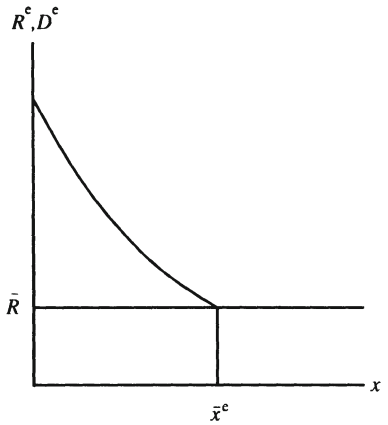

FIGURE 2.2. Equilibrium Rent and Density Gradients.

We cannot sign the second derivative of density with respect to distance as we did in the case of rent. However, empirical evidence presented in the following section 2.3.4 indicates that the spatial distribution of urban population density is similar to the distribution of urban land rent. A sufficient condition for this requirement can be obtained by using (2.20) to derive

$$\left(\frac{dD^e}{dx}\right)/D^e = -\eta_{h:R^e}\left(\frac{dR^e}{dx}\right)/R^e, \tag{2.21}$$

where $\eta_{h:R}$ is the *price elasticity of the compensated demand for housing.* We conclude that, if $\eta_{h:R}$ is constant, the spatial distribution of equilibrium population density will indeed be similar to the corresponding distribution of equilibrium rent. Under these conditions, and taking into account result 2.3, figure 2.2 describes the spatial distribution of both equilibrium land rent and population density.

2.3.4 Negative Exponential Rent and Density

The spatial structure of urban population density has been the subject of considerable empirical analysis. This began with Bleicher (1892), who collected data on the spatial distribution of population in Frankfurt am Main and observed the shape of the population density gradient. His observation was forgotten and rediscovered independently by Stewart (1947) and Clark (1951). Although both these authors proposed the negative exponential as a descriptor of urban population density, it became known as the "Clark formula" probably because Clark was the first one to write it down and test it explicitly. The Clark formula has been established as representing one of the strongest empirical regularities in urban spatial analysis. Here we examine how this specific empirical form can arise within our more general framework and we present a brief overview of the associated empirical literature.

Muth (1961) has derived an exponential density function for land values by assuming that the price elasticity of the compensated demand for housing is unitary.[11] Following Papageorgiou and Pines (1989), and using (2.15), if the expenditure function yields a unitary price elasticity of the compensated demand for housing then it must satisfy

$$\frac{\partial^2 e^e}{\partial R^2}\frac{R^e}{\partial e/\partial R^e} = \frac{\partial h^e}{\partial R}\frac{R^e}{H^e} = 1. \tag{2.22}$$

The solution to the above second-order partial differential equation is

$$e[R^e, \bar{U}^e, \bar{\gamma}] = f[\bar{U}^e, \bar{\gamma}] \ln R^e + F[\bar{U}^e, \bar{\gamma}], \tag{2.23}$$

where f and F are some functions.[12] Using (2.15) and (2.23),

$$H^e = \frac{f[\bar{U}^e, \bar{\gamma}]}{R^e}. \tag{2.24}$$

This, and normality of land, imply $f[\cdot] > 0$ and $\partial f/\partial U > 0$.

Assuming that the transportation cost function is given by $T[x, \bar{\alpha}] = x/\bar{\alpha}$, it follows from (2.24) and result 2.2 that

$$\frac{1}{R^e}\frac{dR^e}{dx} = -\frac{1}{\bar{\alpha} f[\bar{U}^e, \bar{\gamma}]} \equiv -\bar{\delta}. \tag{2.25}$$

A solution of (2.25), in turn, is given by

$$R^e[x] = R^e\langle 0\rangle \exp\left(-\bar{\delta}x\right), \tag{2.26}$$

which is the negative exponential form used in the empirical literature. Finally, taking into account that population density is the inverse of land per capita and substituting (2.26) into (2.24) yields the Clark (1951) formula

$$\begin{aligned} D^e &= \frac{R^e\langle 0\rangle}{f[\bar{U}^e, \bar{\gamma}]} \exp\left(-\bar{\delta}x\right) \\ &= D^e\langle 0\rangle \exp\left(-\bar{\delta}x\right). \end{aligned} \tag{2.27}$$

---

[11] Other theoretical works concerning the exponential density function include Niedercorn (1971), Brueckner (1982), Anas and Kim (1992) and Anas, Arnott and Small (forthcoming).

[12] For example, the utility function

$$U = \frac{\bar{\gamma} Z}{\bar{A} - \bar{B}\ln H}$$

generates

$$e[\cdot] = \frac{\bar{B}U}{\bar{\gamma}} \ln R + U\left(\frac{(\bar{A} - \bar{B}\ln(\bar{B}U))}{\bar{\gamma}} + \bar{B}\right).$$

In a second example by Brueckner (1982), the utility function $U = \bar{A}Z + \bar{B}\ln H$ generates

$$e[\cdot] = \frac{\bar{B}}{\bar{A}} \ln R + \frac{\bar{B}}{\bar{A}}\left(1 - \ln\frac{\bar{B}}{\bar{A}}\right) + \frac{U}{\bar{A}}.$$

Observe that, since the price elasticity of the compensated demand for housing is unitary here, the rent and density gradients are the same by (2.21).

According to (2.27), the spatial distribution of population is empirically described by the two parameters $D^e\langle 0\rangle$ and $\bar{\delta}$ which can be estimated for any city. The parameter $\bar{\delta}$, representing the proportional rate at which population density declines away from the centre, is known in the literature as the *density gradient.* The model predicts that the density gradient must be positive, a fact consistently supported by evidence.[13] The density gradient is also useful as a summary measure of urban concentration or deconcentration. For if we connect the degree of urban sprawl by the average distance of individuals from the centre, $\bar{\delta}$ is inversely related to it.[14] This property of the density gradient is particularly useful for the empirical study of suburbanisation trends around the world, which will be discussed in connection with the comparative statics of chapter three.

Soon after Clark's (1951) seminal paper other authors contributed further empirical evidence in support of the negative exponential. Early studies include Kramer (1955), Muth (1961), Berry, Simmons and Tennant (1963), Clark (1967) and Casetti (1969). These workers estimated the linearised version of (2.27) for a large number of different cities using a sample of small–area density measurements (such as census tract data) on each city. An entirely different estimation approach for the study of urban population densities has been introduced by Mills (1972), which has proven influential since it requires a minimum of information: the areas of the entire city and its central part, as well as the population of the central city. Both approaches have been extensively used in the literature.[15]

2.3.5 Aggregate Relationships at Equilibrium

In this section we discuss a general relationship between aggregate differential land rent and aggregate urban transportation cost, due to Arnott and Stiglitz (1979), and how this relationship is affected by the transportation technology and the shape of the urban area at equilibrium. Considerable intuition can be gained on the subject if we motivate the discussion by using a special case first proposed by Mohring (1961). Let $u[Z,H,\bar{\gamma}] = f[\min\{Z/\bar{A}, H/\bar{B}\},\bar{\gamma}]$ with $\bar{A},\bar{B}$ parameters. For constant level of utility, $\partial z/\partial R = 0$ and $\partial h/\partial R = 0$, *i.e.* both the composite good and housing are consumed at fixed levels over the urban area in equilibrium. Hence this utility function implies zero elasticity of sub-

[13] There are some exceptions related to the phenomenon of population density 'craters' (Newling (1966), Latham and Yeates (1970)). This arises in the central part of some large metropolitan areas in which the quality of the environment has deteriorated to a sufficient degree.

[14] More specifically, with $\bar{R} = 0$, the degree of deconcentration is $2/\bar{\delta}$ (see Ashenfelter (1976) and White (1977)).

[15] An informative review of various population density studies is provided by McDonald (1989). For a more detailed methodological review of the two approaches see also White (1977).

stitution between housing and the composite good. Under these circumstances, the locational costs are constant not only for marginal change, as in result 2.2, but in the global sense also, that is, $\bar{H}^e R^e + T = \bar{H}^e \bar{R} + T\langle \bar{x}^e \rangle$ for $x \in \mathcal{X}^e$.

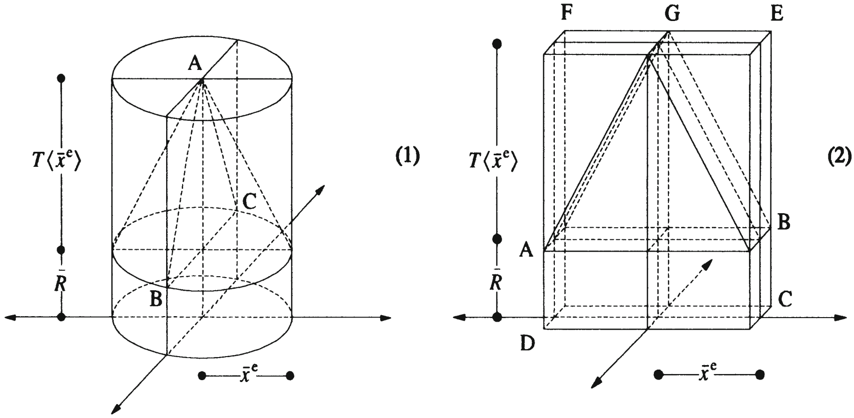

FIGURE 2.3. Aggregate Relationships When Location Cost is Globally Constant.

Let the transportation cost be linear with distance from the centre and the city be circular. Consider figure 2.3(1). The volume of the lower cylinder in this figure represents the *aggregate development cost* $\overline{\text{ADC}}^e$, the volume of the cone represents $\overline{\text{DLR}}^e$ and the volume of the upper cylinder, from which the cone is removed, represents the *aggregate transportation cost* $\overline{\text{ATC}}^e$. Since the volume of the upper cylinder is three times the volume of the cone, $\overline{\text{ATC}}^e = 2\overline{\text{DLR}}^e$. Clearly, when the transportation cost is concave(convex) with distance, $\overline{\text{ATC}}^e > (<) 2\overline{\text{DLR}}^e$. Now let the city be linear. In figure 2.3(2), the volume of the upper parallelepiped is only two times the volume of the prism and $\overline{\text{ATC}}^e = \overline{\text{DLR}}^e$. The same observations can be extended to cases where there are variable returns to scale in transportation. We conclude that the equilibrium relationship between aggregate transportation cost and differential land rent in this example depends on the shape of the city and on the technology of transportation.

We now verify these findings in a general context. Integrating by parts the differential land rent, and taking into account that the aggregate development cost $\overline{\text{ADC}}$ is given by $\Theta\langle\bar{x}\rangle\bar{R}$, we have

$$
\begin{aligned}
\overline{\text{DLR}}^e &= \int_{\mathcal{X}^e} \theta(R^e - \bar{R})dx \\
&= \overline{\text{ADC}}^e - \int_{\mathcal{X}^e} \Theta\frac{dR^e}{dx}dx - \overline{\text{ADC}}^e \\
&\underset{\text{(Result 2.2)}}{=} \int_{\mathcal{X}^e} \Theta\frac{t}{H^e}dx. \\
&\underset{(2.9)}{=} \int_{\mathcal{X}^e} \Theta\frac{n^e t}{\theta}dx \\
&= \int_{\mathcal{X}^e} \frac{\eta_{T:x}}{\eta_{\Theta:x}} n^e T dx \\
&\underset{(2.2)}{\equiv} \int_{\mathcal{X}^e} \eta_{\bar{}\Theta:x}\, n^e T dx.
\end{aligned}
\tag{2.28}
$$

Since the aggregate transportation cost $\overline{\text{ATC}}$ is equal to $\int_{\mathcal{X}} nTdx$, we obtain

Result 2.6 (Arnott and Stiglitz (1979)):

$$
\min_{x\in\mathcal{X}^e} \eta_{\bar{}\Theta:x} \equiv \min_{x\in\mathcal{X}^e}\left(\frac{\eta_{T:x}}{\eta_{\Theta:x}}\right) \le \frac{\overline{\text{DLR}}^e}{\overline{\text{ATC}}^e} \le \max_{x\in\mathcal{X}^e}\left(\frac{\eta_{T:x}}{\eta_{\Theta:x}}\right) \equiv \max_{x\in\mathcal{X}^e} \eta_{\bar{}\Theta:x}.
$$

The elasticity $\eta_{\Theta:x}$ is a measure of shape. For a linear city of unit width, this elasticity is a constant $\bar{\eta}_{\Theta:x} = 1$ while, for a circular city, $\bar{\eta}_{\Theta:x} = 2$.[16] The elasticity $\eta_{T:x}$ measures economies of scale in transportation with respect to distance: $\eta_{T:x} \geq (\leq)\, 1$ implies decreasing (constant,increasing) returns to scale. When the ratio of the two elasticities is constant throughout the urban area, result 2.6 reduces to a single equality. These remarks are consistent with the example discussed at the beginning of this section.

2.4 Appendices

2.4.1 Alternative Choice Problems

The standard *utility–maximisation problem* of an individual located at distance x from the centre, which we explained in section 2.2.2, is to determine an

[16] These are the two extremes of shape. In-between those extremes, there is a continuum of urban shapes for which the urban area increases with distance relatively faster than the linear city and slower than the circular city. Notice that the shape elasticity of any circular wedge is also two. The elasticity of shape becomes less than two as the linear segments that define the wedge turn inwards.

affordable consumption bundle (Z, H) which maximises utility. The resulting maximised utility level V can be represented in

$$V = v[R, \bar{Y} - T, \bar{\gamma}] \equiv \max_{Z,H} u[Z, H, \bar{\gamma}] \text{ subject to } Z + RH \leq \bar{Y} - T. \tag{2.29}$$

The solution for U, Z and H can also be obtained by using two alternative approaches. Firstly, the *expenditure–minimisation problem*

$$\bar{Y} - T = e[R, V, \bar{\gamma}] \equiv \min_{Z,H} Z + RH \text{ subject to } V \leq u[Z, H, \bar{\gamma}], \tag{2.30}$$

where $e[\cdot]$ is the minimum expenditure function. This is extensively used here. Secondly, the *rent-maximisation problem*

$$R = r[Y - T, V, \bar{\gamma}] \equiv \max_{Z,H} (\bar{Y} - Z - T)/H \text{ subject to } V \leq u[Z, H, \bar{\gamma}], \tag{2.31}$$

where $r[\cdot]$ is the *bid rent function.*

Figure 2.4 illustrates the relationships between the two alternatives and the original optimisation problem. Maximising utility subject to an income constraint appears in figure 2.4(1) where, with prices fixed, the objective is to find (Z, H) on the highest feasible indifference curve. Minimising expenditure subject to a utility constraint appears in figure 2.4(2) where, with prices fixed, the objective is to find (Z, H) on the lowest feasible budget line. Maximising land rent subject to income and utility constraints appears in figure 2.4(3) where, with the price of the composite good fixed, the objective is to find (Z, H) on the steepest possible budget line. Lines AA and BB are identical in all three diagrams. Hence point C, which defines the optimal choice in each case, is also the same. Using these figures,

$$v[R, e[R, V, \bar{\gamma}], \bar{\gamma}] = v[r[\bar{Y} - T, V, \bar{\gamma}], \bar{Y} - T, \bar{\gamma}] = V, \tag{2.32}$$

that is, the maximal utility from income $e[\cdot]$ given R, or from income $\bar{Y}$ given $r[\cdot]$, is V; and

$$e[R, v[R, \bar{Y} - T, \bar{\gamma}], \bar{\gamma}] = e[r[\bar{Y} - T, V, \bar{\gamma}], V, \bar{\gamma}] = \bar{Y} - T, \tag{2.33}$$

that is, the minimal expenditure necessary to achieve utility $v[\cdot]$ given R, or to achieve utility V given $r[\cdot]$, is $\bar{Y} - T$.

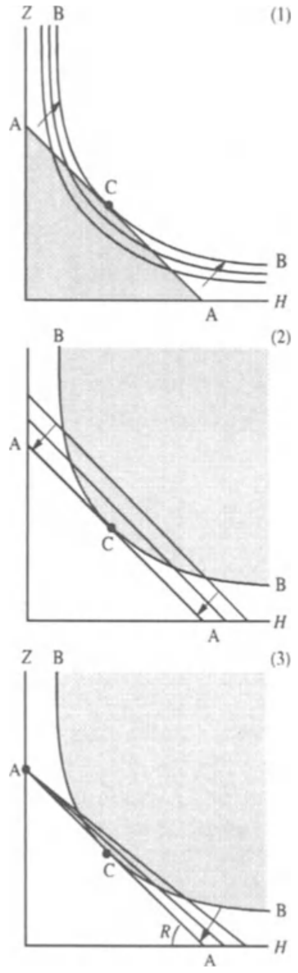

FIGURE 2.4. Relationships Among Choice Problems.

All three choice problems have the same solution. For the utility–maximisation problem, this solution represents *ordinary demand functions* $z^{(o)}\left[R, \bar{Y}-T, \bar{\gamma}\right]$ and $h^{(o)}\left[R, \bar{Y}-T, \bar{\gamma}\right]$, which determine $v[\cdot]$. For the expenditure–minimisation problem, it represents *compensated demand functions* $z[R, V, \gamma]$ and $h[R, V, \gamma]$, which determine $e[\cdot]$. Finally, for the rent–maximisation problem, it represents *bid-demand functions* $z^{(b)}[\bar{Y}-T, V, \bar{\gamma}]$ and $h^{(b)}[\bar{Y}-T, V, \bar{\gamma}]$, which determine $r[\cdot]$.

2.4.2 A Fundamental Issue

The use of calculus as the basic tool in urban economic analysis requires continua of agents distributed over continuous spaces. Yet, in the real world, discrete individuals occupy discrete parcels of land, each characterised by a specific location, shape and finite area. Although the possibility of a model based on parcels of land has been raised explicitly by Alonso (1964), the continuous approach has dominated urban economic theory. The implicit belief of those who have worked in urban economics, as well as in some other areas where analogous continuum models are used, seems to be that such models can serve as reasonable approximations to corresponding finite models. After all, working with densities continuously distributed over some space, rather than with discrete point patterns on that space or even with partitions of it, is both natural and convenient. However, the very foundation of this whole approach has been questioned by Berliant (1985) who has proven that, for any sequence of finite spatial economies which approaches a limiting economy with a continuum of agents, average endowments and consumption of land must tend to zero. Since the continuous model of urban economics calls for a positive amount of land consumed, it cannot serve as a reasonable approximation to large, finite spatial economies. In particular, the equilibrium solutions of a continuous model can be different from those in a discrete model where agents occupy parcels of land.

In the light of this criticism, the appropriateness of continuum modelling in urban economics can no longer be taken as self–evident. What needs to be done is to associate the continuous model of urban economics with some discrete model, and to propose conditions under which the equilibrium solutions and the comparative statics of the two models coincide. On the one hand utility in the continuous model is defined over a single point in space. On the other hand utility in the discrete model is defined over a parcel. How do the characteristics of a parcel affect utility, and how can they be linked with the continuous model? The heart of the matter is to construct a mapping, based on the continuous model utility and defined over consumption characteristics in the discrete model, which determines the discrete model utility in a way that renders the continuous and the discrete models equivalent. Such a mapping has been proposed by Papageorgiou and Pines (1990), and presupposes the existence of continuous model equilibria. According to Fujita and Smith (1987), existence is satisfied by the majority of existing models in urban economics, which associate distance from the centre with transportation cost alone as in chapters two and three of our book. However, in the more general case where utility also depends on location directly, a continuous model equilibrium may not exist. Under these circumstances, equivalence between the continuous and the discrete models becomes conditional upon such existence. This case arises not only through the disutility of travel time and through an uneven exogenous distribution of environmental quality (see sections 5.1.1 and 5.1.2 in chapter six respectively), but also in the presence of major urban externalities and public goods of later chapters. Therefore existence problems in the continuous model cannot be discounted as trivial. The same criticism applies to the discrete model, in the sense that the proposed mapping (which determines the way in which a population continuum is partitioned into discrete individuals) may not exist as well. Furthermore, although the equilibria of the continuous and the discrete model are the same under the proposed mapping, their demands may not be the same. Hence the two models may differ in terms of dynamic behaviour. Nevertheless, since the two models are equivalent, the comparative statics of the continuous model apply to the discrete model under the proposed mapping. It follows that when a continuous model equilibrium does exist, it can be used for studying the equilibrium behaviour of a corresponding discrete model. Conversely, when the proposed mapping does exist, we can study the equilibrium behaviour of its associated discrete model through a corresponding continuous model. These, in our opinion, are enough to justify the use of continuous modelling in urban economics.

2.5 References

W. Alonso, 1964, *Location and Land Use.* Harvard University Press, Cambridge, Mass.

A. Anas, R. J. Arnott and K. A. Small, forthcoming, Urban Spatial Structure. *Journal of Economic Literature.*

A. Anas and J. I. Kim, 1992, Income Distribution and the Residential Density Gradient. *Journal of Urban Economics* 31, 164–180.

R. J. Arnott and J. E. Stiglitz, 1979, Aggregate Land Rents, Expenditure on Public Goods, and Optimal City Size. *Quarterly Journal of Economics* 93, 471–500.

R. J. Arnott and J. E. Stiglitz, 1981, Aggregate Land Rent and Aggregate Transport Cost. *Economic Journal* 91, 331-347.

O. Ashenfelter, 1976, On the Interpretation of Urban Density Functions. *Journal of Urban Economics* 3, 82–84.

M. Berliant, 1985, Equilibrium Models with Land: A Criticism and an Alternative. *Regional Science and Urban Economics* 15, 325–340.

B. J. L. Berry, J. W. Simmons and R. J. Tennant, 1963, Urban Population Densities: Structure and Change. *Geographical Review* 53, 389–405.

H. Bleicher, 1892, *Statistische Beschreibung Der Staadt Frankfurt Am Main Und Ihre Bevölkerung.* Frankfurt Am Main.

J. K. Brueckner, 1982, A Note on Sufficient Conditions for Negative Exponential Population Density. *Journal of Regional Science* 22, 353–359.

E. Casetti, 1969, Alternative Urban Population Density Models: An Analytical Comparison of Their Validity Range. In A. J. Scott (ed.), *Studies in Regional Science.* Pion Limited, London.

C. Clark, 1951, Urban Population Densities. *Journal of the Royal Statistical Society* 114 (Series A), 490–496.

C. Clark, 1967, *Population Growth and Land Use.* St Martin's Press, New York.

M. Fujita and T. Smith, 1987, Existence of Continuous Residential Land–Use Equilibria. *Regional Science and Urban Economics* 17, 549–594.

R. M. Haig, 1926, Toward an Understanding of the Metropolis. *Quarterly Journal of Economics* 40, 421–423.

K. Inada, 1963, On a Two–Sector Model of Economic Growth: Comments and a Generalization. *Review of Economic Studies* 30, 119–127.

Y. Kanemoto, 1980, *Theories of Urban Externalities.* North–Holland, Amsterdam.

C. Kramer, 1955, Population Density Patterns. *CATS Research News* 2, 3–10.

R. Latham and M. Yeates, 1970, Population Density in Metropolitan Toronto. *Geographical Analysis* 2, 177–185.

J. R. Markusen and D. T. Scheffman, 1978, Ownership Concentration and Market Power in Urban Land Markets. *Review of Economic Studies* 45, 519–526.

A. Marshall, 1890, *Principles of Economics.* Reprinted by MacMillan, London, 1946.

E. S. Mills, 1969, The Value of Urban Land, in H. S. Perloff (ed.) *The Quality of Urban Environment.* The Johns Hopkins University Press, Baltimore.

E. S. Mills, 1972, *Studies in the Structure of the Urban Economy.* The Johns Hopkins University Press, Baltimore.

E. S. Mills and J. P. Tan, 1980, A Comparison of Urban Population Density Functions in Developed and Developing Countries. *Urban Studies* 17, 313–321.

H. Mohring, 1961, Land Values and the Measurement of Highway Benefits. *Journal of Political Economy* 69, 236–249.

R. F. Muth, 1961, The Spatial Structure of the Housing Market, *Papers and Proceedings of the Regional Science Association* 7, 207–220.

R. F. Muth, 1969, *Cities and Housing.* The University of Chicago Press, Chicago.

B. E. Newling, 1966, Urban Growth and Spatial Structure: Mathematical Models and Empirical Evidence. *The Geographical Review* 56, 213–225.

J. H. Niedercorn, 1971, A Negative Exponential Model of Urban Land Use Densities and Its Implications for Metropolitan Development. *Journal of Regional Science* 11, 317–326.

Y. Oron, D. Pines and E. Sheshinski, 1973, Optimum Versus Equilibrium Land Use Pattern and Congestion Toll. *Bell Journal of Economics and Management Science* 4, 619–636.

Y. Y. Papageorgiou and D. Pines, 1989, The Exponential Density Function: First Principles, Comparative Statics and Empirical Evidence. *Journal of Urban Economics* 26, 264–268.

Y. Y. Papageorgiou and D. Pines, 1990, The Logical Foundations of Urban Economics Are Consistent. *Journal of Economic Theory* 50, 37–53.

D. Pines and E. Sadka, 1986, A Comparative Statics Analysis of a Fully Closed City. *Journal of Urban Economics* 20, 1–20.

R. M. Solow, 1973, Congestion Cost and the Use of Land for Streets. *Bell Journal of Economics and Management Science* 4, 602–618.

J. Stewart, 1947, Suggested Principles of Social Physics. *Science*, 179–180.

W. C. Wheaton, 1974, A Comparative Statics Analysis of Urban Spatial Structure. *Journal of Economic Theory* 9, 223–237.

L. J. White, 1977, A Further Note on Density Functions and Average Distance from the Center. *Journal of Urban Economics* 4,218–129.

L. Wingo, Jr., 1961, *Transportation and Urban Land*. Resources for the Future, Washington D.C.

3

Comparative Statics

This chapter investigates the effects of the parameters on the characteristics of the urban equilibrium allocation. As we noted in section 2.3, there are in principle 3×2 city types ((renter, mixed, owner city) $\times$ (open, closed city)). To avoid some conceptual problems in defining land ownership when there is migration, we restrict the analysis to the four types shown in table 3.1.

Early works in urban economics were not specified well enough to allow for comparative statics. To our knowledge, the first attempts toward this direction were those of Casetti (1971) and Pines (1972) who dealt with an open renter city. Subsequent works, both positive (see, for example Solow (1972) and (1973)) and normative (see, for example, Dixit (1973) and Oron, Pines and Sheshinki (1973)), were formulated in a way allowing a systematic comparative statics analysis. Yet no such analysis was undertaken until the seminal paper of Wheaton (1974) who studied both cases on the first column of table 3.1. Since Wheaton was confined to a renter city, where $\bar{k} = 0$, his results could not be applied to normative models, such as those previously mentioned, which required an owner city, that is, $\bar{k} = 1$. Although the importance of the distinction between $\bar{k} = 0$ and $\bar{k} = 1$ was already recognised by Solow (1973), Oron, Pines and Sheshinski (1973) and, later on, by Wheaton (1979), a comparative statics analysis of an closed owner city was undertaken much later by Pines and Sadka (1986). In this chapter we present the existing results for both the open and closed city models. However, we restrict the discussion in the text to renter cities and base it mainly on diagrammatic and intuitive explanations. A more formal discussion for the closed mixed and closed owner cities is relegated to the appendix.

TABLE 3.1. City Types.

	RENTER	MIXED	OWNER
OPEN	•		
CLOSED	•	•	•

3.1 Open Renter City

For an open city, equilibrium was defined in section 2.2.7 as $\left(n^e, R^e, \bar{Y}^e, \bar{N}^e, \bar{x}^e\right)$ that satisfies (2.12) for known parameters $\bar{\Omega}, \bar{U}, \bar{R}, \bar{k}, \bar{\alpha}, \bar{\beta}$ and $\bar{\gamma}$. However, as we explained above, we confine our discussion to an open renter city where $\bar{k} = 0$. In such a city, $\bar{Y}^e = \bar{\Omega}$ by (2.11). We therefore investigate the impact of changes in the parameters $\bar{\Omega}, \bar{U}, \bar{R}, \bar{\alpha}, \bar{\beta}$ and $\bar{\gamma}$ on the variables $D^e, R^e, \bar{N}^e$and $\bar{x}^e$, as well as on $\overline{\text{DLR}}^e$, $\overline{\text{ATC}}^e$ and the *aggregate distance travelled* at equilibrium

$$\overline{\text{ADT}}^e \equiv \int_{\mathcal{X}^e} n^e x dx. \tag{3.1}$$

Notice that since population density is the inverse of land per capita, (2.9) implies $n = \theta D$. Thus, determining the comparative statics with respect to D also determines the comparative statics of the original variable n.

3.1.1 Initial Endowment

Since $\bar{Y}^e = \bar{\Omega}$, we have from (2.8) that $e\left[R^e, \bar{U}, \bar{\gamma}\right] = \bar{\Omega} - T$. This and (2.15) imply that the urban rent increases everywhere as the initial endowment increases. Since urban rent must increase at the original border while the opportunity cost of land remains fixed, the spatial extent of the city must also increase by (2.11) and result 2.3 as the initial endowment increases. Now, using the property of compensated demands, we know that $h\left[R^e, \bar{U}, \bar{\gamma}\right]$ decreases everywhere because urban rent increases everywhere. Therefore, since $HD = 1$, the urban population density increases everywhere as the initial endowment increases. Increasing urban rent and urban area imply an increased differential land rent by (2.11) and (2.12(b)). Finally, since the number of individuals at a particular distance from the centre increases with the density there, increasing urban density and urban area imply an increased total urban population by (2.10) and, by definition, increased aggregate transportation cost and aggregate distance travelled.

An increasing initial endowment raises utility in the city, which causes immigration from the countryside. Higher demand for urban land follows from two distinct reasons. The first is the increase in the demand per capita resulting from the increase in the initial endowment. This demand increase tends to raise both the urban land rent and the per–capita housing consumption, thus reducing density. The second is the demand of the immigrants which raises the land

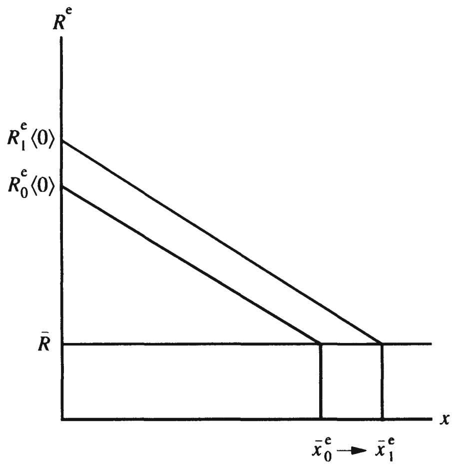

FIGURE 3.1. Effects of Increasing Initial Endowment on an Open Renter City.

rent but suppresses per–capita housing consumption, thus increasing density. The second effect, which restores urban utility to its original lower level, dominates so that urban rent increases as in figure 3.1[1] Higher land values, in turn, extend the urban area outwards. In summary,

Result 3.1 (Wheaton (1974)):

(i) $$\frac{dD^e}{d\bar{\Omega}} > 0; \quad \frac{dR^e}{d\bar{\Omega}} > 0; \quad \frac{d\bar{N}^e}{d\bar{\Omega}} > 0;$$

$$\frac{d\bar{x}^e}{d\bar{\Omega}} > 0.$$

(ii) $$\frac{d\overline{\text{DLR}}^e}{d\bar{\Omega}} > 0; \quad \frac{d\overline{\text{ATC}}^e}{d\bar{\Omega}} > 0; \quad \frac{d\overline{\text{ADT}}^e}{d\bar{\Omega}} > 0.$$

3.1.2 Agricultural Rent

Using $e\left[R^e, \bar{U}, \bar{\gamma}\right] = \bar{\Omega} - T$, we conclude that the urban land rent does not change as the agricultural rent increases with every other parameter remaining fixed. Since rent does not change, land per capita and, therefore, population density do not change as well. However, since agricultural rent increases at the original border while the urban rent remains fixed, the spatial extent of the city must become smaller. Unchanged urban rent and smaller urban area imply a decreased differential land rent. Finally, unchanged urban density and smaller

[1] In order to obtain linear graphs we employ the case of constant locational costs discussed in section 2.3.5.

urban area imply a smaller total urban population, aggregate transportation cost and aggregate distance travelled.

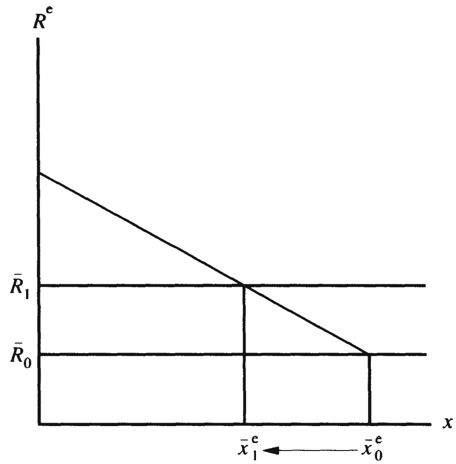

FIGURE 3.2. Effects of Increasing Agricultural Rent on an Open Renter City.

An increasing agricultural rent only affects urban residents close enough to the countryside. Since migration is costless, those who cannot afford the competition for land in agriculture leave the city. Since nothing else changes, the city retracts as in figure 3.2. In summary,

Result 3.2 (Wheaton (1974)):

(i) $$\frac{dD^e}{d\bar{R}} = 0; \quad \frac{dR^e}{d\bar{R}} = 0; \quad \frac{d\bar{N}^e}{d\bar{R}} < 0;$$

$$\frac{d\bar{x}^e}{d\bar{R}} < 0.$$

(ii) $$\frac{d\overline{\text{DLR}}^e}{d\bar{R}} < 0; \quad \frac{d\overline{\text{ATC}}^e}{d\bar{R}} < 0; \quad \frac{d\overline{\text{ADT}}^e}{d\bar{R}} < 0.$$

3.1.3 Utility Level

Using $e\left[R^e, \bar{U}, \bar{\gamma}\right] = \bar{\Omega} - T$, we know that urban rent must decline everywhere in order to keep expenditure fixed as utility increases. Since urban rent must decrease at the original border while the opportunity cost of land remains fixed, the spatial extent of the city must also decrease as utility increases. Furthermore, using the property of compensated demands, we know that land per capita increases everywhere because urban rent decreases everywhere. Therefore urban population density decreases everywhere as utility increases. Decreasing urban rent and urban area imply a decrease in differential land rent. Finally, since the number of individuals at any particular distance from the centre increases with the density there, decreasing urban density and urban area imply decreasing total urban population, aggregate transportation cost and aggregate distance travelled.

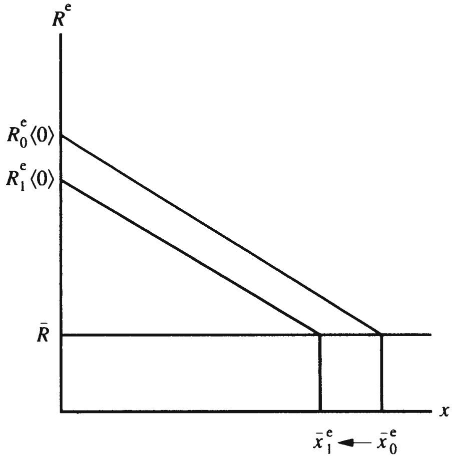

FIGURE 3.3. Effects of Increasing Utility on an Open Renter City.

An increasing level of utility outside of the city causes emigration to the countryside, which raises the utility level in the city to that in the alternative sector. Once again there are two distinct changes in the demand for housing. The first is the increase in the per–capita demand resulting from higher utility. The second is the decrease in the demand caused by the emigration. The net effect is an increased housing consumption and, consequently, a decrease in both population density and urban land rent as in figure 3.3. Lower urban land values, in turn, move the urban border inwards. In summary,

Result 3.3 (Wheaton (1974)):

(i) $$\frac{dD^e}{d\bar{U}} < 0; \quad \frac{dR^e}{d\bar{U}} < 0; \quad \frac{d\bar{N}^e}{d\bar{U}} < 0;$$

$$\frac{d\bar{x}^e}{d\bar{U}} < 0.$$

(ii) $$\frac{d\overline{\text{DLR}}^e}{d\bar{U}} < 0; \quad \frac{d\overline{\text{ATC}}^e}{d\bar{U}} < 0; \quad \frac{d\overline{\text{ADT}}^e}{d\bar{U}} < 0.$$

3.1.4 Transportation Technology

Suppose that $T\langle 0\rangle = 0$. Then, using $e\left[R^e\langle 0\rangle, \bar{U}, \bar{\gamma}\right] = \bar{\Omega}$, we conclude that the urban rent at the centre and, therefore, the population density there do not change as $\bar{\alpha}$ increases. However, at any other distance from the centre, improved transportation technology raises the disposable income. For fixed levels of utility and other parameters, $e\left[R^e, \bar{U}, \bar{\gamma}\right] = \bar{\Omega} - T$ implies that urban rent must increase everywhere in $(0, \bar{x}^e]$ as the transportation technology improves. The difference in urban rent will be stronger further away because more distant locations benefit more from transportation improvements. Consequently, the urban population density in $(0, \bar{x}^e]$ will rise as urban rent does. Since at the original border urban rent increases while agricultural rent remains fixed, the border moves outward. Increasing urban rent and urban area imply an increased differential land rent. Finally, increasing urban density and urban area imply an increased total urban population, aggregate transport cost and aggregate distance travelled.

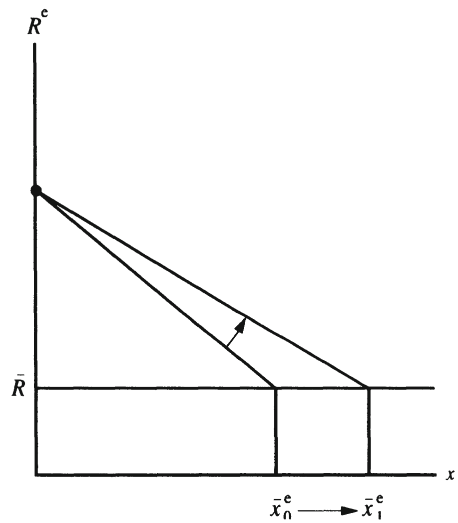

FIGURE 3.4. Effects of Improving Transportation Technology on an Open Renter City.

Lower transportation cost increases urban disposable income and, therefore raises utility in the city. This attracts immigration from the countryside which restores the original utility level. As previously, housing demand increases because of two reasons. First, the increase in disposable income raises per–capita demand everywhere excepting the centre. This tends to increase both rent and housing consumption. Second, the demand of the immigrants raises urban rent everywhere and depresses housing consumption. The net effect is an increase in urban rent as in figure 3.4 and a decrease in housing consumption, which is equivalent to an increase in population density everywhere excepting the centre. Higher urban land values, in turn, extend the urban area outwards. In summary,

Result 3.4:

(i) $$\frac{dD^e\langle 0\rangle}{d\bar{\alpha}} = 0; \quad \frac{dR^e\langle 0\rangle}{d\bar{\alpha}} = 0.$$

(ii) $$\frac{dD^e}{d\bar{\alpha}} > 0; \quad \frac{dR^e}{d\bar{\alpha}} > 0 \qquad \text{for} \quad x \in (0, \bar{x}^e].$$

(iii) $$\frac{d\bar{N}^e}{d\bar{\alpha}} > 0; \quad \frac{d\bar{x}^e}{d\bar{\alpha}} > 0.$$

(vi) $$\frac{d\overline{\text{DLR}}^e}{d\bar{\alpha}} > 0; \quad \frac{d\overline{\text{ATC}}^e}{d\bar{\alpha}} > 0; \quad \frac{d\overline{\text{ADT}}^e}{d\bar{\alpha}} > 0.$$

3.1.5 Land Reclamation

When $\bar{\beta}$ increases, nothing changes in $e\left[R^e, \bar{U}, \bar{\gamma}\right] = \bar{\Omega} - T$. Therefore both urban rent and population density remain the same as available land is reclaimed to become urban. It follows that the urban border also remains the same. However, since land per capita remains fixed as land supply increases, the number of individuals at that particular distance from the centre must increase. Therefore total urban population becomes larger. Furthermore, increasing the urban land supply while keeping the same urban land values and urban area implies that differential land rent increases. Finally, a larger total population distributed so that population density remains unchanged implies an increased aggregate transportation cost and aggregate distance travelled.

When urban land is reclaimed at some distance from the centre, the excess supply of land is absorbed by immigrants from the surrounding region. Conditions in the existing residential area are not affected because utility and disposable income in the city remain as before. In summary,

Result 3.5:

(i) $$\frac{dD^e}{d\bar{\beta}} = 0; \quad \frac{dR^e}{d\bar{\beta}} = 0; \quad \frac{d\bar{N}^e}{d\bar{\beta}} > 0;$$

$$\frac{d\bar{x}^e}{d\bar{\beta}} = 0.$$

(ii) $$\frac{d\overline{\text{DLR}}^e}{d\bar{\beta}} > 0; \quad \frac{d\overline{\text{ATC}}^e}{d\bar{\beta}} > 0; \quad \frac{d\overline{\text{ADT}}^e}{d\bar{\beta}} > 0.$$

3.1.6 Public Good

Using $e\left[R^e, \bar{U}, \bar{\gamma}\right] = \bar{\Omega} - T$, if the public good service increases, land rent must increase everywhere. We then follow the argument of section 3.1.1 to conclude that the effects of a better public good service are precisely analogous to the effects of a higher initial endowment. Thus figure 3.1 holds. In summary,

Result 3.6:

(i) $$\frac{dD^e}{d\bar{\gamma}} > 0; \quad \frac{dR^e}{d\bar{\gamma}} > 0; \quad \frac{d\bar{N}^e}{d\bar{\gamma}} > 0;$$

$$\frac{d\bar{x}^e}{d\bar{\gamma}} > 0.$$

(ii) $$\frac{d\overline{\text{DLR}}^e}{d\bar{\gamma}} > 0; \quad \frac{d\overline{\text{ATC}}^e}{d\bar{\gamma}} > 0; \quad \frac{d\overline{\text{ADT}}^e}{d\bar{\gamma}} > 0.$$

3.2 Closed Renter City

For the closed renter city model, we investigate the impact of changes in the parameters $\bar{\Omega}, \bar{N}, \bar{R}, \bar{\alpha}, \bar{\beta}$ and $\bar{\gamma}$ on the variables $D^e, R^e, \bar{U}^e$ and $\bar{x}^e$, as well as on $\overline{\text{DLR}}^e$, $\overline{\text{ATC}}^e$ and $\overline{\text{ADT}}^e$. Basically, we do this by using supply and demand analysis for residential sites as functions of some representative price of housing (land rent). We choose the rent at the centre to represent housing rents in general.

The *total demand for residential sites* $\bar{N}^d$ in a closed city model is infinitely inelastic because the total population size is fixed by definition. When utility is equalised across locations, the *total supply for residential sites* $\bar{N}^s$ is given by

$$\begin{aligned}
& \bar{N}^{s,e}\left[R^e\langle 0\rangle ; \bar{N}, \bar{R}, \bar{\alpha}, \bar{\beta}, \bar{\gamma}, \bar{\Omega}\right] \\
= & \int_{\mathcal{X}^e}\left(\frac{\theta\left[x, \bar{\beta}\right]}{h\left[r\left[x, R^e\langle 0\rangle ; \bar{\alpha}, \bar{\gamma}, \bar{\Omega}\right], v\left[R^e\langle 0\rangle, \bar{\Omega}, \bar{\gamma}\right], \bar{\gamma}\right]}\right) dx \\
\underset{(\text{result } 2.2)}{=} & -\int_{\mathcal{X}^e} \sigma \frac{\partial r}{\partial x} dx \\
\underset{(\text{integration by parts})}{=} & \sigma\langle 0\rangle R^e\langle 0\rangle-\sigma\langle\bar{x}^e\rangle \bar{R}+\int_{\mathcal{X}^e} \frac{\partial \sigma}{\partial x} r[\cdot] dx
\end{aligned} \tag{3.2}$$

where $\sigma \equiv \partial\tilde{\Theta}/\partial T$, $v\left[R^e, \bar{\Omega} - T, \bar{\gamma}\right]$ is the *indirect utility function* at x and $r\left[x, R^e\langle 0\rangle, \bar{\alpha}, \bar{\gamma}, \bar{\Omega}\right]$ is the solution for R^e of

$$v\left[R^e, \bar{\Omega} - T[x, \bar{\alpha}], \bar{\gamma}\right] = v\left[R^e\langle 0\rangle, \bar{\Omega}, \bar{\gamma}\right]. \tag{3.3}$$

The equilibrium supply for residential sites is an increasing function of central rent because both the integrand and the domain of the integral on the second line of (3.2) increase with central rent. To see this observe first by using (3.3) that $r[\cdot]$ increases with central rent and decreases with distance from the centre, while $v[\cdot]$ decreases with central rent. Then (i) the integrand increases because its denominator $h[\cdot]$ must decrease as a result of the increase in urban rent and the decline in utility which results from an increase in central rent. (ii) Since $r[\cdot]$ decreases with distance from the centre and increases with central rent, it follows from (2.11) that

$$\frac{d\bar{x}^e}{dR^e\langle 0\rangle} = -\left(\frac{\partial r^e/\partial R\langle 0\rangle}{\partial r/\partial \bar{x}^e}\right) > 0. \tag{3.4}$$

This, and assumption 2.1, ensure that the domain of the integral increases with central rent as required. Therefore we have

$$\frac{\partial \bar{N}^{s,e}}{\partial R\langle 0\rangle} > 0. \tag{3.5}$$

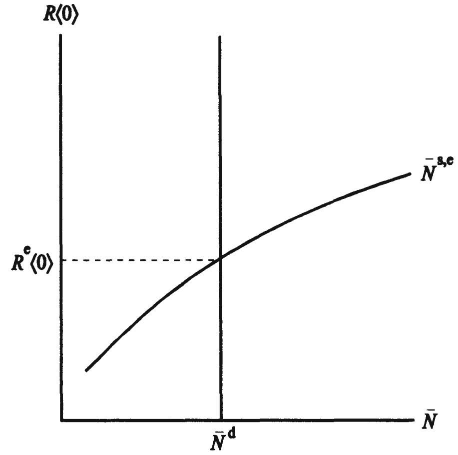

FIGURE 3.5. Determination of the Equilibrium Central Rent in a Closed Renter City.

The determination of the equilibrium rent at the centre of the city can be represented by the intersection of supply $\bar{N}^{s,e}$ and demand $\bar{N}^d = \bar{N}$ in figure 3.5. We shall use this figure to evaluate the effects of changes in the parameters $\bar{N}, \bar{R}, \bar{\alpha}, \bar{\beta}$ and $\bar{\gamma}$ on the variables $D^e, R^e, \bar{U}^e, \bar{x}^e$, $\overline{\text{DLR}}^e$, $\overline{\text{ATC}}^e$ and $\overline{\text{ADT}}^e$. Since we cannot use the supply and demand approach to evaluate the effects of the initial endowment, we study $\bar{\Omega}$ last.

3.2.1 Population Size

Increasing urban population size causes a rightward shift in the demand for residential sites as reflected in figure 3.6, which raises the equilibrium rent at

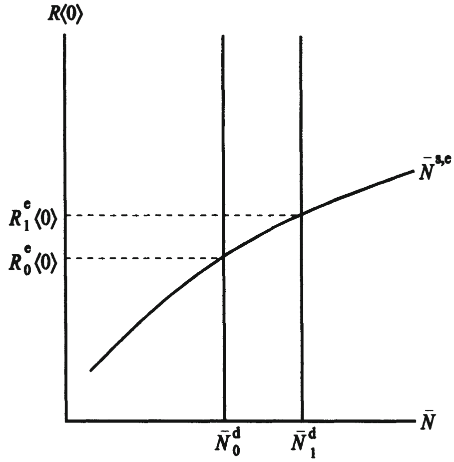

FIGURE 3.6. The Impact of Increasing Population on Central Rent in a Closed Renter City.

the city center from $R_0^e\langle 0\rangle$ to $R_1^e\langle 0\rangle$. It follows that the equilibrium rent increases over the entire urban area. Consequently the per-capita demand for land declines, the equilibrium population density increases everywhere, the equilibrium utility level decreases by (3.3) and the city area expands by (3.4). Some of these changes are shown in figure 3.7. Since the rent increases everywhere and the city area expands, $\overline{\text{DLR}}^e$ must increase. Furthermore, with increased

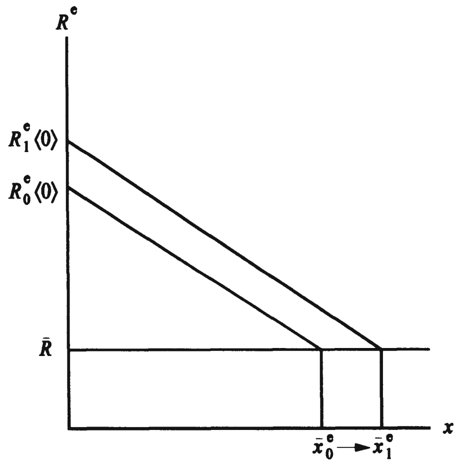

FIGURE 3.7. Effects of Increasing Population Size on a Closed Renter City.

density everywhere and with a larger city area, $\overline{\text{ATC}}^e$ and $\overline{\text{ADT}}^e$ also increase.

Result 3.7 (Wheaton (1974)):

(i) $\frac{dD^e}{d\bar{N}} > 0; \quad \frac{dR^e}{d\bar{N}} > 0; \quad \frac{d\bar{U}^e}{d\bar{N}} < 0; \quad \frac{d\bar{x}^e}{d\bar{N}} > 0.$

(ii) $\frac{d\overline{\mathrm{DLR}}^e}{d\bar{N}} > 0; \quad \frac{d\overline{\mathrm{ATC}}^e}{d\bar{N}} > 0; \quad \frac{d\overline{\mathrm{ADT}}^e}{d\bar{N}} > 0.$

3.2.2 Agricultural Rent

An increase in the agricultural land rent causes a decline in the supply of residential sites and a resulting increase in central rent. To see this observe that a fixed central rent implies by (3.3) that the urban rent profile remains unchanged. Differentiate (2.11) to obtain $\partial\bar{x}^e/\partial\bar{R} = 1/\left(\partial R^e\left\langle\bar{x}^e\right\rangle/\partial x\right) < 0$, i.e. the city area contracts. Hence the domain of the integral on the second line of (3.2) decreases and so does the supply of residential sites, i.e. $\partial\bar{N}^{s,e}/\partial\bar{R} < 0$. This is reflected in figure 3.8 by the shift of the supply function from $\bar{N}_0^{s,e}$ to $\bar{N}_1^{s,e}$ and the resulting increase of the central rent from $R_0^e\langle 0\rangle$ to $R_1^e\langle 0\rangle$.

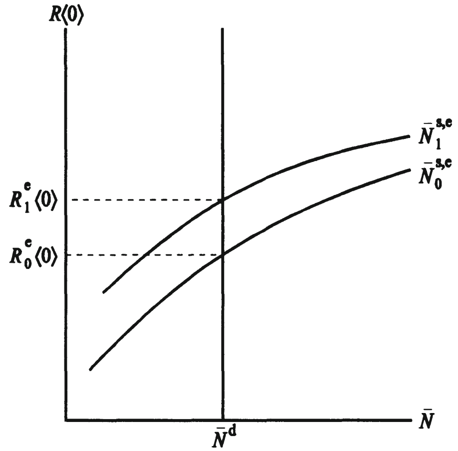

FIGURE 3.8. The Impact of Increasing Agricultural Rent on Central Rent in a Closed Renter City.

We know that if the central rent increases so does the equilibrium urban rent everywhere. Therefore, since per capita income remains fixed, equilibrium utility decreases, housing consumption also decreases and the corresponding density increases within the contracted urban area. Some of these changes are shown in figure 3.9. Because density increases everywhere while the population remains the same, it must be the case that both the aggregate transport cost and the aggregate distance travelled decline. However, we cannot determine the impact of an increasing agricultural rent on the differential land rent because it is subject to three conflicting effects. First, the urban rent increases everywhere.

Second, the agricultural rent, which is subtracted from the gross rent to derive

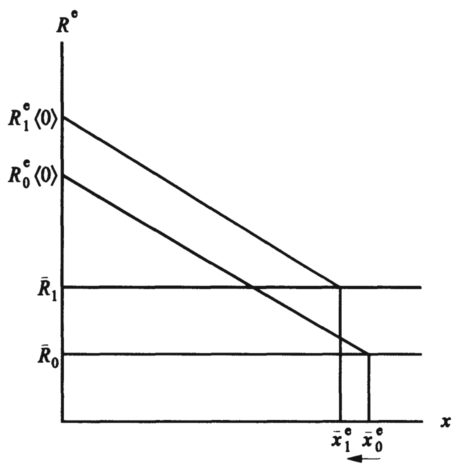

FIGURE 3.9. Effects of Increasing Agricultural Rent on a Closed Renter City.

the differential land rent, increases. Third, the city area contracts. In general our assumptions are not sufficient to specify the combined outcome of those three effects, and stronger conditions are required.[2] In summary

Result 3.8 (Wheaton (1974)):

(i) $$\frac{dD^e}{d\bar{R}} > 0; \quad \frac{dR^e}{d\bar{R}} > 0; \quad \frac{d\bar{U}^e}{d\bar{R}} < 0; \quad \frac{d\bar{x}^e}{d\bar{N}} < 0.$$

(ii) $$\frac{d\overline{\mathrm{ATC}}^e}{d\bar{R}} < 0; \quad \frac{d\overline{\mathrm{ADT}}^e}{d\bar{R}} < 0.$$

3.2.3 Transportation Technology

Given $R^e\langle 0\rangle$, the utility at the centre $v\left[R^e\langle 0\rangle, \bar{\Omega}, \bar{\gamma}\right]$ does not change with $\bar{\alpha}$. It then follows from (3.3) that R^e increases everywhere else in the urban area as $\bar{\alpha}$ increases and, by (2.11), that the city area expands. Furthermore, with unchanged utility and increased land rent, the denominator on the second line of (3.2) declines due to the pure substitution effect. Thus the supply of residential sites increases. In terms of figure 3.8, the supply shifts from $\bar{N}_1^{s,e}$ to $\bar{N}_0^{s,e}$ and this results to a decline of the central rent from $R_1^e\langle 0\rangle$ to $R_0^e\langle 0\rangle$. Consequently, the equilibrium utility increases.

[2]For example, if the elasticity of the transportation cost shape of the city $\eta_{\dot{\Theta}:T}$ is fixed with respect to x (e.g. if the elasticities of T and Θ with respect to x are fixed) then the ratio $\overline{\mathrm{DLR}}^e/\overline{\mathrm{ATC}}^e$ is also fixed so that $\overline{\mathrm{DLR}}^e$ declines as does $\overline{\mathrm{ATC}}^e$.

Can the urban rent decline everywhere in response to an improvement in transportation? If it could, the housing demand would increase and the density would decrease everywhere. Since the urban area would also become smaller, the city would be unable to accommodate its population. Thus the rent must increase somewhere, so that there exists a location $x^* \in (0, \bar{x}^e)$ where the rent does not change, i.e. $dR\langle x^*\rangle / d\bar{\alpha} = 0$. Now consider the change in the slope of the rent gradient at x^*. Toward this end differentiate result 2.2 with respect to $\bar{\alpha}$ at x^* to obtain

$$\frac{d}{d\bar{\alpha}}\left(\frac{dR^e\langle x^*\rangle}{dx}\right) = \left(\frac{t}{H^{e2}}\frac{\partial h^e}{\partial \bar{U}}\frac{d\bar{U}^e}{d\bar{\alpha}} - \frac{1}{H^e}\frac{\partial t}{\partial \bar{\alpha}}\right)\Bigg|_{x=x^*} > 0. \tag{3.6}$$

The inequality follows since housing is a normal good, $v\left[R^e, \bar{\Omega} - T\left[x, \bar{\alpha}\right], \bar{\gamma}\right]$ must increase at x^* (hence $\bar{U}^e$ must also increase) and, by assumption 2.2, the transportation cost rate decreases. Thus there exists a unique $x^* \in (0, \bar{x}^e)$ such that $dR^e/d\bar{\alpha} \leq (\geq) 0$ as $x \leq (\geq) x^*$, so that the rent gradient flattens as in figure 3.10.

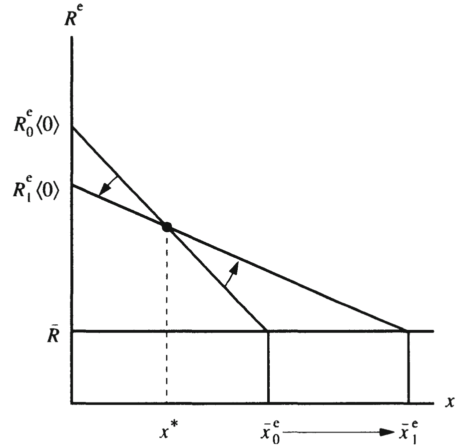

FIGURE 3.10. Effects of Improving Transportation Technology on a Closed Renter City.

We are unable to sign the effect of an improvement in transportation technology on the differential land rent and on the aggregate transportation cost. It is actually possible to construct examples for both an increase and a decrease of those variables in response to better transportation. By contrast, we can determine the effect of transportation improvements on the aggregate distance travelled because we know that the urban population disperses. Since the rent decreases within x^*, housing demand increases there due to both effects of reduced price and increased utility. In consequence density decreases, implying that within any radius smaller than x^* fewer individuals reside. Taking into account that total population is fixed, this change can be described as a reallocation of individuals from more to less accessible sites implying that the aggregate distance travelled increases with an improvement in transportation technology. More precisely, by lemma 3.4 in the appendix, we know that a larger number of individuals lives outside any particular radius over the original $(0, \bar{x}_0^e]$ after transportation has improved:

$$\frac{dN^e}{d\bar{\alpha}} < 0 \text{ for } x \in (0, \bar{x}_0^e]. \tag{3.7}$$

Also, by differentiating the definition of the aggregate distance travelled

$$\begin{aligned} \overline{\text{ADT}}^e &\equiv \int_{\mathcal{X}^e} n^e x dx \\ &= \bar{N}\bar{x}^e - \int_{\mathcal{X}^e} N^e dx \end{aligned} \tag{3.8}$$

with respect to $\bar{\alpha}$, we have

$$\frac{d\overline{\text{ADT}}^e}{d\bar{\alpha}} = -\int_{\mathcal{X}^e} \frac{dN^e}{d\bar{\alpha}} dx. \tag{3.9}$$

This and (3.7) imply that the aggregate distance travelled increases as transportation improves. Notice that the same argument cannot apply to the aggregate transportation cost because the higher propensity to travel as transportation improves is countered by a lower transportation rate. Thus whether the aggregate transportation cost will increase or decrease with an improvement in transportation, depends on the relative importance of these two effects.[3] In summary, we have

---

[3] Arnott, Pines and Sadka (1986) , consider this problem in the special case where $T[x, \bar{\alpha}] = f[x]/\bar{\alpha}$ with $f' > 0$ and $\theta = \bar{\beta}x$. Then

$$\overline{\text{ATC}}^e = \frac{1}{\bar{\alpha}} \int_{\mathcal{X}^e} n^e f dx \equiv \frac{1}{\bar{\alpha}} \overline{\text{ATD}}^e,$$

where $1/\bar{\alpha}$ can be interpreted as the price of transportation and $\overline{\text{ATD}}$, the *aggregate transportation demand*, as the corresponding quantity demanded. In this case assumption 3.1, hence lemma 3.4, holds. Since, similarly to (3.9),

$$\frac{d\overline{\text{ATD}}^e}{d\bar{\alpha}} = -\int_{\mathcal{X}^e} \frac{dN^e}{d\bar{\alpha}} f' dx,$$

the quantity of transportation demanded falls as its price increases. Now integrate by parts the aggregate transportation cost to obtain

$$\overline{\text{ATC}}^e = \bar{N}T\langle \bar{x}^e \rangle - \int_{\mathcal{X}^e} N^e t dx = \int_{\mathcal{X}^e} (\bar{N} - N^e) t dx.$$

Upon differentiation with respect to $\bar{\alpha}$,

$$\frac{d\overline{\text{ATC}}^e}{d\bar{\alpha}} = \frac{d\overline{\text{ATD}}^e}{d\bar{\alpha}} + \int_{\mathcal{X}^e} (\bar{N} - N^e) \frac{dt}{d\bar{\alpha}} dx.$$

The first term on the RHS represents the increase in the demand for transportation caused by the transportation improvement. The second, which is negative by assumption 2.2, reflects

Result 3.9 (Wheaton (1974), Pines and Sadka (1986)):

(i) $$\frac{dD^e\langle 0\rangle}{d\bar{\alpha}} < 0; \quad \frac{d\bar{U}^e}{d\bar{\alpha}} > 0; \quad \frac{d\bar{x}^e}{d\bar{\alpha}} > 0.$$

(ii) There exists a unique $x^* \in (0, \bar{x}^e)$ such that
$$\frac{dR^e}{d\bar{\alpha}} \leq (\geq)\, 0 \text{ as } x \leq (\geq)\, x^*.$$

(iii) $$\frac{d\overline{\mathrm{ADT}}^e}{d\bar{\alpha}} > 0.$$

3.2.4 Land Reclamation

Land reclamation is represented by an increase in $\bar{\beta}$ which increases $\theta\left[x, \bar{\beta}\right]$. Therefore land reclamation increases the numerator on the second line of (3.2). Keeping the central rent fixed we know by (3.3) that the equilibrium rent function does not change, which implies that the denominator on the second line of (3.2) does not change as well. It follows that increasing $\bar{\beta}$ while keeping the central rent fixed increases the equilibrium supply of residential sites, i.e. $\partial N^{s,e}/\partial\bar{\beta} > 0$. In terms of figure 3.8, the supply shifts from $\bar{N}_1^{s,e}$ to $\bar{N}_0^{s,e}$ and the central rent declines from $R_1^e\langle 0\rangle$ to $R_0^e\langle 0\rangle$. It follows from (3.3) that the equilibrium rent declines everywhere, so that housing consumption increases together with the utility level, population density declines and, by (2.11), the city contracts. These changes are shown in figure 3.11.

In general, we cannot determine the effect of land reclamation on the aggregate transportation cost and the aggregate distance travelled. The reason is that, there are three effects, two of which tend to reduce the magnitude of these variables and one which tends to increase it. The contraction of the city boundary, $\bar{x}^e$, and the new supply of land inside the old boundary contribute to the decline of aggregate transportation cost and aggregate distance travelled. However, these effects can be offset by the increase in housing demand and the resulting decline in density. Finally, we cannot determine the effect of land reclamation on the differential land rent because the rent declines but the supply of land increases everywhere at the same time.

In summary,

Result 3.10:

$$\frac{dD^e}{d\bar{\beta}} < 0; \quad \frac{dR^e}{d\bar{\beta}} < 0; \quad \frac{d\bar{U}^e}{d\bar{\beta}} > 0; \quad \frac{d\bar{x}^e}{d\bar{\beta}} < 0.$$

---

the effect of a declining transportation rate on the initial demand for transportation. The net effect depends on the elasticity of $\overline{\mathrm{ATD}}^e$ with respect to its price $1/\bar{\alpha}$. This elasticity, and the factors determining it, is the subject of Arnott, Pines and Sadka (1986) .

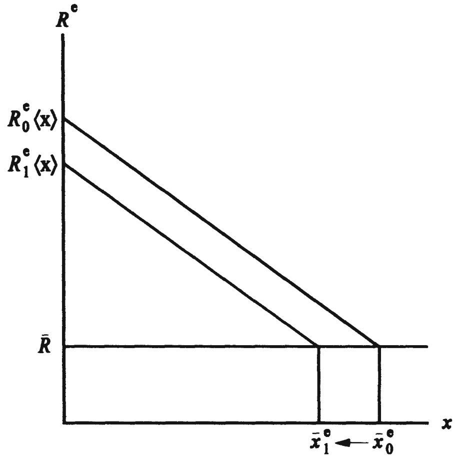

FIGURE 3.11. Effects of Land Reclamation on a Closed Renter City.

3.2.5 Public Good

We use (3.2) and assumption 2.3 regarding the transportation cost shape of the city in order to characterise the effects of an increase in $\bar{\gamma}$. We will first show that, following an increase in $\bar{\gamma}$, the rent neither can increase everywhere nor can it decrease everywhere. Differentiating the last equation on the RHS of (3.2) with respect to $\bar{\gamma}$, one obtains

$$\begin{aligned} \frac{\partial \bar{N}^{s,e}}{\partial \bar{\gamma}} &= \int_{\mathcal{X}^e} \frac{\partial \sigma}{\partial x} \frac{\partial r}{d\bar{\gamma}} dx \\ &= 0. \end{aligned} \tag{3.10}$$

The last equality in (3.10) holds because $\bar{N}^{s,e} = \bar{N}$. Since $\partial\sigma/\partial x > 0$ by assumption 2.3, (3.10) establishes our claim that the equilibrium rent cannot increase or decrease everywhere. Hence, as in section 3.2.3, there is a location x^* where the rent does not change. The change in the slope of the rent gradient at x^* can be determined by differentiating result 2.2 once again with respect to $\bar{\gamma}$ at x^*, which gives

$$\frac{d}{d\bar{\gamma}}\left(\frac{dR^e\langle x^*\rangle}{dx}\right) = \left(\frac{t}{H^{e2}}\left(\frac{\partial h^e}{\partial \bar{U}}\frac{d\bar{U}^e}{d\bar{\gamma}} + \frac{\partial h}{\partial \bar{\gamma}}\right)\right)\Bigg|_{x=x^*}. \tag{3.11}$$

The first term in parenthesis on the RHS of (3.11) is positive because of assumption 2.2 and because, with unchanged rent and income, utility increases with $\bar{\gamma}$. The second term, on the other hand, is positive (negative) depending upon whether housing and the public good are complements (substitutes). If they are complements, the rent flattens about x^* as in figure 3.10 and, therefore, the central rent declines and the urban area expands. Moreover, since utility increases and the rent decreases within radius x^*, population density also decreases within this radius. In contrast, whenever housing and the public good are substitutes, as in the case of a public park which reduces the demand for private gardens, the sign of (3.11) remains ambiguous and nothing more can be said about the impact of the public good on the equilibrium urban structure.

It is clear that if housing and the public good are complements then the population disperses with an increasing level of the public good because the rent decreases over the range $(0, x^*]$, so that the housing demand increases due to the reduced price effect, increased utility and the increased supply of the public good. Since the population disperses both the aggregate distance travelled and the aggregate transportation cost increase with the level of the public good. For the aggregate distance travelled, differentiating (3.8) with respect to $\bar{\gamma}$ gives

$$\frac{d\overline{\text{ADT}}^e}{d\bar{\gamma}} = -\int_{\mathcal{X}^e} \frac{dN^e}{d\bar{\gamma}} dx. \tag{3.12}$$

Our claim follows because $dN^e/d\bar{\gamma} < 0$, which we can establish by substituting $\bar{\gamma}$ for $\bar{\Omega}$ in the proof of lemma 3.1 in the appendix. The effect of the public good on $\overline{\text{ATC}}^e$ is very similar to that on $\overline{\text{ADT}}^e$. Using the definition

$$\begin{aligned} \overline{\text{ATC}}^e &\equiv \int_{\mathcal{X}^e} n^e T dx \\ &= \bar{N} T \langle \bar{x}^e \rangle - \int_{\mathcal{X}^e} N^e t dx, \end{aligned} \tag{3.13}$$

we get

$$\frac{d\overline{\text{ATC}}^e}{d\bar{\gamma}} = -t \int_{\mathcal{X}^e} \frac{dN^e}{d\bar{\gamma}} dx. \tag{3.14}$$

This and $dN^e/d\bar{\gamma} < 0$ imply that if housing and the public good are complements then the aggregate transportation cost increases with the level of the public good. A stronger condition is required to sign the effect on the differential land rent.

In summary,

Result 3.11:

(i) There exists $x^* \in (0, \bar{x}^e)$ such that $\dfrac{dR^* \langle x^* \rangle}{d\bar{\gamma}} = 0.$

If housing and the public good are complements:

(ii) $\dfrac{dD^e \langle 0 \rangle}{d\bar{\gamma}} < 0;\quad \dfrac{d\bar{x}^e}{d\bar{\gamma}} > 0.$

(iii) x^* is unique and such that $\dfrac{dR^*}{d\bar{\gamma}} \leq (\geq) 0$ as $x \leq (\geq) x^*$.

(iv) $\dfrac{d\overline{\text{ATC}}^e}{d\bar{\gamma}} > 0;\quad \dfrac{d\overline{\text{ADT}}^e}{d\bar{\gamma}} > 0.$

3.2.6 Initial Endowment

If we substitute $\bar{\Omega}$ for $\bar{\gamma}$ in (3.10) and use the same argument we conclude that the rent can neither increase nor decrease everywhere as the initial endowment increases. Therefore there exists at least one location $x^* \in (0, \bar{x}^e)$ where $R^e \langle x^* \rangle$ does not change with $\bar{\Omega}$. We then differentiate result 2.2 with respect to $\bar{\Omega}$ at x^* to obtain

$$\frac{d}{d\bar{\Omega}}\left(\frac{dR^e \langle x^* \rangle}{dx}\right) = \left(\frac{t}{H^{e2}}\frac{\partial h^e}{\partial \bar{U}}\frac{dv}{d\bar{\Omega}}\right)\Bigg|_{x=x^*} > 0. \tag{3.15}$$

The inequality follows because housing is normal and, since $R \langle x^* \rangle$ remains invariant as $\bar{\Omega}$ increases, the indirect utility level $v\left[R^e, \bar{\Omega} - T, \bar{\gamma}\right]$ must increase at x^*—hence $\bar{U}^e$ must increase. Consequently the equilibrium rent flattens at x^*, and there exists a unique $x^* \in (0, \bar{x}^e)$ such that $dR^e/d\bar{\Omega} \leq (\geq) 0$ as $x \leq (\geq) x^*$, which also implies $d\bar{x}^e/d\bar{\Omega} > 0$.[4] Therefore figure 3.10 also applies for an increasing initial endowment as well. Notice that although we have presented the effect of increasing initial endowment on the equilibrium rent profile after those of transportation technology and the public good, this important implication was the first of its kind to be discovered and presented in Wheaton's seminal paper (1974).

In order to gain some intuition about this result recall that as the initial endowment increases, so does income. Hence, using the normality of land, per capita consumption of land also increases. This, in turn, implies that the saving in housing cost associated with a move away from the centre, $H^e(dR^e/dx)$, must increase as the initial endowment increases. Since the transportation rate t remains unaffected, the condition of Muth in result 2.2 is upset: the total marginal cost of location $H^e(dR^e/dx) + t$ becomes negative, indicating that utility can be increased by moving away from the centre. Consequently, the demand for land declines in central locations and increases in peripheral locations.

Since income increases and the rent decreases within radius x^* as the initial endowment increases, housing consumption increases and, consequently, population density decreases within radius x^*. This follows from the normality of land and the observation that density is the inverse of land consumption. Outside radius x^*, both income and rent increase. Since these have opposing effects on the demand for land, we are unable to determine in a precise manner how population density behaves everywhere beyond x^* as the initial endowment increases. However, we know that population disperses in the sense that more individuals live outside any particular radius over the original $(0, \bar{x}^e]$ after the initial endowment has increased. In particular, from lemma 3.1 in the appendix

[4] Increasing urban area as the initial endowment increases can be verified directly from (2.11). Upon variation of $\bar{x}^e$ and $\bar{\Omega}$, we have

$$\frac{d\bar{x}^e}{d\bar{\Omega}} = -\left.\frac{dR^e}{d\bar{\Omega}}\Big/\frac{dR^e}{dx}\right|_{x=\bar{x}^e} > 0$$

because of result 2.1 and the observation that the equilibrium rent increases at the border as the initial endowment increases.

we have

$$\frac{dN^e}{d\bar{\Omega}} < 0 \text{ for } x \in (0, \bar{x}^e]. \tag{3.16}$$

For the exponential density function, we also know from (2.25) that the distance–decay parameter $\bar{\delta}$ decreases as the utility level increases. Since utility, in turn, increases with the initial endowment, it follows that the equilibrium distance–decay parameter decreases as the initial endowment increases. This explains why, on average, the urban population density surface is flatter in the case of more developed economies, an observation documented by, among others, Clark (1951), Muth (1969), Mills (1972), and Mills and Tan (1980).

Using (3.16), we can establish that the aggregate distance travelled increases with the initial endowment. Differentiating (3.8) with respect to $\bar{\Omega}$ gives

$$\frac{d\overline{\text{ADT}}^e}{d\bar{\Omega}} = -\int_{\mathcal{X}^e} \frac{dN^e}{d\bar{\Omega}} dx. \tag{3.17}$$

Our claim then follows from (3.16) and (3.17). Similarly, using (3.13) we obtain

$$\frac{d\overline{\text{ATC}}^e}{d\bar{\Omega}} = -\int_{\mathcal{X}^e} \frac{dN^e}{d\bar{\Omega}} t dx > 0. \tag{3.18}$$

A stronger condition is required to characterise the effect of $\bar{\Omega}$ on the equilibrium differential land rent (see the appendix).

In summary,

Result 3.12 (Wheaton (1974), Pines and Sadka (1986)):

(i) $\frac{dD^e\langle 0\rangle}{d\bar{\Omega}} < 0; \quad \frac{d\bar{U}^e}{d\bar{\Omega}} > 0; \quad \frac{d\bar{x}^e}{d\bar{\Omega}} > 0.$

(ii) There exists a unique $x^* \in (0, \bar{x}^e)$ such that $\frac{dR^e}{d\bar{\Omega}} \leq (\geq)\, 0$ as $x \leq (\geq)\, x^*$.

(iii) $\frac{d\overline{\text{ATC}}^e}{d\bar{\Omega}} > 0; \quad \frac{d\overline{\text{ADT}}^e}{d\bar{\Omega}} > 0.$

3.3 Theoretical Implications and Empirical Evidence

Tables 3.2 and 3.3 summarise the comparative statics results for the renter city ($\bar{k} = 0$). The former describes an open renter city and the latter a closed renter city. A row in those tables specifies the impact of parameters on a given variable. For example, a positive sign means that the total derivative of the variable on the row with respect to the parameter on the column is positive. A parenthesis in the case of $\bar{\gamma}$ for the closed renter city refers to the case where housing and the public good are complements. Absence of sign implies ambiguity.

TABLE 3.2. Comparative Statics of an Open Renter City.

	$\bar{\Omega}$	$\bar{R}$	$\bar{U}$	$\bar{\alpha}$	$\bar{\beta}$	$\bar{\gamma}$
$D^e\langle 0\rangle$	+	0	−	0	0	+
$D^e\langle \bar{x}^e\rangle$	+	0	−	+	0	+
$R^e\langle 0\rangle$	+	0	−	0	0	+
$R^e\langle \bar{x}^e\rangle$	+	0	−	+	0	+
$\bar{Y}^e$	+	0	0	0	0	0
$\bar{N}^e$	+	+	−	+	+	+
$\bar{x}^e$	+	0	−	+	0	+
$\overline{\text{DLR}}^e$	+	+	−	+	+	+
$\overline{\text{ATC}}^e$	+	+	−		+	+
$\overline{\text{ADT}}^e$	+	+	−	+	+	+

The comparative statics results of this chapter help to explain some aspects of how urban structure has typically evolved over time.[5] Urban change often occurs as an adjustment to general trends in the larger environment within which cities are found. Some general trends in the real world can be represented by changes in the exogenous parameters of our model. Thus comparative statics can be used for explaining observed urban adjustments to those general trends. In the remainder of this section we present three examples in which the comparative statics results we discussed are used to generate empirical hypotheses about urban adjustments to changing conditions. Our theoretical arguments imply that some of these adjustments must vary in a systematic way over different stages of urban development. In particular, as we mentioned in section 2.2.7, early stages of urbanisation are represented by the open city model and late stages by the closed city model. We can therefore evaluate the comparative statics of this chapter by using empirical evidence about how cities grow during the early and late stages of urbanisation.

[5] The comparative statics results for a closed mixed or owner city are presented in the appendix. They differ from those of the closed renter city in two ways. First, the impact of all the parameters on the density at the boundary of the city is ambiguous. Second, as in Pines and Sadka (1986), the effect of the population and the agricultural rent on the city area is ambiguous. In all these cases the ambiguity is caused by the additional income effect which is absent in the case of a renter city. However the effects of all the variables on utility, on the density and the rent at the center, and on the differential land rent, are the same for a closed renter, mixed or owner city.

TABLE 3.3. Comparative Statics of a Closed Renter City.

	$\bar{\Omega}$	$\bar{R}$	$\bar{N}$	$\bar{k}$	$\bar{\alpha}$	$\bar{\beta}$	$\bar{\gamma}$
$D^e\langle 0\rangle$	−	+	+	−	−	−	(−)
$D^e\langle \bar{x}^e\rangle$		+	+			−	
$R^e\langle 0\rangle$	−	+	+	−	−	−	(−)
$R^e\langle \bar{x}^e\rangle$	+	+	+	+	+	−	(+)
$\bar{Y}^e$	+	0	0	+		0	0
$\bar{U}^e$	+	−	−	+	+	+	+
$\bar{x}^e$	+	−	+	+	+	−	(+)
$\overline{\text{DLR}}^e$	(+)	(−)		(+)			(+)
$\overline{\text{ATC}}^e$	+	−	+	+			(+)
$\overline{\text{ADT}}^e$	+	−	+	+	+		(+)

3.3.1 Agricultural Rent

The effect of economic growth on agricultural land rent is ambiguous. One however might expect that, with economic growth, higher income would cause the relative prices of agricultural products to decline—if indeed the income elasticity of food is lower than that of urban products. Under these circumstances, economic growth suppresses agricultural rent in relative terms. A lower $\bar{R}$, according to table 3.2, does not affect population density during early stages of urbanisation. During late stages, however, it contributes to suburbanisation according to table 3.3. Moreover the effect of urban land reclamation, which increases the supply of land through an increase in $\bar{\beta}$, is similar to the effect of decreasing agricultural rent in those tables. The main comparative statics difference is to be found in the closed model where, although decreasing agricultural rent is bound to cause suburbanisation, a larger supply of land at any distance from the centre may even reduce the urban area. This will happen if the higher demand for land induced by lower bid rents is dominated by the corresponding higher supply.

For the negative exponential function, we know from (2.25) that the equilibrium density gradient decreases as the corresponding utility level increases. This and table 3.3 imply that the equilibrium density gradient of a closed city becomes steeper as the agricultural rent increases. It is reasonable to expect that, on average, the agricultural rent is higher in regions where overall population densities are higher. Thus, on average, one should expect steeper density gradients in regions of higher overall population densities. This is supported by facts. For example, Mills and Tan (1980, p.317) report an average gradient of .67 per km for Korean metropolitan areas in 1970, while the corresponding figure for the USA is .21 per km according to Edmonston (1975, p.67) . The average density in Korea was sixteen times that of the USA, while per capita income in the USA was seventeen times that of Korea (Mills and Tan (1980, p.319)).

Although both such density and income differences may account for the steeper Korean gradient in our context, there is some indirect evidence suggesting that average density in itself must have a significant impact on the gradient. For example, Mills and Tan (1980, pp. 319–320) report similar average gradients for Brazil and the USA, an average density for Brazil equal to one–half of the USA density and a ninefold income difference in favour of the USA.

3.3.2 Economic Growth and Suburbanisation

For cities, economic growth has been characterised by increased productivity and accumulation of capital, including housing and infrastructure. This general trend can be reflected by exogenous increases in $\bar{\Omega}, \bar{\alpha}$ and $\bar{\gamma}$. The increase in $\bar{\Omega}$ is the result of both increased productivity and the accumulation of human and non–human capital. The increase in $\bar{\alpha}$ reflects the remarkable improvements in transportation technology and the sustained growth of urban transportation infrastructure. Finally, the increase in $\bar{\gamma}$ could represent the provision of better public goods.

Economic growth applies to both less developed and developed economies. For a particular region, growth shifts the economy from a less developed, primarily agrarian state to a developed, primarily urban state. During the early stages of urbanisation, table 3.2 suggests that increases in $\bar{\Omega}, \bar{\alpha}$ and $\bar{\gamma}$ compound each other to create a larger, more densely populated city. Continuing economic growth eventually brings the region to a late stage of urbanisation, for which table 3.3 applies. This table suggests that increases in $\bar{\Omega}, \bar{\alpha}$ and $\bar{\gamma}$ once again compound each other to raise utility and to cause suburbanisation, that is, a lowering of population density and residential land values at the centre, an increasing demand for residential land at the periphery and consequent urban sprawl.[6] Our comparative statics of the open and closed city models further indicate that the same external changes also generate an increase in the demand for transportation and, usually, in the aggregate cost of transportation and the differential land rent. The exception arises when sprawl is caused by a lower transportation rate. Then, higher quantity demanded for transportation is countered by lower price to generate ambiguity; and the same is true for the differential land rent in a closed city, since better transportation does not necessarily imply higher per capita income for $\bar{k} > 0$, only higher disposable income.

Significant comparative statics implications arise from the systematic differences in the pattern of urban growth for different stages of a region's development. Whereas rising per capita income and better provision of public goods raise density everywhere in the open city model, they cause suburbanisation in the closed city model. This sharp difference is blurred to some extent by the effect of a transportation improvement which causes a flattening of the rent and

---

[6] For the closed city model, increasing share in profits from land development is another cause of suburbanisation. The increase in $\bar{k}$ can perhaps be taken to account for the higher rate of home–ownership observed in more developed economies.

density gradients in both open and closed cities. Since, however, an increasing $\bar{\alpha}$ cannot lower population density anywhere in an open city, the flattening of the density gradient must be stronger during late stages of urbanisation.

These implications can be readily verified in the case of the negative exponential function. On the one hand we know from (2.25) that the equilibrium density gradient of an open renter city remains unchanged as the initial endowment increases because utility is parametric. On the other hand, using table 3.3, we know that the gradient of a closed renter city flattens with rising initial endowment because the equilibrium density gradient decreases as the corresponding utility level increases.[7] This systematic difference between less developed and developed economies has been verified by, among others, Berry, Simmons and Tennant (1963) who compared data of western and non–western cities from various sources. It is interesting to note that in their early study, which was done before the comparative statics of this chapter were known, Berry et al. (p. 404) attributed such systematic differences to "...the inverted locational patterns of socioeconomic groups within Western and non–Western cities, and attendant contrasts in demands for residential land"—rather than to differences in their stage of economic development.

A similar inference across time periods, rather than across regions, can be drawn from Edmonston (1975) who has undertaken the most detailed study of density gradients in the USA. Using the estimation method proposed by Mills (1972), Edmonston (1975, p. 67) found that the average density gradient for metropolitan areas of the USA has remained approximately the same at .46 per km until 1930, has increased to .54 per km during the 1930s and has steadily declined since then to reach .21 per km in 1970. Within our framework, this roughly implies a corresponding transition from the early stages toward the late stages of urbanisation for the USA.[8] More generally, suburbanisation appears to be a global phenomenon which has accelerated during the second half of the twentieth century. This conclusion is supported by a large number of diverse studies including Davis (1965), Clark (1967) and Mills and Tan (1980) for cities around the world, Mills and Ohta (1976) and Glickman (1979) for Japan, and Ingram and Carroll (1981) for Latin America.

3.3.3 Population Growth and Suburbanisation

Global population growth causes larger cities and increases the demand for agricultural output. These exogenous effects can be represented in our framework for a closed renter city model by an increase in both $\bar{N}$ and $\bar{R}$ which, according to table 3.3, work in the same direction to lower utility and to raise popula-

[7] Using once again (2.25), we also know that improving transportation technology flattens the gradient in both the open and closed city cases.

[8] Similar conclusions can be drawn from the evidence presented in the early study of Bogue (1953), as well as in Goldberg and Mercer (1986).

tion density and residential land values.[9] Although the urban area of a closed city expands with increasing population, so that the aggregate distance travelled increases, *average* distance may still decrease.[10] To the extent that average distance travelled serves as a criterion for comparing spatial distributions (see Mills (1972)), this suggests that larger cities may become more concentrated, rather than dispersed, at equilibrium.

For the negative exponential function, using table 3.3, we know that the gradient of a closed renter city steepens with rising population because the equilibrium density gradient decreases as the corresponding utility level increases. However both Muth (1969) and Mills (1972) observed that larger North American cities are, on average, more suburbanised. In other words, that the density gradient is a decreasing rather than an increasing function of population size. There have been several attempts to explain this inconsistency between theory and empirical evidence. Muth (1969) maintains that a less–than–unit elasticity of substitution of land for other factors in the production of housing could explain this inconsistency. Papageorgiou (1971) and Mills (1972) point out to the polycentric structure of large cities. As we explain in chapter eight, since the negative exponential function describes the general trend away from the main centre of such cities, increasing relative importance of peripheral subcentres in larger cities may account for the flattening of the general density trend. Finally Adler (1987) uses two distinct gradients: a steeper one near the centre produced by renters and a flatter, peripheral one produced by owners. Since larger cities imply a higher relative ownership share according to Adler (1987), the average gradient flattens.

The above studies invoke additional factors in order to explain the alleged inconsistency between theory and reality. By contrast, in chapter five, we discuss an extension which resolves this serious problem within the context of the monocentric city model. In particular we introduce production characterised by scale economies. Under this extension, the present comparative statics change and a *negative* correlation between the density gradient and population size becomes possible.

---

[9]Urban immigration is also caused by a lower exogenous utility level in the open city model. There, the city becomes larger because of a difference in urban and regional utility levels favouring the city.

[10]For the exponential density function (2.27) and for $\bar{R} = 0$, we know from Ashenfelter (1976) and White (1977) that the average distance travelled is given by

$$\frac{\overline{\mathrm{ADT}}^e}{\bar{N}} = \frac{2}{\bar{\delta}}$$

in the limit where $\bar{x}^e \rightarrow \infty$. Upon differentiation and use of result 3.13 in the appendix, we have

$$\frac{d}{d\bar{N}}\left(\frac{\overline{\mathrm{ADT}}^e}{\bar{N}}\right) = -\frac{2}{\bar{\delta}^2}\frac{d\bar{\delta}}{d\bar{U}^e}\frac{d\bar{U}^e}{d\bar{N}} = \frac{2}{\bar{\alpha}(\bar{\delta}\bar{U}^e)^2}\frac{d\bar{U}^e}{d\bar{N}} < 0.$$

3.4 Appendix: Closed Mixed and Owner Cities

In this appendix we study in detail the comparative statics of a closed city in which at least part of the aggregate profit from urban land transactions accrues to the city residents, i.e. $\bar{k} \in (0, 1]$. Our exposition is based on Pines and Sadka (1986). It is convenient to determine first the impact of all parameters on the equilibrium utility level. This is done in section 3.4.1. Each subsequent section of the appendix discusses the impact of a single parameter on all remaining variables.

3.4.1 Effects on Utility

Using (2.6) and (2.12(a)) we obtain upon total differentiation

$$\begin{aligned}&\frac{\partial e^e}{\partial R}dR^e + \frac{\partial e^e}{\partial \bar{U}}d\bar{U}^e + \frac{\partial e}{\partial \bar{\gamma}}d\bar{\gamma} + \frac{\partial T}{\partial \bar{\alpha}}d\bar{\alpha} = \\ &d\bar{\Omega} - \frac{\bar{k}}{\bar{N}^2}\overline{\mathrm{DLR}}^e d\bar{N} + \frac{\bar{k}}{\bar{N}}d\overline{\mathrm{DLR}}^e + \frac{1}{\bar{N}}\overline{\mathrm{DLR}}^e d\bar{k}\end{aligned} \tag{3.19}$$

Multiplying both sides of (3.19) by n^e, integrating the result over $\mathcal{X}^e$ and taking into account $n^e \partial e^e / \partial R = n^e H^e = \theta$ by (2.15) and (2.9), we get

$$\begin{aligned}&(1-\bar{k})d\overline{\mathrm{DLR}}^e + \left(\int_{\mathcal{X}^e} n^e \frac{\partial e^e}{\partial \bar{U}}dx\right)d\bar{U}^e = \\ &\bar{N}d\bar{\Omega} - \left(\int_{\mathcal{X}^e} n^e \frac{\partial e}{\partial \bar{\gamma}}dx\right)d\bar{\gamma} - \frac{\bar{k}}{\bar{N}}\overline{\mathrm{DLR}}^e d\bar{N} - \Theta\left\langle \bar{x}^e\right\rangle d\bar{R} \\ &-\left(\int_{\mathcal{X}^e} n^e \frac{\partial T}{\partial \bar{\alpha}}dx\right)d\bar{\alpha} + \left(\int_{\mathcal{X}^e} \frac{\partial \theta}{\partial \bar{\beta}}(R^e - \bar{R})dx\right)d\bar{\beta} + \overline{\mathrm{DLR}}^e d\bar{k},\end{aligned} \tag{3.20}$$

where

$$d\overline{\mathrm{DLR}}^e = \left(\int_{\mathcal{X}^e} \frac{\partial \theta}{\partial \bar{\beta}}(R^e - \bar{R})dx\right)d\bar{\beta} + \int_{\mathcal{X}^e} \theta dR^e dx - \Theta\left\langle \bar{x}^e\right\rangle d\bar{R}. \tag{3.21}$$

We can employ (3.20) to assess the impact of changing parameters on the equilibrium utility level of a closed owner city. Recall that since $\bar{U}$ and $\bar{\gamma}$ are parameters of the minimum expenditure function, the envelope theorem gives

$$\begin{aligned}\frac{\partial e}{\partial \bar{U}} &= 1/\left(\frac{\partial u}{\partial Z^i}\right) && (a) \\ \frac{\partial e}{\partial \bar{\gamma}} &= -\left(\frac{\partial u}{\partial \bar{\gamma}}\right)/\left(\frac{\partial u}{\partial Z^i}\right). && (b)\end{aligned} \tag{3.22}$$

Using (3.22) in conjunction with (2.11) and assumptions 2.1 and 2.2 on (3.20) we can establish that, for a closed owner city, the derivatives $d\bar{U}^e/d\bar{\Omega}$, $d\bar{U}^e/d\bar{\alpha}$, $d\bar{U}^e/d\bar{\beta}$ and $d\bar{U}^e/d\bar{\gamma}$ are positive while the derivatives $d\bar{U}^e/d\bar{N}$ and $d\bar{U}^e/d\bar{R}$ are negative (see Pines and Sadka (1986)). As we have argued in chapter three, the same is true for a closed renter city. However, for a closed mixed city, we cannot sign these derivatives directly from (3.20) because the loss of resources to absentee investors at the rate of $(1-\bar{k})d\overline{\mathrm{DLR}}^e$ must also be taken into account.

We will establish that the same results hold for the entire $\bar{k} \in [0,1]$. Toward this end, we follow the approach developed by Wheaton (1974) which has proven to be very useful for the comparative statics of this appendix. Starting with (2.10), we have

$$\begin{aligned} \bar{N} &= \int_{\mathcal{X}^e} n^e dx \\ &\underset{(2.8)}{=} \int_{\mathcal{X}^e} \frac{\theta}{H^e} dx \\ &\underset{(\text{Result } 2.2)}{=} -\int_{\mathcal{X}^e} \sigma \frac{dR^e}{dx} dx. \end{aligned} \tag{3.23}$$

Upon total differentiation, we get

$$\begin{aligned} d\bar{N} = & \\ & -\sum\nolimits_i \left(\int_{\mathcal{X}^e} \frac{\partial \sigma}{\partial i} \frac{dR^e}{dx} dx \right) di \\ & -\sum\nolimits_j \left(\int_{\mathcal{X}^e} \sigma \frac{\partial}{\partial j} \left(\frac{dR^e}{dx} \right) dx + \left(\sigma \frac{dR^e}{dx} \right) \bigg|_{x=\bar{x}^e} \frac{\partial \bar{x}^e}{\partial j} \right) dj. \end{aligned} \tag{3.24}$$

where $i = \bar{\alpha}, \bar{\beta}$ and $j = \bar{\Omega}, \bar{N}, \bar{k}, \bar{\alpha}, \bar{\beta}$ and $\bar{\gamma}$. Integration by parts gives

$$\begin{aligned} d\bar{N} = & \\ & -\sum_i \left(\int_{\mathcal{X}^e} \frac{\partial \sigma}{\partial i} \frac{dR^e}{dx} dx \right) di + \sum_j \left(\left(\sigma \frac{\partial R^e}{\partial j} \right) \bigg|_{x=0} + \int_{\mathcal{X}^e} \frac{\partial \sigma}{\partial x} \frac{\partial R^e}{\partial j} \right) dj \\ & - \sigma \langle \bar{x}^e \rangle \sum_j \left(\frac{\partial R^e \langle \bar{x}^e \rangle}{\partial j} + \frac{dR^e \langle \bar{x}^e \rangle}{dx} \frac{\partial \bar{x}^e}{\partial j} \right) dj. \end{aligned} \tag{3.25}$$

Taking into account that

$$\sum_j \left(\frac{\partial R^e \langle \bar{x}^e \rangle}{\partial j} + \frac{dR^e \langle \bar{x}^e \rangle}{dx} \frac{\partial \bar{x}^e}{\partial j} \right) dj = d\bar{R} \tag{3.26}$$

leads to

$$d\bar{N} = -\sum_i \left(\int_{\mathcal{X}^e} \frac{\partial \sigma}{\partial i}\frac{dR^e}{dx}dx\right) di + \sum_j \left(\left(\sigma\frac{dR^e}{dj}\right)\bigg|_{x=0} + \int_{\mathcal{X}^e} \frac{\partial \sigma}{\partial x}\frac{dR^e}{dj}dx\right) dj - \sigma\left\langle \bar{x}^e \right\rangle d\bar{R}. \tag{3.27}$$

Since $\partial\theta/\partial\bar{\beta} \geq 0$ by assumption 2.1 and $\partial t/\partial\bar{\alpha} < 0$ by assumption 2.2, it must be that $\partial\sigma/\partial i \geq 0$. Since in addition $dR^e/dx < 0$ by result 2.3, we conclude that the first term on the RHS of (3.27) is nonnegative. The sign of the second term, on the other hand, depends on dR^e/dj.

Now hold $\bar{N}$ and $\bar{R}$ fixed. If one of $\bar{\Omega}, \bar{k}, \bar{\alpha}, \bar{\beta}$ or $\bar{\gamma}$ increases, (3.27) and assumption 2.3 implies that the equilibrium rent must decrease somewhere in $\mathcal{X}^e$. Consider such a location. Assume that the equilibrium utility level does not increase there as one of $\bar{\Omega}, \bar{k}, \bar{\alpha}, \bar{\beta}$ or $\bar{\gamma}$ increases. Under these circumstances (3.20) implies that $\overline{\mathrm{DLR}}^e$ must increase. Therefore equilibrium income must also increase. We are forced to infer that since the price of land is lower and income is higher, utility must have increased at the location considered—a contradiction. It follows that the results of Pines and Sadka (1986), on the impact of $\bar{\Omega}, \bar{\alpha}, \bar{\beta}$ and $\bar{\gamma}$ on $\bar{U}^e$, apply for $\bar{k} \in [0, 1]$; and that $d\bar{U}^e/d\bar{k} > 0$ for $0 \leq \bar{k} < 1$.

We next turn to the impact of $\bar{N}$ and $\bar{R}$ on the equilibrium utility level when $0 \leq \bar{k} < 1$. Suppose that $\bar{U}^e$ does not decrease when one of $\bar{N}, \bar{R}$ increases. Then $\bar{Y}^e$ cannot increase because, by (3.20), $\overline{\mathrm{DLR}}^e$ must decrease. On the other hand if one of $\bar{N}, \bar{R}$ increases, (3.27) and assumption 2.3 imply that the equilibrium rent must increase somewhere in $\mathcal{X}^e$. At such location utility must decrease because the price of land is higher and income is not—a contradiction. In summary,

Result 3.13 (Wheaton (1974), Pines and Sadka (1986)): For $\bar{k} \in [0, 1]$

(i) $\dfrac{d\bar{U}^e}{d\bar{\Omega}} > 0; \quad \dfrac{d\bar{U}^e}{d\bar{N}} < 0; \quad \dfrac{d\bar{U}^e}{d\bar{R}} < 0.$

(ii) $\dfrac{d\bar{U}^e}{d\bar{k}} > 0; \quad \dfrac{d\bar{U}^e}{d\bar{\alpha}} > 0; \quad \dfrac{d\bar{U}^e}{d\bar{\beta}} > 0; \quad \dfrac{d\bar{U}^e}{d\bar{\gamma}} > 0.$

This result will be used to determine the effects of each parameter, in turn, on the remaining variables.

3.4.2 Initial Endowment

When only the initial endowment varies (3.27) is written as

$$\left(\sigma\frac{dR^e}{d\bar{\Omega}}\right)\bigg|_{x=0} + \int_{\mathcal{X}^e} \frac{\partial\sigma}{\partial x}\frac{dR^e}{d\bar{\Omega}}dx = 0. \tag{3.28}$$

This, together with assumption 3.1, implies that either $dR^e/d\bar{\Omega} = 0$ everywhere in $\mathcal{X}^e$ or that $dR^e/d\bar{\Omega}$ changes sign. Then since R^e is continuous in $\mathcal{X}^e$, there is $x^* \in (0, \bar{x}^e)$ such that $dR^e \langle x^* \rangle / d\bar{\Omega} = 0$. We then apply the corresponding argument in section 3.2.6 to establish that x^* is unique and that the equilibrium rent flattens at x^* as in figure 3.6.

Consider now the impact of $\bar{\Omega}$ on the indirect utility function $v[R^e, \bar{Y}^e - T, \bar{\gamma}]$ at x^*. Total differentiation simply gives

$$\frac{d\bar{U}^e}{d\bar{\Omega}} = \left(\frac{\partial v^e}{\partial \bar{Y}} \frac{d\bar{Y}^e}{d\bar{\Omega}} \right) \Bigg|_{x=x^*} > 0 \tag{3.29}$$

by result 3.7. Then $d\bar{Y}^e/d\bar{\Omega} > 0$ follows from the property of indirect utility functions $\partial v^e / \partial \bar{Y} > 0$. Therefore, as in the case of a closed renter city, increasing initial endowment implies increasing income, and the intuition provided in section 3.2.6 about figure 3.6 also applies in the more general case of this appendix.

We next prove that the population disperses after the initial endowment has increased for any $\bar{k} \in [0, 1]$:

Lemma 3.1 (Pines and Sadka (1986)): $\dfrac{dN^e}{d\bar{\Omega}} < 0$ for $x \in (0, \bar{x}^e]$.

Proof: We know there exists $x^* \in (0, \bar{x}^e)$, such that $dR^e/d\bar{\Omega} \leq (\geq) 0$ as $x \leq (\geq) x^*$. Therefore, since $d\bar{U}^e/d\bar{\Omega} > 0$ by result 3.13, $dH^e/d\bar{\Omega} > 0$ for $x \in (0, x^*]$ by assumption 2.2. It follows that $dn^e/d\bar{\Omega} < 0$ and $dN^e/d\bar{\Omega} < 0$ for $x \in (0, x^*]$.

Suppose there exists $x^1 \in (x^*, \bar{x}^e]$ such that $dN^e \langle x^1 \rangle / d\bar{\Omega} \geq 0$. Using (2.1),

$$\int_0^{x^1} \frac{\partial \sigma}{\partial x} \frac{dR^e}{d\bar{\Omega}} dx + \int_{x^1}^{\bar{x}^e} \frac{\partial \sigma}{\partial x} \frac{dR^e}{d\bar{\Omega}} dx = - \left(\sigma \frac{\partial R^e}{\partial I} \right) \Bigg|_{x=0}. \tag{3.30}$$

Integrating by parts (3.30) and after cancellation of terms we have

$$\left(\sigma \frac{\partial R^e}{\partial I} \right) \Bigg|_{x=x^1} - \int_0^{x^1} \sigma \frac{d}{dx} \left(\frac{dR^e}{d\bar{\Omega}} \right) dx + \int_{x^1}^{\bar{x}^e} \frac{\partial \sigma}{\partial x} \frac{dR^e}{d\bar{\Omega}} dx = 0. \tag{3.31}$$

Now

$$\begin{aligned} \frac{d}{dx} \left(\frac{dR^e}{d\bar{\Omega}} \right) &= \frac{d}{d\bar{\Omega}} \left(\frac{dR^e}{dx} \right) \\ &\underset{\text{(Result 2.2)}}{=} -t \frac{d}{d\bar{\Omega}} \left(\frac{1}{H^e} \right) \\ &= -\frac{t}{\theta} \frac{d}{d\bar{\Omega}} \left(\frac{\theta}{H^e} \right) \\ &\underset{(2.8)}{=} -\frac{1}{\sigma} \frac{dn^e}{d\bar{\Omega}}. \end{aligned} \tag{3.32}$$

Therefore (3.31) can be written as

$$\begin{aligned} \left(\sigma\frac{dR^e}{d\bar{\Omega}}\right)\Bigg|_{x=x^1} + \int_0^{x^1}\frac{dn^e}{d\bar{\Omega}}dx + \int_{x^1}^{\bar{x}^e}\frac{\partial\sigma}{\partial x}\frac{dR^e}{d\bar{\Omega}}dx &= \\ \left(\sigma\frac{dR^e}{d\bar{\Omega}}\right)\Bigg|_{x=x^1} + \frac{dN^e\langle x^1\rangle}{d\bar{\Omega}} + \int_{x^1}^{\bar{x}^e}\frac{\partial\sigma}{\partial x}\frac{dR^e}{d\bar{\Omega}}dx &= 0. \end{aligned} \tag{3.33}$$

Since $x^1 > x^*$ by construction, $dR^e/d\bar{\Omega} > 0$ for $x \in [x^1, \bar{x}^e]$. It follows that the first term on the LHS of (3.33) is positive; the second is nonnegative by premise; and the third is also nonnegative because of assumption 3.1. Therefore the LHS of (3.33) must be positive—a contradiction. □

The effects of the initial endowment on aggregate distance travelled and aggregate transport cost can be readily determined using lemma 3.1 as in section 3.2.6 of chapter three. We conclude that result 3.7 holds in the general case $\bar{k} \in [0,1]$.

It remains to determine the effect initial endowment on differential land rent. Toward this end we need the following lemma.

Lemma 3.2: If $f[x]$ and $g[x]$ satisfy

(i) $f > 0$ and $f' > 0$

(ii) $\int_{\mathcal{X}^e} g dx \geq 0$

(iii) $g[x] \leq (\geq)\, 0$ as $x \leq (\geq)\, x^*$ for $0 < x^* < \bar{x}^e$

then $\int_{\mathcal{X}^e} fg dx > 0.$

PROOF: First notice that

$$\int_{\mathcal{X}^e} f[x^*]g[x]dx = f[x^*]\int_{\mathcal{X}^e} g[x]dx \geq 0 \tag{3.34}$$

by (i) and (ii). Secondly, since (i) also implies $f[x] - f[x^*] \leq (\geq)\, 0$ as $x \leq (\geq)\, x^*$, it follows from (iii) that

$$(f[x] - f[x^*])g[x] > 0 \text{ for } x \in [0, \bar{x}^e]. \tag{3.35}$$

Now

$$\begin{aligned} \int_{\mathcal{X}^e} f[x]g[x]dx &= \\ \int_{\mathcal{X}^e}(f[x] - f[x^*])g[x]dx + \int_{\mathcal{X}^e} f[x^*]g[x]dx &> 0 \end{aligned} \tag{3.36}$$

by (3.34) and (3.35). □

Using the definition of differential land rent, we have

$$\begin{aligned}\frac{d\overline{\mathrm{DLR}}^e}{d\bar{\Omega}} &= \int_{\mathcal{X}^e} \theta \frac{dR^e}{d\bar{\Omega}} dx + \left(\theta(R^e - \bar{R})\right)\Big|_{x=\bar{x}^e} \frac{d\bar{x}^e}{d\bar{\Omega}} \\ &\underset{(2.11)}{=} \int_{\mathcal{X}^e} \frac{\theta}{\partial\sigma/\partial x} \frac{\partial \sigma}{\partial x} \frac{dR^e}{d\bar{\Omega}} dx. \end{aligned} \tag{3.37}$$

Let $f[x] \equiv \theta/(\partial\sigma/\partial x)$ and $g[x] \equiv (\partial\sigma/\partial x)(dR^e/d\bar{\Omega})$. Because of assumptions 2.1 and 2.3, we have $f[x] > 0$. If, in addition,

$$\frac{\partial}{\partial x}\left(\frac{\theta}{\partial\sigma/\partial x}\right) \leq 0 \tag{3.38}$$

is satisfied, we have $f'[x] > 0$ and (i) of lemma 3.2 is satisfied. Condition (3.38) holds, for example, in the case of a circular city under linear transportation cost. On the other hand, (2.1) and the observation that $dR^e\langle 0\rangle/d\bar{\Omega} < 0$ (see figure 3.6) ascertain that (ii) of lemma 3.2 is satisfied. Finally, we have already established that there exists a unique $x^* \in (0, \bar{x}^e)$, such that $dR^e/d\bar{\Omega} \leq (\geq) 0$ as $x \leq (\geq) x^*$. Therefore, by assumption 2.3, (iii) of lemma 3.2 is also satisfied. Under these circumstances, the RHS of (3.37) is positive so that differential land rent increases with increasing initial endowment.

3.4.3 Population Size

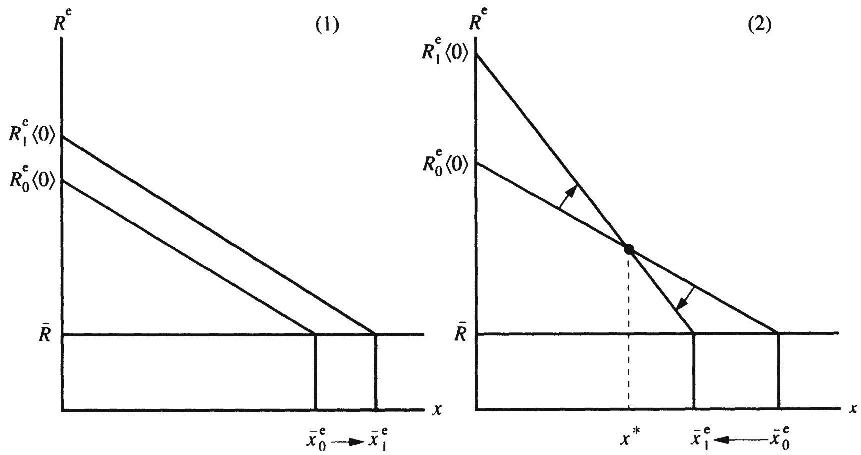

FIGURE 3.12. Effects of Increasing Population Size on a Closed City.

For urban types other than a closed renter city some effects of increasing population are ambiguous. In particular, the equilibrium rent may either increase everywhere (figure 3.12(1)) or increase at the centre and decrease at the boundary (figure 3.12(2)). To see this, vary only population in (3.27) and obtain

$$\left(\sigma\frac{dR^e}{d\bar{N}}\right)\bigg|_{x=0}+\int_{\mathcal{X}^e}\frac{\partial\sigma}{\partial x}\frac{dR^e}{d\bar{N}}dx=1. \tag{3.39}$$

Thus equilibrium rent cannot decline everywhere as population increases. If it changes sign, there exists a unique $x^* \in (0,\bar{x}^e)$ such that $dR^e/d\bar{N} \leq (\geq)\, 0$ as $x \leq (\geq)\, x^*$. This can be seen upon differentiation of result 2.2 at x^*, which gives

$$\frac{d}{d\bar{N}}\left(\frac{dR^e\langle x^*\rangle}{dx}\right)=\left(\frac{t}{H^{e2}}\frac{\partial h^e}{\partial\bar{U}}\frac{d\bar{U}^e}{d\bar{N}}\right)\bigg|_{x=x^*}<0, \tag{3.40}$$

that is, the equilibrium rent gradient steepens. In order to establish the effect of increasing population size on the spatial extent of the city, vary $\bar{x}^e$ and N^e in (2.11) to obtain

$$\frac{d\bar{x}^e}{d\bar{N}}=-\left(\frac{dR^e}{d\bar{N}}/\frac{dR^e}{dx}\right)\bigg|_{x=\bar{x}^e}. \tag{3.41}$$

Thus since $dR^e\langle\bar{x}^e\rangle/dx<0$, the city expands or retracts according to whether the equilibrium rent increases or decreases at the border. That both possibilities can in fact occur, has been shown by numerical example in Pines and Sadka (1986). A city retracting in spite of an increased demand for accommodation defies, at first glance, intuition. However, as pointed out by Pines and Sadka, such behaviour can be accounted for through the income effect generated by an increasing population. In particular, result 3.13 implies that increasing population operates to reduce real income with respect to the utility level. Now, as discussed in section 3.4.2, a reduced initial endowment lowers the rent at the boundary, hence it causes the city to retract as the per capita demand for land decreases. If this effect is strong enough to overcome the increased demand for accommodation, figure 3.12(2) applies.

Although differential land rent increases in the case of figure 3.12(1), we cannot determine the effect on income when $\bar{k}>0$ since the urban share in profits from the development of land is now distributed among a larger number of individuals. In the case of figure 3.12(2), however, income must decrease since the decline in utility at x^* can occur only through an income decrease. Finally, differentiating aggregate distance travelled gives

$$\frac{d\overline{\mathrm{ADT}}^e}{d\bar{N}}=\int_{\mathcal{X}^e}\frac{dn^e}{d\bar{N}}x\,dx+n^e\langle\bar{x}^e\rangle\,\bar{x}^e\frac{d\bar{x}^e}{d\bar{N}}, \tag{3.42}$$

while differentiating aggregate transportation cost gives

$$\frac{d\overline{\mathrm{ATC}}^e}{d\bar{N}}=\int_{\mathcal{X}^e}\frac{dn^e}{d\bar{N}}T\,dx+n^e\langle\bar{x}^e\rangle\,T\langle\bar{x}^e\rangle\,\frac{d\bar{x}^e}{d\bar{N}}. \tag{3.43}$$

If the city expands both aggregate distance travelled and aggregate transportation cost increase as population increases. This happens because, with declining utility and increasing rent, the consumption of land per capita must decrease—hence the number of individuals at a particular distance from the centre must increase. If on the other hand the city retracts, increasing population has an ambiguous effect on these two aggregate measures.

In summary,

Result 3.14 (Wheaton (1974), Pines and Sadka (1986)): For $\bar{k} \in [0,1]$

(i) $\dfrac{dD^e\langle 0\rangle}{d\bar{N}} > 0.$

(ii) Either (1) $\dfrac{dR^e}{d\bar{N}} > 0$ for $x \in \mathcal{X}^e$

or (2) there exists a unique $x^* \in (0, \bar{x}^e)$ such that $\dfrac{dR^e}{d\bar{N}} \geq (\leq) 0$ as $x \leq (\geq) x^*$.

(iii) In case (ii).1:
$$\begin{cases} \dfrac{dD^e\langle \bar{x}^e\rangle}{d\bar{N}} > 0; & \dfrac{d\bar{x}^e}{d\bar{N}} > 0. \\ \dfrac{d\overline{\mathrm{DLR}}^e}{d\bar{N}} > 0; & \dfrac{d\overline{\mathrm{ATC}}^e}{d\bar{N}} > 0; \\ \dfrac{d\overline{\mathrm{ADT}}^e}{d\bar{N}} > 0. & \end{cases}$$

(iv) In case (ii).2: $\dfrac{d\bar{Y}^e}{d\bar{N}} < 0;\quad \dfrac{d\bar{x}^e}{d\bar{N}} < 0.$

3.4.4 Agricultural Rent

In this section we show that result 3.8 holds for $\bar{k} > 0$ with only one exception. Namely, the equilibrium rent may decrease rather than increase at the boundary. As Pines and Sadka (1986) have shown by example both possibilities described in figure 3.13 can occur.

When only the agricultural rent varies (3.27) can be written as

$$\left(\sigma\frac{dR^e}{d\bar{R}}\right)\Big|_{x=0} + \int_{\mathcal{X}^e} \frac{\partial\sigma}{\partial x}\frac{dR^e}{d\bar{R}}dx - \sigma\langle\bar{x}^e\rangle =$$
$$\left(\left(\sigma\frac{dR^e}{d\bar{R}}\right)\Big|_{x=0} - 1\right) + \int_{\mathcal{X}^e} \frac{\partial\sigma}{\partial x}\left(\frac{dR^e}{d\bar{R}} - 1\right)dx = 0. \tag{3.44}$$

Since $\sigma > 0$, (3.44) implies that the rent cannot decline everywhere as the agricultural rent increases. Thus $dR^e/d\bar{R} > 0$ at least somewhere. If it is not

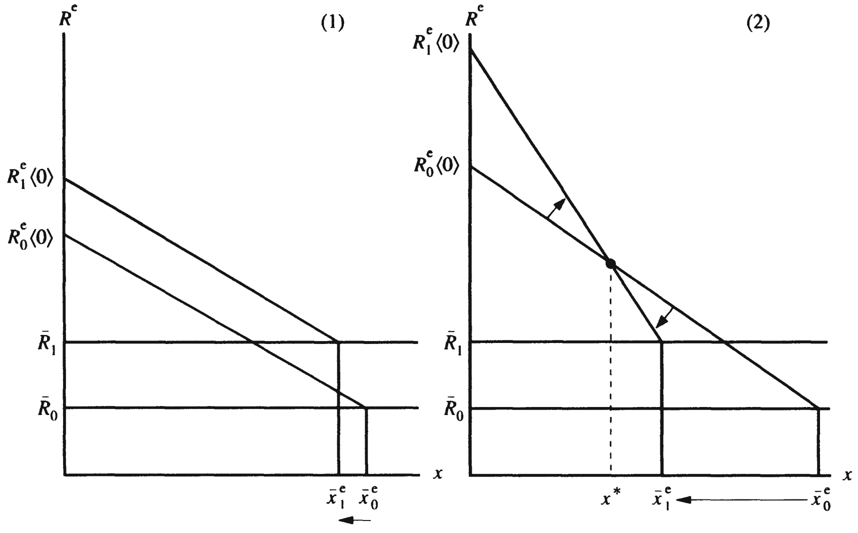

FIGURE 3.13. Effects of Increasing Agricultural Rent on a Closed City.

everywhere, there is $x^* \in (0, \bar{x}^e)$ such that $dR^e \langle x^* \rangle / d\bar{R} = 0$. Differentiating the result 2.2 with respect to $\bar{R}$ at x^* we obtain

$$\frac{d}{d\bar{R}}\left(\frac{\partial R^e \langle x^* \rangle}{\partial x}\right) = \left.\left(\frac{t}{H^{e2}}\frac{\partial h^e}{\partial \bar{U}}\frac{d\bar{U}^e}{d\bar{R}}\right)\right|_{x=x^*} < 0 \tag{3.45}$$

because of assumptions 2.3 and 2.4, and result 3.13. As before x^* must be unique. Hence if $dR^e/d\bar{R}$ changes sign the rent gradient becomes steeper when the agricultural rent increases. Clearly, in this case, the city retracts (figure 3.13(2)). The same holds when $dR^e/d\bar{R}$ is positive everywhere: since utility decreases and the price of land increases at the same time, the compensated demand function implies that the demand for land decreases everywhere—hence n^e increases everywhere. Thus, using (2.10), $\bar{x}^e$ must decrease as $\bar{R}$ increases. This case is illustrated in figure 3.13(1). Furthermore, in any case, central locations are associated with both higher rent and lower utility. Therefore population density increases at the centre.

When the case of figure 3.13(2) applies differential land rent must decline as the agricultural rent increases. This happens because utility decreases everywhere, including x^*, where the price of land remains unaltered. Therefore, using once more the indirect utility function, income must decrease as the agricultural rent increases. From (2.12) it follows that differential land rent also declines. Next we establish that the same is true when the rent increases everywhere. This can be seen from

$$\frac{d\overline{\text{DLR}}^e}{d\bar{R}} = \int_{\mathcal{X}^e} \frac{\theta}{\partial\sigma/\partial x}\frac{\partial\sigma}{\partial x}\left(\frac{dR^e}{d\bar{R}} - 1\right)dx \tag{3.46}$$

which is derived similarly to (3.37). Now, if $dR^e/d\bar{R} < 1$ for $x \in \mathcal{X}^e$, it follows that $d\overline{\text{DLR}}^e/d\bar{R} < 0$ by assumption 2.3. Suppose now that $dR^e/d\bar{R} > 1$ for some, but not all, $x \in \mathcal{X}^e$. Let $f[x] \equiv \theta/(\partial\sigma/\partial x)$ and $-g[x] \equiv (\partial\sigma/\partial x)(dR^e/d\bar{R} - 1)$. We shall argue that lemma 3.2 applies. Assumptions 2.1 and 2.3 imply $f[x] > 0$. If, in addition, (3.38) holds, (i) of lemma 3.2 is satisfied. After differentiation of (2.8) first with respect to $\bar{R}$ and then with respect to x, we obtain

$$\frac{d}{d\bar{R}}\left(\frac{dR^e}{dx}\right) = \frac{t}{H^{e2}}\left(\frac{\partial h^e}{\partial \bar{U}}\frac{d\bar{U}^e}{d\bar{R}} + \frac{\partial h^e}{\partial R}\frac{dR^e}{d\bar{R}}\right) < 0 \tag{3.47}$$

for $x \in \mathcal{X}^e$ since, by premise, $dR^e/d\bar{R} > 0$ everywhere and using the property of compensated demand functions $\partial h^e/\partial R < 0$. Thus if $dR^e/d\bar{R} > 1$ for some x then $dR^e\langle 0\rangle/d\bar{R} > 1$. It follows from (3.44) that (ii) of lemma 3.2 is satisfied. Finally, if $dR^e/d\bar{R} > 1$ for some, but not all, $x \in \mathcal{X}^e$, (3.47) implies that there exists a unique $x^1 \in (0, \bar{x}^e)$, such that $dR^e/d\bar{R} \geq (\leq)\, 1$ as $x \leq (\geq)\, x^1$. With (iii) of lemma 3.2 also satisfied, we know that the RHS of (3.46) is negative, that is, differential land rent decreases in any case—provided that the additional conditions imposed are true. Consequently, income also decreases for $\bar{k} > 0$ according to (2.12).

In order to determine the effect on aggregate distance travelled and aggregate transportation cost, we must first establish that population agglomerates in the sense that more individuals live inside a particular radius over the urban area after the agricultural rent has increased:

Lemma 3.3 (Pines and Sadka (1986)): $\dfrac{dN^e}{d\bar{R}} > 0$ for $x \in (0, \bar{x}^e]$.

PROOF: The proof is similar to that of lemma 3.1.

When the land rent increases everywhere $h\left[R^e, \bar{U}^e, \bar{\gamma}\right]$ decreases everywhere and, therefore, n^e increases everywhere because $\bar{Y}^e$ does not increase anywhere. Thus, in this case, our claim holds.

Suppose now there exists $x^* \in (0, \bar{x}^e)$, such that $dR^e/d\bar{R} \geq (\leq)\, 0$ as $x \leq (\geq)\, x^*$. Then $dN^e/d\bar{R} > 0$ for $x^* \in (0, \bar{x}^e)$. Also suppose there exists $x^1 \in (x^*, \bar{x}^e]$ such that $dN^e[x = x^1]/d\bar{R} \leq 0$. Using (2.1), we arrive at

$$\left.\left(\sigma\frac{dR^e}{d\bar{R}}\right)\right|_{x=x^1} + \frac{dN^e\langle x^1\rangle}{d\bar{R}} + \int_{x^1}^{\bar{x}^e}\frac{\partial\sigma}{\partial x}\frac{dR^e}{d\bar{R}}dx = \sigma\langle\bar{x}^e\rangle > 0 \tag{3.48}$$

following the same calculations as those of lemma 3.1. Since $x^1 > x^*$ by construction, $dR^e/d\bar{R} < 0$ for $x \in [x^1, \bar{x}^e]$. It follows that the first term on the LHS of (3.48) is negative; the second cannot be positive by premise; and the third is also nonpositive because of assumption 2.3. Therefore the LHS of (3.48) must be negative—a contradiction. □

Lemma 3.3, in conjunction with expressions analogous to (3.17) and (3.18), implies that both the aggregate distance travelled and the aggregate transportation cost decrease as the agricultural rent increases. In summary

Result 3.15 (Wheaton (1974), Pines and Sadka (1986)): For $\bar{k} \in [0,1]$

(i) $\dfrac{dD^e\langle 0\rangle}{d\bar{R}} > 0.$

(ii) Either (1) $\dfrac{dR^e}{d\bar{R}} > 0$ for $x \in \mathcal{X}^e$

or (2) there exists a unique $x^* \in (0, \bar{x}^e)$ such that $\dfrac{dR^e}{d\bar{R}} \geq (\leq) 0$ as $x \leq (\geq) x^*$.

(iii) In case (ii).1 $\dfrac{dD^e\langle \bar{x}^e\rangle}{d\bar{R}} > 0.$

(iv) If there are $x \in \mathcal{X}^e$ for which $\dfrac{dR^e}{d\bar{R}} \leq 1$ and if $\dfrac{\partial}{\partial x}\left(\dfrac{\theta}{\partial\sigma/\partial x}\right) \leq 0$ then $\dfrac{d\bar{Y}^e}{d\bar{R}} < 0$ for $\bar{k} > 0$; $\dfrac{d\overline{\text{DLR}}^e}{d\bar{R}} < 0.$

(v) $\dfrac{d\overline{\text{ATC}}^e}{d\bar{R}} < 0;$ $\dfrac{d\overline{\text{ADT}}^e}{d\bar{R}} < 0$

3.4.5 Share in Profits from Land Development

When the share in profits from converting the land to urban use increases, real income increases, hence it is as if the initial endowment has increased. The similarity between these two effects can be verified by observing that if $\bar{\Omega}$ is replaced by $\bar{k}$ in (3.15)–(3.18) of section 3.2.6, as well as in the entire section 3.4.2, the new set of equations is exactly the one which would have been derived through the corresponding direct argument on $\bar{k}$. Therefore, using result 3.14, we obtain

Result 3.16 (Pines and Sadka (1986)): For $\bar{k} \in [0,1)$

(i) $\dfrac{dD^e\langle 0\rangle}{d\bar{k}} < 0;$ $\dfrac{d\bar{Y}^e}{d\bar{k}} > 0;$ $\dfrac{d\bar{x}^e}{d\bar{k}} > 0.$

(ii) There exists a unique $x^* \in (0, \bar{x}^e)$ such that $\dfrac{dR^e}{d\bar{k}} \leq (\geq) 0$ as $x \leq (\geq) x^*$.

(iii) If $\dfrac{\partial}{\partial x}\left(\dfrac{\theta}{\partial\sigma/\partial x}\right) \leq 0$ then $\dfrac{d\overline{\text{DLR}}^e}{d\bar{\Omega}} > 0.$

(iv) $\dfrac{d\overline{\text{ATC}}^e}{d\bar{k}} > 0;$ $\dfrac{d\overline{\text{ADT}}^e}{d\bar{k}} > 0.$

3.4.6 Transportation Technology

When only $\bar{\alpha}$ varies (3.27) can be written as

$$\left(\sigma\frac{dR^e}{d\bar{\alpha}}\right)\bigg|_{x=0} + \int_{\mathcal{X}^e} \frac{\partial\sigma}{\partial x}\frac{dR^e}{d\bar{\alpha}}dx = \int_{\mathcal{X}^e} \frac{\partial\sigma}{\partial\bar{\alpha}}\frac{dR^e}{dx}dx < 0 \tag{3.49}$$

because $\partial t/\partial\bar{\alpha} < 0$ by assumption 2.2, hence $\partial\sigma/\partial\bar{\alpha} > 0$. It follows that $dR^e/d\bar{\alpha}$ cannot be positive everywhere. Furthermore, it cannot be negative everywhere: since utility increases by result 3.13, $h[R^e, \bar{U}^e, \bar{\gamma}]$ must increase everywhere, hence n^e must decrease everywhere and this violates (2.10). In consequence (3.6) still holds and there is a unique $x^* \in (0, \bar{x}^e)$ such that $dR^e/d\bar{\alpha} \leq (\geq) 0$ as $x \leq (\geq) x^*$. Therefore, as in section 3.2.3, figure 3.6 continues to apply, the central density declines and the city area expands with an improvement in transportation. We also know that the population disperses:

Lemma 3.4 (Pines and Sadka (1986)): $\dfrac{dN^e}{d\bar{\alpha}} < 0$ for $x \in (0, \bar{x}^e]$.

PROOF: The proof is similar to that of lemma 3.1.

Since there exists $x^* \in (0, \bar{x}^e)$ such that $dR^e/d\bar{\alpha} \leq (\geq) 0$ as $x \leq (\geq) x^*$, $dN^e/d\bar{\alpha} < 0$ for $x^* \in (0, \bar{x}^e)$. Suppose now there exists $x^1 \in (x^*, \bar{x}^e]$, such that $dN^e\langle x^1\rangle/d\bar{\alpha} \geq 0$. Using (3.49) and the calculations in the proof of lemma 3.1, we get

$$\left(\sigma\frac{dR^e}{d\bar{\alpha}} + \frac{dN^e}{d\bar{\alpha}}\right)\bigg|_{x=x^1} + \int_{x^1}^{\bar{x}^e} \frac{\partial\sigma}{\partial x}\frac{dR^e}{d\bar{\alpha}}dx = \int_{\mathcal{X}^e} \frac{\partial\sigma}{\partial\bar{\alpha}}\frac{dR^e}{dx}dx < 0. \tag{3.50}$$

Since $x^1 > x^*$ by construction, $dR^e/d\bar{\alpha} > 0$ for $x \in [x^1, \bar{x}^e]$. It follows that the LHS of (3.9) cannot be negative because $dN^e\langle x^1\rangle/dx$ is nonnegative, which is a contradiction. □

Thus the remaining conclusions in section 3.2.3 are also valid and result 3.10 applies to the general case $\bar{k} \in [0, 1]$.

3.4.7 Land Reclamation

When only $\bar{\beta}$ varies (3.27) becomes

$$\left(\sigma\frac{dR^e}{d\bar{\beta}}\right)\bigg|_{x=0} + \int_{\mathcal{X}^e} \frac{\partial\sigma}{\partial x}\frac{dR^e}{d\bar{\beta}}dx = \int_{\mathcal{X}^e} \frac{\partial\sigma}{\partial\bar{\beta}}\frac{dR^e}{dx}dx \leq 0 \tag{3.51}$$

because of assumption 2.1 and result 2.3. This implies that either dR^e/dx is negative somewhere, or that it is zero everywhere. If R^e does not decline everywhere, there exists $x^* \in (0, \bar{x}^e)$, with $dR^e\langle x^*\rangle/d\bar{\beta} = 0$. Differentiating result 2.2 with respect to $\bar{\beta}$ at x^*, we obtain

$$\frac{d}{d\bar{\beta}}\left(\frac{dR^e\langle x^*\rangle}{dx}\right) = \left(\frac{t}{H^{e2}}\frac{\partial h^e}{\partial\bar{U}}\frac{d\bar{U}^e}{d\bar{\beta}}\right)\bigg|_{x=x^*} > 0 \tag{3.52}$$

because of assumptions 2.2 and 2.4, and result 3.13. Thus when the rent does not decrease everywhere, there exists a unique $x^* \in (0, \bar{x}^e)$, such that $dR^e/d\bar{\beta} \leq (\geq) 0$ as $x \leq (\geq) x^*$. The two possibilities appear in figure 3.14.[11] Clearly the city expands or retracts according to whether the equilibrium rent increases or decreases at the border.

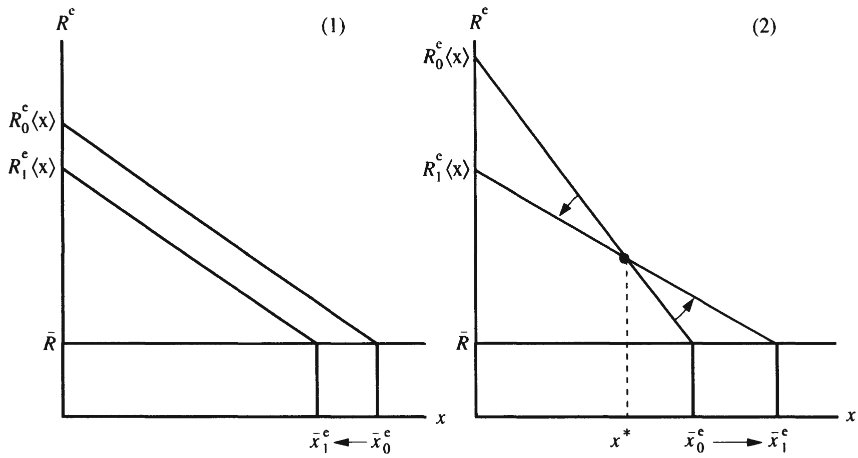

FIGURE 3.14. Effects of Land Reclamation on a Closed City.

We cannot determine the effect of land reclamation on aggregate transportation cost and aggregate distance travelled. Moreover, we cannot determine the effect of land reclamation on differential land rent, hence on income, when the rent decreases everywhere.[12] Nevertheless, when the rent increases toward the periphery, income must increase since, at x^*, only this can explain the increasing utility. Consequently the differential land rent also increases.

In summary,

---

[11] To see that both possibilities can in fact occur, consider the special case where $\theta = g[x]\bar{\beta}$ with $g' > 0$. In this case the definition of a competitive equilibrium reduces to

$$E^e = \bar{Y}^e - T$$

$$\int_{\mathcal{X}^e} \frac{g}{H^e} dx = \frac{\bar{N}}{\bar{\beta}}$$

$$R^e \geq \bar{R} \quad \text{and} \quad R^e \langle \bar{x}^e \rangle = \bar{R}$$

$$Y^e = \bar{\Omega} + \frac{\bar{\beta}\bar{k}}{\bar{N}} \int_{\mathcal{X}^e} g(R^e - \bar{R}) dx.$$

This specification implies that an increase in $\bar{\beta}$ by $\lambda\bar{\beta}$ is equivalent to a decrease in $\bar{N}$ by $\lambda\bar{N}/(1+\lambda)$. As it has been argued in section 3.4.3, an increase in $\bar{N}$ can result in a decrease of R^e close to the periphery. Therefore an increase in $\bar{\beta}$, which is equivalent to a decrease in $\bar{N}$, can result in an increase in R^e close to the periphery.

[12] The reasons are discussed in section 3.2.4.

Result 3.17: For $\bar{k} \in [0,1]$

(i) $\dfrac{dD^e\langle 0\rangle}{d\bar{\beta}} < 0.$

(ii) Either (1) $\dfrac{dR^e}{d\bar{\beta}} > 0$ for $x \in \mathcal{X}^e$

or (2) there exists a unique $x^* \in (0, \bar{x}^e)$ such that $\dfrac{dR^e}{d\bar{\beta}} \leq (\geq) 0$ as $x \leq (\geq) x^*$.

(iii) In case (ii).1 $\dfrac{d\bar{x}^e}{d\bar{\beta}} < 0.$

(iv) In case (ii).2 $\dfrac{d\bar{Y}^e}{d\bar{\beta}} > 0;\quad \dfrac{d\bar{x}^e}{d\bar{\beta}} > 0;\quad \dfrac{d\overline{\text{DLR}}^e}{d\bar{\beta}} > 0.$

3.4.8 Public Good

When only the level of the public good varies (3.27) simplifies into

$$\left(\sigma\frac{dR^e}{d\bar{\gamma}}\right)\bigg|_{x=0} + \int_{\mathcal{X}^e} \frac{\partial\sigma}{\partial x}\frac{dR^e}{d\bar{\gamma}}dx = 0. \tag{3.53}$$

Therefore, as in section 3.4.2, there exists $x^* \in (0, \bar{x}^e)$, where $dR^e\langle x^*\rangle/d\bar{\gamma} = 0$. We then follow the arguments of section 3.2.5 leading us to conclude that result 3.12 holds in the general case $\bar{k} \in [0,1]$. Finally, the effect on differential land rent is given by

$$\frac{d\overline{\text{DLR}}^e}{d\bar{\gamma}} = \int_{\mathcal{X}^e} \frac{\theta}{\partial\sigma/\partial x}\frac{\partial\sigma}{\partial x}\frac{dR^e}{d\bar{\gamma}}dx \tag{3.54}$$

as with (3.37). We also know from (3.53) that

$$\int_{\mathcal{X}^e} \frac{\partial\sigma}{\partial x}\frac{dR^e}{d\bar{\gamma}}dx > 0 \tag{3.55}$$

if housing and the public good are complements. Therefore if (3.38) holds, the RHS of (3.54) satisfies the conditions of lemma 2.2 and differential land rent increases with the level of the public good. Under these conditions income also increases.

3.5 References

M. Adler, 1987, The Location of Owners and Renters in the City. *Journal of Urban Economics* 21, 347–363.

R. J. Arnott, D. Pines and E. Sadka, 1986, The Effects of Equiproportional Transport Improvement in a Closed Monocentric City. *Regional Science and Urban Economics* 16, 387–406.

O. Ashenfelter, 1976, On the Interpretation of Urban Density Functions. *Journal of Urban Economics* 3, 82–84.

B. J. L. Berry, J. W. Simmons and R. J. Tennant, 1963, Urban Population Densities: Structure and Change. *Geographical Review* 53, 389–405.

D. J. Bogue, 1953, *Population Growth in Standard Metropolitan Areas.* Housing and Home Finance Agency, Washington D.C.

E. Casetti, 1971, Equilibrium Values and Population Densities in an Urban Setting. *Economic Geography* 47, 16–20.

C. Clark, 1951, Urban Population Densities. *Journal of the Royal Statistical Society* 114 (Series A), 490–496.

C. Clark, 1967, *Population Growth and Land Use.* St Martin's Press, New York.

K. Davis, 1965, The Urbanization of Human Populations. *Scientific American* 213, 41–53.

A. K. Dixit, 1973, The Optimum Factory Town. *Bell Journal of Economics and Management Science* 4, 637–651.

B. Edmonston, 1975, *Population Distribution in American Cities.* Heath and Company, Lexington.

N. Glickman, 1979, *The Growth and Management of the Japanese Urban System.* Academic Press, New York.

M. Goldberg and J. Mercer, 1986, *The Myth of the North American City.* The University of British Columbia Press, Vancouver.

G. K. Ingram and A. Carroll, 1981, The Spatial Structure of Latin American Cities. *Journal of Urban Economics* 9, 257–273.

E. S. Mills, 1972, *Studies in the Structure of the Urban Economy.* The Johns Hopkins University Press, Baltimore.

E. S. Mills and K. Ohta, 1976, Urbanization and Urban Problems. In H. Patrick and H. Rosovsky (eds) *Asia's New Giant: How the Japanese Economy Works.* Brookings Institution, Washington D.C., 673–751.

E. S. Mills and J. P. Tan, 1980, A Comparison of Urban Population Density Functions in Developed and Developing Countries. *Urban Studies* 17, 313–321.

R. F. Muth, 1969, *Cities and Housing.* The University of Chicago Press, Chicago.

Y. Oron, D. Pines and E. Sheshinski, 1973, Optimum Versus Equilibrium Land Use Pattern and Congestion Toll. *Bell Journal of Economics and Management Science* 4, 619–636.

Y. Y. Papageorgiou, 1971, The Population Density and Rent Distribution Models Within a Multicentre Framework. *Environment and Planning* 3, 267–282.

Y. Y. Papageorgiou and D. Pines, 1989, The Exponential Density Function: First Principles, Comparative Statics and Empirical Evidence. *Journal of Urban Economics* 26, 264–268.

D. Pines, 1972, The Equilibrium Utility Level and City Size: A Comment. *Economic Geography* 48, 439–443.

D. Pines and E. Sadka, 1986, A Comparative Statics Analysis of a Fully Closed City. *Journal of Urban Economics* 20, 1–20.

R. M. Solow, 1972, Congestion, Density, and the Use of Land in Transporta tion. *The Swedish Journal of Economics* 74, 161–173.

R. M. Solow, 1973, Congestion Cost and the Use of Land for Streets. *Bell Journal of Economics and Management Science* 4, 602–618.

W. C. Wheaton, 1974, A Comparative Statics Analysis of Urban Spatial Structure. *Journal of Economic Theory* 9, 223–237.

W. C. Wheaton, 1979, Monocentric Models of Urban Land Use: Contributions and Criticisms. In P. Mieszkowski and M. Straszheim (eds) *Current Issues in Urban Economics*. Johns Hopkins University Press, Baltimore, 107–129.

L. J. White, 1977, A Further Note on Density Functions and Average Distance from the Center. *Journal of Urban Economics* 4, 218–219.

4

Optimality of the Equilibrium Allocation

In this chapter we introduce normative considerations using the framework developed in chapter two. Since no agent in the city has market power and since there are no external economies or diseconomies, one should expect that the first welfare theorem applies to the urban spatial context as well, that is, the equilibrium allocation of chapter two is Pareto efficient. As we show in the appendix this, indeed, is the case. Thus the spatial aspect of the model does not matter in that sense. But when we adopt the more restrictive concept of optimality using, say, a social welfare function which is concave and symmetrical in the commodities consumed by the individuals, the spatial aspect of the model does matter: whereas in a spaceless context the equilibrium allocation is optimal, in a spatial context it is not—excepting a special case. More specifically, whereas the equilibrium allocation of chapter two yields the same utility level for identical individuals everywhere, the optimal allocation of this chapter yields different utility levels for identical individuals at different locations unless the criterion for optimality is to maximise minimum welfare. In other words, optimality requires unequal treatment of equals.

4.1 Allocations that Maximise Social Welfare

4.1.1 Social Welfare Function

We begin by introducing some fundamental concepts for optimality in a spaceless framework closely analogous to the spatial framework of chapter two. We shall adopt the criterion of maximising a *social welfare function* (SWF) which specifies how individual utilities are aggregated to produce a measure of social welfare.[1] Formally, a SWF is a mapping W from a set of individual utilities $\{U_j;\ j=1,...,\bar{N}\}$ to a real number $\bar{W}$ called the *aggregate level of social welfare* that represents a measure of social satisfaction: $\bar{W} = W\left[U_j;\ j=1,...,\bar{N}\right]$. The SWF obeys the Pareto criterion, that is, $W[\cdot]$ increases if the utility of an individual increases while no other individual utility level has decreased. If we restrict ourselves to identical individuals consuming a private good, housing and a public good as in chapter two, we obtain the SWF in the form $W\left[u\left[Z_j, H_j, \bar{\gamma}\right];\ j=1,...,\bar{N}\right]$ which depends indirectly on the consumption bundles of the individuals. We shall further restrict the class of SWFs by admitting only those which are additive in some transformation of the individual utilities $w\left[U_j\right]$, where w represents the *weight* assigned to the utility of an individual by the SWF.[2]

> **Assumption 4.1:** The SWF $W\left[u\left[Z_j, H_j, \bar{\gamma}\right];\ j=1,...,\bar{N}\right] = \sum_j w[u[Z_j, H_j, \bar{\gamma}]]$, where $w[U]$ is differentiable, non–decreasing and concave in $u[\cdot]$, and strictly concave in Z and H.[3]

In a spaceless context, a competitive equilibrium allocation maximises any SWF satisfying assumption 4.1. Given that the equilibrium allocation is Pareto efficient by the first welfare theorem, this assertion must be true if maximising social welfare implies equal utility levels across identical individuals. To illustrate, consider a simple economy with $\bar{N}$ individuals, $\bar{Z}$ units of the numéraire commodity Z and $\bar{H}$ units of commodity H at a unit price P_H. We know that

$$\left.\begin{aligned} Z_j^e &= \bar{Z}/\bar{N} \\ H_j^e &= \bar{H}/\bar{N} \end{aligned}\right\} \text{ for } j=1,...,\bar{N} \qquad (4.1)$$
$$P_H^e = \text{MRS}\left\langle \bar{Z}/\bar{N}, \bar{H}/\bar{N}\right\rangle$$

determines the unique competitive equilibrium.[4] We have to show that the equilibrium allocation in (4.1), where everyone has an identical consumption bundle, maximises any SWF which is consistent with assumption 4.1. Consider another allocation that satisfies material balance and where two individuals, say 1 and 2, have different consumption bundles $(Z_1, H_1) \neq (Z_2, H_2)$. By the

---

[1] We confine our study to individualistic SWFs, where the arguments are the utilities as perceived by the individuals. The well–known fundamental difficulties involved in defining a SWF will not be discussed here (see, for example, Arrow (1951) and Sen (1977)).

[2] If, for example, it is socially preferable to assign the same priority for improving the condition of any individual then $w[\cdot]$ is the identity function.

[3] This does not require that $u[Z_j, H_j, \bar{\gamma}]$ itself is concave, but if it is convex, that $w[U]$ is sufficiently concave to offset the convexity of $u[Z_j, H_j, \bar{\gamma}]$.

[4] Using the framework of chapter two, if we interpret our spaceless economy as the central zone where $T=0$, impose $\bar{k}=0$, assume $\bar{\Omega} = \left(\bar{Z} + P_H^e \bar{H}\right)/\bar{N}$ and recognise that, in a spaceless context, there is no restriction on equality of rents at the border, we can see immediately that (4.1) is consistent with the equilibrium conditions (2.13) of section 2.2.6.

definition of strict concavity, we have

$$\begin{aligned} & w\left[u\left[Z_1, H_1, \bar{\gamma}\right]\right] + w\left[u\left[Z_2, H_2, \bar{\gamma}\right]\right] \\ < \quad & 2w\left[u\left[\frac{Z_1+Z_2}{2}, \frac{H_1+H_2}{2}, \bar{\gamma}\right]\right]. \end{aligned} \tag{4.2}$$

Therefore if the bundle $(Z_1 + Z_2, H_1 + H_2)$ is redistributed equally between the individuals 1 and 2, the contribution of these individuals to the SWF will increase while that of all other individuals remains the same. This observation directly implies that the SWF is maximised when the resources available are divided equally among the individuals.

The same result can be generalised to any spaceless context where no agent has market power and there are no external economies or diseconomies. Under these circumstances the competitive allocation is not only Pareto efficient but it also maximises any SWF which satisfies the strict concavity and other requirements of assumption 4.1.[5] However, as we shall establish later on, the same result is no longer valid in the case of an urban spatial economy. Namely, even though the equilibrium allocation of chapter two is Pareto efficient (see the appendix), optimality precludes equal utilities across the urban area unless the criterion for optimality is to maximise minimum welfare.

4.1.2 Conditions for Maximising Social Welfare

We are now ready to define the SWF in the spatial context of chapter two. We know that, in a spaceless context, maximizing a SWF which satisfies assumption 4.1 yields identical bundles for identical individuals. Hence, given the additive structure of the admissible SWFs to which the present discussion is restricted, individuals who live at the same distance from the centre must have the same consumption. Using this fact and (2.9), we can write the SWF as

$$\bar{W} \equiv \int_{\mathcal{X}} \frac{\theta}{H} w[u[Z, H, \bar{\gamma}]] dx. \tag{4.3}$$

Since the welfare of absentee land–owners is not included in (4.3), the optimum pertains to a closed owner city where $\bar{k} = 1$. Thus, for $x \in \mathcal{X}$, $(Z^\circ, H^\circ, \bar{x}^\circ)$ is an *optimum* if and only if it maximises $\bar{W}$ in (4.3) subject to

$$\begin{aligned} \int_{\mathcal{X}} \frac{\theta}{H} dx &= \bar{N} \qquad (a) \\ \int_{\mathcal{X}} \frac{\theta}{H} (Z+T)\, dx + \Theta\langle \bar{x}\rangle \bar{R} &= \bar{N}\bar{\Omega}. \quad (b) \end{aligned} \tag{4.4}$$

[5] Assumption 4.1 can be extended to include different utilities $U_j = u^j\left[Z_j, H_j, \bar{\gamma}\right]$.

In this definition w, T, θ and Θ are known functions, and $\bar{\Omega}, \bar{N}, \bar{R}$,and $\bar{\gamma}$ are given parameters.[6]

Assumption 4.1 does not guarantee an internal solution for maximising (4.3) subject to (4.4). On the contrary, Hartwick (1982) has constructed degenerate cases where it is optimal to concentrate the entire urban population at a single distance from the centre even though $w[U]$ is strictly concave in Z and H. Internal solutions require a degree of concavity sufficiently strong to counterbalance the advantages born by the proximity to the centre.[7] We therefore impose

> **Assumption 4.2**: $w[U]$ has a degree of concavity in Z and H sufficiently strong to rule out depopulated zones at the optimum.

The first-order conditions for an optimum are

$$\begin{aligned} \bar{\lambda}^{\circ} &= \frac{dw^{\circ}}{dU}\frac{\partial u^{\circ}}{\partial Z} && (a) \\ Z^{\circ} + H^{\circ}\mathrm{MRS}\langle Z^{\circ}, H^{\circ}\rangle + T &= \frac{1}{\bar{\lambda}^{\circ}}\left(w[u[Z^{\circ}, H^{\circ}, \bar{\gamma}]] - \bar{\mu}^{\circ}\right) && (b) \\ \left(n^{\circ}\left(Z^{\circ} + T\right) + \theta\bar{R}\right)\Big|_{x=\bar{x}^{\circ}} &= \left(\frac{n^{\circ}}{\bar{\lambda}^{\circ}}\left(w[u[Z^{\circ}, H^{\circ}, \bar{\gamma}]] - \bar{\mu}^{\circ}\right)\right)\Bigg|_{x=\bar{x}^{\circ}} && (c) \end{aligned} \quad (4.5)$$

---

[6] Comparing with the definition of equilibrium (2.13), the optimum satisfies the composite good constraint that corresponds to $\bar{k} = 1$ (see footnote 6 in chapter two), the land constraint (2.9) and the population constraint (2.10). The requirement for equal rents at the border, which appears in the definition of equilibrium (2.13), does not apply as a constraint in the case of an optimum. Whether or not urban rent equals agricultural rent at the city border will be derived as a property of the optimal solution.

[7] Consider a simple model with only two zones, centre and suburb. The cost of transportation at the centre is zero. Let $F[Z_j, H_j, \bar{\gamma}] \equiv w[u[\cdot]]$, $j = 1, 2$, be homogeneous of degree $\bar{a}$. At the optimum, $\bar{\mathbf{Z}}_j$ units of the composite good are allocated to zone j to be partitioned among the individuals there. For individuals in the suburb, some of this must be used for transportation. Then, at the optimum,

$$W \equiv n_1F[Z_1, H_1] + n_2F[Z_2, H_2] = n_1F\left[\frac{\bar{\mathbf{Z}}_1}{n_1}, \frac{\theta_1}{n_1}\right] + (\bar{N} - n_1)F\left[\frac{\bar{\mathbf{Z}}_2 - (\bar{N} - n_1)T_2}{\bar{N} - n_1}, \frac{\theta_2}{\bar{N} - n_1}\right]$$
$$= n_1^{1-\bar{a}}F[\bar{\mathbf{Z}}_1, \theta_1] + (\bar{N} - n_1)^{1-\bar{a}}F[\bar{\mathbf{Z}}_2 - (\bar{N} - n_1)T_2, \theta_2].$$

The border between strict convexity and strict concavity of the weighting function corresponds to $\bar{a} = 1$. In this case, redistributing the consumption bundles of individuals within a particular zone will not affect social welfare. However, for $\bar{a} = 1$,

$$\frac{\partial W}{\partial n_1} = \frac{\partial F}{\partial Z_2}T_2 > 0.$$

Therefore social welfare increases by moving individuals from the suburb to the centre when the weighting function is linearly homogeneous. It follows that a non-degenerate optimum requires the entire population at the centre when $\bar{a} = 1$. The optimal depopulation of the suburb can be reversed only if $\bar{a}$ becomes sufficiently smaller than unity, that is, only if the concavity of the weighting function becomes sufficiently strong.

where $\bar{\mu}$ and $\bar{\lambda}$ represent Lagrangean multipliers which are strictly positive and which correspond to the population and resource constraints in (4.4) respectively.

4.1.3 Properties of the Optimal Allocation

Conditions (4.5(b)) and (4.5(c)) can be written as

$$\begin{aligned} Z^\circ + H^\circ R^\circ + T &= \frac{1}{\bar{\lambda}^\circ}\left(w[u[Z^\circ, H^\circ, \bar{\gamma}]] - \bar{\mu}^\circ\right) && (a) \\ \left(Z^\circ + \frac{\theta}{n^\circ}\bar{R} + T\right)\bigg|_{x=\bar{x}^\circ} &= \left(\frac{1}{\bar{\lambda}^\circ}\left(w[u[Z^\circ, H^\circ, \bar{\gamma}]] - \bar{\mu}^\circ\right)\right)\bigg|_{x=\bar{x}^\circ} && (b) \end{aligned} \tag{4.6}$$

respectively, where

$$R^\circ = \mathrm{MRS}\,\langle Z^\circ, H^\circ\rangle\,. \tag{4.7}$$

Using (2.9), comparison of the two expressions in (4.6) yields

Result 4.1 (Mirrlees (1972)): $R^\circ\,\langle\bar{x}^\circ\rangle = \bar{R}$.

Therefore, as in the case of equilibrium, the optimal urban rent at the border of the city equals the cost of developing a unit of land. Now differentiate (4.6(a)) with respect to distance:

$$\frac{dZ^\circ}{dx} + H^\circ\frac{dR^\circ}{dx} + R^\circ\frac{dH^\circ}{dx} + t = \frac{1}{\bar{\lambda}^\circ}\frac{dw^\circ}{dU}\frac{\partial u^\circ}{\partial Z}\left(\frac{dZ^\circ}{dx} + R^\circ\frac{dH^\circ}{dx}\right). \tag{4.8}$$

Substituting (4.7) into (4.8 implies

Result 4.2 (Mirrlees (1972)): $H^\circ\dfrac{dR^\circ}{dx} + t = 0$.

Therefore, as in the case of equilibrium, the imputed optimal rent satisfies the principle of zero marginal location costs and it decreases with distance. Moreover, applying the argument of section 2.3.5 on result 4.2, we obtain the exact analogue of the equilibrium result 2.6:

Result 4.3: $\min\limits_{x\in\mathcal{X}^\circ}\left(\dfrac{\eta_{T:x}}{\eta_{\Theta:x}}\right) \leq \dfrac{\overline{\mathrm{DLR}}^\circ}{\overline{\mathrm{ATC}}^\circ} \leq \max\limits_{x\in\mathcal{X}^\circ}\left(\dfrac{\eta_{T:x}}{\eta_{\Theta:x}}\right)$.

We now arrive at the core of this chapter. Recall that $\partial u/\partial Z = 1/\left(\partial e/\partial U\right)$.[8] Substitute this into (4.5(a)) and differentiate the resulting expression with respect to distance:

$$\frac{d^2w^\circ}{dU^2}\frac{dU^\circ}{dx} = \lambda^\circ\left(\frac{\partial^2 e^\circ}{\partial U\partial R}\frac{dR^\circ}{dx} + \frac{\partial^2 e^\circ}{\partial U^2}\frac{dU^\circ}{dx}\right). \tag{4.9}$$

[8] If ν represents the Lagrangean multiplier of an individual's expenditure minimisation problem (2.30) in appendix B, the first-order condition for Z is $\partial u/\partial Z = 1/\nu$ while $\partial e/\partial U = \nu$ follows from the envelope theorem.

After replacing the value of $\bar{\lambda}^{\circ}$ from (4.5(a)) into (4.9) and rearrangement, we obtain

$$\left(\frac{d^2w^{\circ}}{dU^2} - \left(\frac{dw^{\circ}}{dU}\frac{\partial^2 e^{\circ}}{\partial U^2}\right) / \left(\frac{\partial e^{\circ}}{\partial U}\right)\right)\frac{dU^{\circ}}{dx} = \left(\frac{dw^{\circ}}{dU}\frac{\partial^2 e^{\circ}}{\partial U \partial R}\frac{dR^{\circ}}{dx}\right) / \left(\frac{\partial e^{\circ}}{\partial U}\right) \underset{(2.15)}{=} \left(\frac{dw^{\circ}}{dU}\frac{\partial h^{\circ}}{\partial U}\frac{dR^{\circ}}{dx}\right) / \left(\frac{\partial e^{\circ}}{\partial U}\right) < 0. \tag{4.10}$$

The last inequality follows provided $dw^{\circ}/dU > 0$ for all U because land is a normal good and $dR^{\circ}/dx < 0$ by result 4.1. We also know that the coefficient of dU°/dx on the LHS of is negative.[9] Consequently,

> **Result 4.4** (Mirrlees (1972)): If $dw^{\circ}/dU > 0$ for all U then the optimal utility level increases away from the centre.

The result states that the optimal utility level in a city where individuals have identical preferences must increase with distance from the centre. It was considered in the 1970s to be one of the most intriguing results in urban economics (see, for example, Mills and MacKinnon (1973)). This characteristic of optimal urban resource allocation was first discovered by Mirrlees (1972) in the case of a Benthamite SWF. Subsequently, Dixit (1973) and Riley (1973) among others, established that the same characteristic follows from the maximisation of any SWF which is both symmetric and quasi–concave in the utilities if land is a normal good.

The intuition behind result 4.4 is straightforward (see Arnott and Riley (1977), Levhari, Oron, and Pines (1978), and Wildasin (1983)). Consider (4.5(a)). A uniform spatial distribution of utility implies that $w[U]$ is constant across locations and, therefore, so must be $\partial u^{\circ}/\partial Z$. But we know that $\partial u^{\circ}/\partial Z$, hence $(dw^{\circ}/dU)(\partial u^{\circ}/\partial Z)$, increases with distance from the centre when evaluated for

---

[9] Differentiating

$$\frac{\partial U}{\partial Z} = 1/\frac{\partial e}{\partial U}$$

with respect to Z yields

$$\text{I.}\quad \frac{\partial^2 U}{\partial Z^2} = -\frac{(\partial^2 e/\partial U^2)\,(\partial U/\partial Z)}{(\partial e/\partial U)^2}.$$

The concavity of $w[U]$ with respect to Z implies

$$\text{II.}\quad 0 > \frac{d^2w}{dZ^2} = \frac{d^2w}{dU^2}\left(\frac{\partial U}{\partial Z}\right)^2 + \frac{dw}{dU}\left(\frac{\partial^2 U}{\partial Z^2}\right).$$

If we combine I and II and use $\partial e/\partial U = 1/(\partial U/\partial Z)$, he have

$$\frac{d^2w}{dU^2} - \left(\frac{dw}{dU}\frac{\partial^2 e}{\partial U^2}\right) / \left(\frac{\partial e}{\partial U}\right) < 0.$$

a competitive equilibrium allocation.[10] We can use this fact to understand why, contrary to an equilibrium allocation, an optimal allocation that maximises a SWF satisfying assumption 4.1 does not imply a uniform utility over the city. Define any two distinct rings accommodating the same number of individuals in equilibrium, the central ring and the peripheral ring. Match one–to–one all individuals in the central ring with individuals in the peripheral ring. Transfer one unit of the composite good from each individual in the central ring to the corresponding individual in the peripheral ring. Then $w[U]$ will decline in the central ring and will increase in the peripheral ring, but the decline will be smaller in absolute terms than the corresponding increase because $(dw^e/dU)(\partial u^e/\partial Z)$ increases away from the centre. We conclude that, in general, starting with a competitive equilibrium and provided that $dw/dU > 0$ for all U, the aggregate level of social welfare $\bar{W}$ can be increased by reducing consumption in central locations and increasing consumption in peripheral locations. The only exception is when the SWF exhibits zero elasticity of substitution between the individual utilities. In that case, unequal distribution of welfare would cause a strict social loss because utility gains in the peripheral ring do not count socially by definition. To summarise, excepting the case of zero elasticity of substitution between the utilities in the SWF, an urban allocation which maximises a SWF satisfying assumption 4.1 implies unequal treatment of equals. As we explain later on, zero elasticity of substitution corresponds to the social criterion of maximising minimum welfare articulated by Rawls (1971).

Further insight into result 4.4 can be gained from figures 4.1(1) and 4.1(2). They are used to illustrate the above arguments diagrammatically, as well as for deriving additional results. Consider two locations $x_1, x_2 \in \mathcal{X}^\circ, x_1 < x_2$.[11] Let $dw/dU > 0$ for all U. At the optimum, the associated consumption bundles are (Z_1°, H_1°) and (Z_2°, H_2°). Using results 4.2 and 4.4, we know that $R_1^\circ > R_2^\circ$ and $U_1^\circ < U_2^\circ$. Since, at x_2, the price of land is lower and the utility is higher, $H_1^\circ < H_2^\circ$. Now, given H_1° and H_2°, we can define the constrained *utility possibility frontier* UPF which relates the utility of an individual at x_1 to that of an individual at x_2 when $H_i = H_i^\circ$ $(i = 1, 2)$ and $\bar{Z} = Z_1^\circ + Z_2^\circ$ is continuously redistributed. It is clear that since $H_1^\circ < H_2^\circ$, $u[Z, H_1^\circ] < u[Z, H_2^\circ]$ for all Z. Thus, if individual a lives at x_1 and individual b lives at x_2, the constrained UPF is represented by AA′ in figure 4.1(1); and BB′ in the same figure represents the UPF when the two individuals exchange locations. Now consider the point C in figure 4.1(2), where the division of $\bar{Z}$ is skewed in favor of a so that the utility

[10] Since $\partial u^e/\partial Z$ is the reciprocal of $\partial e^e/\partial U$, we have

$$\begin{aligned}\frac{d}{dx}\left(\frac{\partial u^e}{\partial Z}\right) &= \frac{d}{dx}\left(1/\frac{\partial e^e}{\partial U}\right) = -\frac{\partial^2 e^e/\partial U \partial R}{(\partial e^e/\partial U)^2}\frac{dR^e}{dx} \\ &\underset{(2.15)}{=} -\frac{\partial h^e/\partial U}{(\partial e^e/\partial U)^2}\frac{dR^e}{dx} > 0\end{aligned}$$

because housing is normal and equilibrium rent decreases away from the centre.

[11] This exposition draws from Arnott and Riley (1977), and Levhari, Oron and Pines (1978).

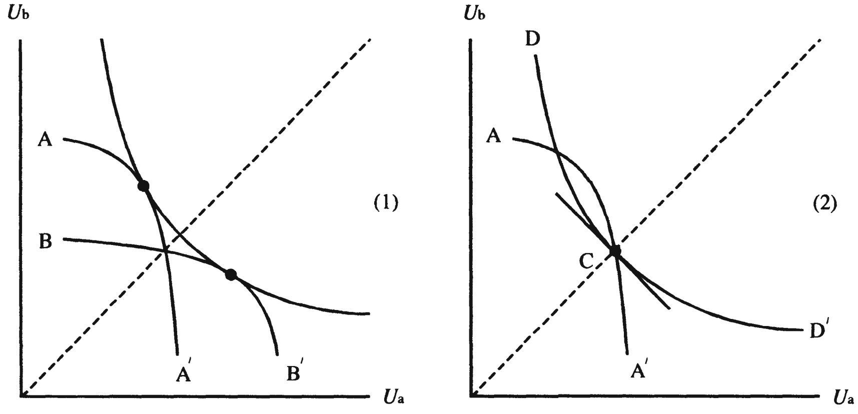

FIGURE 4.1. Unequal Treatment of Equals.

is equalised. At that point a and b are on the same indifference curve with b having more housing than a. According to the previous analysis, it must be true that the slope of the constrained UPF, $dU_b/dU_a = -(\partial u/\partial Z_b)\,/\,(\partial u/\partial Z_a)$, must be steeper than 45°. However, the slope of the social indifference curve DD passing through C must be equal to 45° because the SWF is symmetric. It follows that the slope of the constrained UPF at C is steeper than that of the social indifference curve passing through C. This, of course, indicates that social welfare can be increased by moving to the locus above the 45°, thus implying that b enjoys higher utility than a at the optimum.

4.2 Decentralisation

We can imagine that the optimal city is created by a planner who personifies social institutions. According to the definition of an optimal city in section 4.1.2 the planner, equipped with a full tool menu, determines $\mathcal{X}^{\circ}$ and, for each location, Z° and H°. Thus the planner decides in every detail the feasible urban structure that best fits the optimality criterion embodied in the SWF. Is it though necessary to use such 'brute force' policy, or could perhaps the same outcome be achieved with a smaller set of instruments? The question here is about the nature of the minimal intervention required to maximise the SWF while letting the market to perform the rest. It turns out that this can be achieved by just redistributing income in a specific manner. In other words there exists a price system which can support the SWF–maximising allocation given some feasible income distribution. Under these circumstances,

1. the income redistribution is feasible in the sense that the net balance of taxes and subsidies at the optimum is zero, and

2. individuals can do no better with their disposable income by changing consumption bundle and/or location as determined by the planner.

To show that these two conditions can be achieved, consider the following cumulative income redistribution: For all $x \in \mathcal{X}^{\circ}$,

$$\bar{N} - N = \int_{x}^{\overline{x}^{\circ}} \frac{\theta}{H\langle x'\rangle^{\circ}} dx' \tag{4.11}$$

individuals receive a subsidy of more than

$$\Phi^{\circ}\langle x\rangle = \frac{1}{\bar{\lambda}^{\circ}}\left(w[U^{\circ}\langle x\rangle] - \bar{\mu}^{\circ}\right) - \left(\overline{\Omega} + \frac{\overline{\mathrm{DLR}^{\circ}}}{\bar{N}}\right). \tag{4.12}$$

Condition 1 follows from (4.4) and (4.5) which imply that

$$\int_{0}^{\overline{x}^{\circ}} \frac{\theta}{H^{\circ}} \Phi^{\circ} dx = 0. \tag{4.13}$$

We now turn to the second condition. We have to show that, under the above income redistribution, there exists a price function P such that an individual at distance $x \in \mathcal{X}^{\circ}$ from the centre who receives a (positive or negative) subsidy $\Phi^{\circ}\langle x\rangle$ can achieve $U^{\circ}\langle x\rangle$ and cannot achieve a higher utility elsewhere. Let $P = R^{\circ}$. Then we have

$$\begin{aligned}
Z^{\circ} &= z\left[R^{\circ}, U^{\circ}, \bar{\gamma}\right] && (a)\\
H^{\circ} &= q\left[R^{\circ}, U^{\circ}, \bar{\gamma}\right] && (b)\\
e\left[R^{\circ}, U^{\circ}, \bar{\gamma}\right] + T &\equiv && \\
Z^{\circ} + R^{\circ}H^{\circ} + T &= \bar{\Omega} + \frac{1}{\bar{N}}\left(\int_{\mathcal{X}^{\circ}} \theta R^{\circ} dx - \Theta\langle \bar{x}^{\circ}\rangle \bar{R}\right) + \Phi^{\circ} && (c)\\
R^{\circ} &\geq \bar{R} && (d)
\end{aligned} \tag{4.14}$$

where (4.14(c)) follows directly from the definition of redistributed income in a closed owner city, while (4.14(d)) follows from results 4.1 and 4.2. Equation (4.14(c)) implies that the individual at x can just afford $U^{\circ}\langle x\rangle$. Can the same individual afford higher utility elsewhere with the same income? We will show that this is impossible, in other words, that

$$e\left[R^{\circ}\langle x_1\rangle, U^{\circ}\langle x\rangle, \bar{\gamma}\right] + T\langle x_1\rangle \geq e\left[R^{\circ}\langle x\rangle, U^{\circ}\langle x\rangle, \bar{\gamma}\right] + T\langle x\rangle \tag{4.15}$$

for $x_1 \neq x$. Suppose $x_1 > x$. Then, using the normality of H, we have

$$
\begin{aligned}
& e\left[R^\circ\langle x_1\rangle, U^\circ\langle x\rangle, \bar{\gamma}\right] + T\langle x_1\rangle - e\left[R^\circ\langle x\rangle, U^\circ\langle x\rangle, \bar{\gamma}\right] + T\langle x\rangle \\
= \;& \int_x^{x_1} \left\{ \frac{\partial e\left[R^\circ[x'], U^\circ\langle x\rangle, \bar{\gamma}\right]}{\partial R} \frac{dR^\circ}{dx'} + t[x'] \right\} dx' \\
\underset{(2.15)}{=} \;& \int_x^{x_1} \left\{ H^\circ\left[R^\circ[x'], U^\circ\langle x\rangle, \bar{\gamma}\right] \frac{dR^\circ}{dx'} + t[x'] \right\} dx' \\
\underset{(\text{result } 4.4)}{\geq} \;& \int_x^{x_1} \left\{ H^\circ\left[R^\circ[x'], U^\circ[x'], \bar{\gamma}\right] \frac{dR^\circ}{dx'} + t[x'] \right\} dx' \\
\underset{(\text{result } 4.2)}{=} \;& 0.
\end{aligned}
\tag{4.16}
$$

We can use a similar proof for the case $x_1 < x$. Thus we obtain

> **Result 4.5** (Mirrlees (1972)): The optimum can be decentralised under an appropriate income redistribution.

Of course, the optimal taxes or subsidies cannot be applied after the individuals choose their location because, then, everyone would prefer to live at the boundary. Rather, the optimal tax or subsidy scheme should be determined independently of the location choice, say, by lottery as Mirrlees (1972) has suggested. Then competition among individuals will establish the optimal rent schedule and will ensure that everybody chooses the optimal consumption bundle and location as intended by the planner.

4.3 How to Treat Equals?

Since the SWF is symmetric and quasi–concave in the utilities, the shape of social indifference curves can range from line BB′ to CC′ to a limit distribution approaching DD′ in figure 4.2. The range of optimal distributions of utility between individuals a and b, when individual a lives at x_1 and individual b lives at x_2, corresponds to the segment EG on AA′ in that figure. When the optimality criterion is to maximise average utility, the SWF is Benthamite, the social indifference curves are represented by BB′, and optimal inequality is maximised at E. On the other extreme, when the optimality criterion is to maximise minimum utility, the SWF is Rawlsian (1971), the social indifference curves are represented by DD′, and the optimal solution is given by G where the two utilities are equal. This is the only case where the requirement $dw/dU > 0$ for all U fails, hence result 4.4 does not apply. In–between these two extremes, various

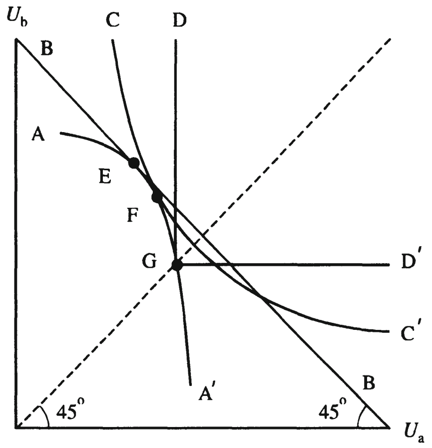

FIGURE 4.2. The Impact of Aversion to Inequality on the Optimal Allocation.

optimality criteria reflect various types of compromise between the average level of utility and the degree of inequality to be permitted.[12]

The unequal treatment of equals is a disturbing consequence of optimising resource allocation over space. However Levhari, Oron and Pines (1978) have shown that the problem is resolved when the preferences of individuals satisfy the von Neumann–Morgenstern expected utility assumptions. To see this

[12] A simple way of classifying optimality criteria, introduced by Riley (1972), is through the *degree of aversion to inequality* ζ. Under this classification, $w[u[\cdot]]$ obeys

$$-\frac{d}{dU}\left(\frac{dw}{dU}\right) / \left(\frac{1}{U}\frac{dw}{dU}\right) \equiv \eta_{w':U} = \zeta$$

where $\eta_{w':U}$ is the utility elasticity of the change in the weighted utility. Consequently, as the degree of aversion to inequality increases, weighted utility decreases relatively faster for relatively higher utility: there is an increasing bias in favour of the less advantaged. Disregarding arbitrary constants, the solution to this differential equation is given by

$$w[u[\cdot]] = \frac{1}{1-\zeta}u[\cdot]^{1-\zeta} \text{ for } \zeta \neq 1$$

$$w[u[\cdot]] = \ln u[\cdot] \text{ for } \zeta = 1.$$

The two extreme optimality criteria are given by zero aversion to inequality, which corresponds to a Benthamite SWF, and by infinite aversion to inequality, which is equivalent to the Rawlsian principle of maximising the utility of the least advantaged.

Within this framework, social indifference curves are determined by

$$dW = \frac{dw}{dU_a}dU_a + \frac{dw}{dU_b}dU_b = U_a^{-\zeta}dU_a + U_b^{-\zeta}dU_b = 0,$$

which implies

$$\frac{dU_b}{dU_a} = -\left(\frac{U_b}{U_a}\right)^{\zeta}.$$

When $\zeta = 0$, $dU_b/dU_a = -1$. When $\zeta \to \infty$, $dU_b/dU_a \to -\infty$ for $U_a < U_b$ and $dU_b/dU_a \to 0$ for $U_a > U_b$.

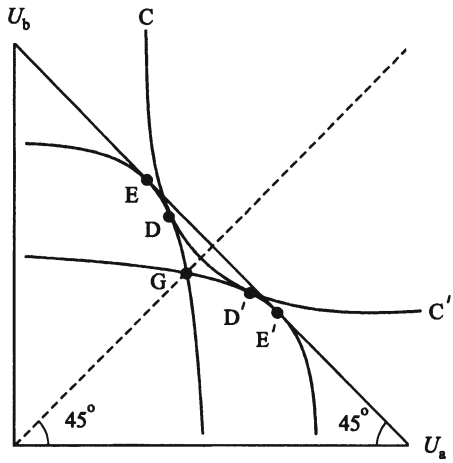

FIGURE 4.3. The Case of von Neumann-Morgenstern Utilities.

consider a strictly quasi–concave SWF defined on certain prospects and the resulting optima.[13] As before, choose any two locations at different distances from the centre and two individuals. By exchanging individuals between locations we obtain the constrained UPF BGB′ in figure 4.3. If CC′ is an indifference curve of the SWF, the two optimal allocations are represented by D and D′.

Now introduce lotteries on uncertain prospects with varying probabilities on the outcomes D and D′. Under the von Neumann–Morgenstern expected utility assumptions, the constrained UPF extends outwards by including a set of uncertain prospects to become (D, D′; $\kappa \in [0,1]$). This set is represented by the straight line connecting D and D′ in figure 4.3. The SWF, now defined on uncertain prospects, is maximised at the midpoint between D and D′ implying that, *ex ante*, equals are treated equally by receiving a prospect (D, D′; .5). Of course, the prospect (D, D′; .5) itself is Pareto inferior because, with housing supply fixed (which keeps the UPF unchanged), we can imagine a SWF exhibiting less aversion to inequality that would shift the line segment DD′ outwards along the 45° line. For such constrained UPF, the limit is reached when a Benthamite SWF has been applied to determine the optimal outcomes E and E′ which yield the set of uncertain prospects (E, E′; $\kappa \in [0,1]$). As before, the SWF is maximised at the midpoint between E and E′ where each individual receives an equal prospect (E, E′; .5).[14] We conclude that under the von Neumann–Morgenstern expected utility assumptions (1) *ex ante*, individuals are treated

[13] We define prospect as a set of outcomes and associated probabilities. A certain prospect is a degenerate prospect with one outcome and unit probability.

[14] Notice that even (E, E′; .5) need not maximise the SWF over uncertain prospects: the UPF is constrained since it depends on the set (H_1°, H_2°) which is a result of maximising the original SWF over certain prospects only. There may exist another set of feasible housing allocations generating a UPF which dominates (H_1°, H_2°). Such housing allocation is obtained when we maximise a Benthamite SWF directly.

equally in the sense that everyone has an identical prospect with equal chance for the good and the bad bundles, and (2) the good and the bad bundles, which are the two uncertain outcomes for the individual, are determined by maximising the Benthamite SWF. In essence, optimal inequality happens because the UPF is non–concave when only certain prospects are admissible. With lotteries on prospects, the frontier becomes concave thus guaranteeing equal *ex ante* utilities.

It seems that the equal *ex ante* treatment result just explained has little practical consequence. If, indeed, the expected utility assumptions were valid for a city, we should expect developers to apply lotteries in marketing homes. One may thus wonder why contractors do not charge the same price for all the apartments in a building and allocate them to the purchasers by lottery. Lottery is not uncommon in public housing but rarely used, if at all, in private projects.

If we abandon the expected utility approach, we are left with an infinite array of choices—each one representing a particular trade–off between the aggregate level of social welfare and the degree of social inequality. In this context, deciding on a particular trade–off must be based on a belief about what constitutes a good criterion for social justice. One well–known philosophical theory concerning the choice of an optimality criterion has been proposed by Rawls (1971). According to this theory, participation in society is akin to participation in a game of chance with respect to assets that largely determine a way of life. Since the existing distribution of assets may well be unfair, enquiries on the nature of just institutions cannot depend on any particular outcome of this game. Therefore, individuals must step behind a 'veil of ignorance', and decide on the rules of the game as if they were uncertain about their actual positions in society. When this is done, two principles emerge. The first calls for maximum personal freedom compatible with the freedom of others. The second calls for inequalities to be arranged so that they are to the greatest benefit of the least advantaged. This last is equivalent to maximising minimum utility which, here, calls for equal utility levels at the optimum.

Although equals are treated equally *ex ante* in the case of the von Neumann–Morgenstern expected utility assumptions, their *ex post* treatment is unequal. Treating identical individuals equally *ex post* seems to be a natural principle of distributive justice, especially in the presence of doubts about the validity of the von Neumann–Morgenstern expected utility assumptions as they apply to the choice of one's relative position in the society. For these reasons, from now on, we choose to apply the Rawlsian criterion of equal utilities at the optimum.

4.4 Appendix: Equilibrium Allocation and Pareto Efficiency

In this appendix we show that the equilibrium allocation of chapter two is Pareto efficient. The total resources measured in units of the composite good which are used in equilibrium by the city's residents are

$$\bar{\Omega}\bar{N} - \left(1-\bar{k}\right)\overline{\mathrm{DLR}}^{e} = \int_{\mathcal{X}^e} \theta\left(\frac{z\left[R^e,\bar{U}^e,\bar{\gamma}\right]+T}{h\left[R^e,\bar{U}^e,\bar{\gamma}\right]}+\bar{R}\right)dx, \tag{4.17}$$

while $\left(1-\bar{k}\right)\overline{\mathrm{DLR}}^{e}$ units are transferred to the absentee landowners.[15] The equilibrium allocation is efficient if and only if, given the equilibrium utility provided to the city population $\bar{U}^e$, $\left(1-\bar{k}\right)\overline{\mathrm{DLR}}^{e}$ is the maximum quantity of resources which can be transferred to the absentee landowners. This is equivalent to requiring that the minimum quantity of resources necessary to provide the $\bar{N}$ city residents with a utility level $\bar{U}^e$ is $\bar{\Omega}\bar{N} - \left(1-\bar{k}\right)\overline{\mathrm{DLR}}^{e}$. Formally, we have to show that $(R^e, \mathcal{X}^e)$ minimises

$$\int_{\mathcal{X}} \theta\left(\frac{z\left[R,\bar{U}^e,\bar{\gamma}\right]+T}{h\left[R,\bar{U}^e,\bar{\gamma}\right]}+\bar{R}\right)dx \tag{4.18}$$

subject to

$$\int_{\mathcal{X}} \frac{\theta}{h\left[R,\bar{U}^e,\bar{\gamma}\right]}dx = \bar{N} \tag{4.19}$$

where z, h, T and θ are known functions, and $\bar{U}^e, \bar{N}$ and $\bar{\gamma}$ are known parameters.

Suppose that $\left(R^{\mathrm{P}}, \mathcal{X}^{\mathrm{P}}\right)$ solve this constrained minimisation problem for all $x \in \mathcal{X}^{\mathrm{P}}$. The associated necessary conditions can be written as

$$\begin{aligned} e\left[R^{\mathrm{P}},\bar{U}^e,\bar{\gamma}\right] &= \bar{\mu}^{\mathrm{P}} - T \quad (a) \\ R^{\mathrm{P}}\left\langle \bar{x}^{\mathrm{P}}\right\rangle &= \bar{R}. \quad (b) \end{aligned} \tag{4.20}$$

where $\bar{\mu}$ is the Lagrangean multiplier that corresponds to the population constraint.[16]

---

[15] See footnote 6 of chapter two.

[16] The Lagrangean function is given by

$$\int_{\mathcal{X}} \theta\left(\frac{Z+T}{H}+\bar{R}\right)dx - \bar{\mu}\left(\int_{\mathcal{X}}\frac{\theta}{H}dx - \bar{N}\right).$$

The necessary conditions for rent and the city border are respectively

$$-\theta\left(\left(\frac{\partial z^e}{\partial R} - \left(Z^{\mathrm{P}}+T\right)\frac{\partial h^{\mathrm{P}}}{\partial R}\right) + \bar{\mu}^{\mathrm{P}}\frac{\partial h^{\mathrm{P}}}{\partial R} + \right)/H^{\mathrm{P}\,2} = 0$$

$$-\left.\frac{\theta}{H^{\mathrm{P}}}\left(\left(Z^{\mathrm{P}}+T\right)+\bar{R}H^{\mathrm{P}}-\bar{\mu}^{\mathrm{P}}\right)\right|_{x=\bar{x}^{\mathrm{P}}} = 0.$$

By using the derivative property of compensated demands, and observing that $\partial h^{\mathrm{P}}/\partial R$ is negative, the first equality reduces immediately to (4.20(a)). Introducing that into the second equality yields (4.20(b)).

We shall first establish that, given $\bar{U}^e$, $(R^e, \mathcal{X}^e)$ is the unique set which simultaneously satisfies (4.19) and (4.20(b)) such that $R^{\text{P}} = R^e$ for all $x \in \mathcal{X}^e$. Suppose not. Then, consider each of the following three exhaustive possibilities:

1. $R^{\text{P}} > R^e$ for all $x \in \mathcal{X}^e$;
2. $R^{\text{P}} < R^e$ for all $x \in \mathcal{X}^e$;
3. $R^{\text{P}} > R^e$ for some $x \in \mathcal{X}^e$ and $R^{\text{P}} < R^e$ for some $x \in \mathcal{X}^e$.

In the first case $h\left[R^{\text{P}}, \bar{U}^e, \bar{\gamma}\right] < h\left[R^e, \bar{U}^e, \bar{\gamma}\right]$ for all $x \in \mathcal{X}^e$ and in the second case $h\left[R^{\text{P}}, \bar{U}^e, \bar{\gamma}\right] > h\left[R^e, \bar{U}^e, \bar{\gamma}\right]$ for all $x \in \mathcal{X}^e$. Thus in both cases (4.19), which is satisfied by $(R^e, \mathcal{X}^e)$, cannot be satisfied by $\left(R^{\text{P}}, \mathcal{X}^{\text{P}}\right)$. To eliminate the third case as well, notice that (4.20(a)) is equivalent to (2.13(a)) with $\bar{\mu}^{\text{P}}$ replacing income. We can therefore apply the argument of section 2.3.2 and establish that result 2.2 also applies in the case of R^{P}. Furthermore notice that, in the third case, the two rent schedules must intersect at some location. Denote any such location by x_1, so that $R^{\text{P}}\langle x_1\rangle = R^e\langle x_1\rangle$. Then $h\left[R^{\text{P}}\langle x_1\rangle, \bar{U}^e, \bar{\gamma}\right] = h\left[R^e\langle x_1\rangle, \bar{U}^e, \bar{\gamma}\right]$. This however implies that the slopes of the two rent schedules must be the same at x_1:

$$\frac{dR^{\text{P}}\langle x_1\rangle}{dx} = -\left(\frac{1}{H^{\text{P}}}\frac{\partial T}{\partial x}\right)\Bigg|_{x=x_1} = -\left(\frac{1}{H^e}\frac{\partial T}{\partial x}\right)\Bigg|_{x=x_1} = \frac{dR^e\langle x_1\rangle}{dx}, \tag{4.21}$$

which is a contradiction. We conclude that the two rent schedules coincide, $R^{\text{P}} = R^e$ for all $x \in \mathcal{X}^e$, as claimed. Consequently $z\left[R^{\text{P}}, \bar{U}^e, \bar{\gamma}\right] = z\left[R^e, \bar{U}^e, \bar{\gamma}\right]$ and $h\left[R^{\text{P}}, \bar{U}^e, \bar{\gamma}\right] = h\left[R^e, \bar{U}^e, \bar{\gamma}\right]$ for all $x \in \mathcal{X}^e$, which implies

$$\bar{\Omega}\bar{N} - \left(1-\bar{k}\right)\overline{\text{DLR}}^e = \int_{\mathcal{X}^e} \theta\left(\frac{z\left[R^{\text{P}}, \bar{U}^e, \bar{\gamma}\right] + T}{h\left[R^{\text{P}}, \bar{U}^e, \bar{\gamma}\right]} + \bar{R}\right) dx \tag{4.22}$$

by (4.17).

We have demonstrated that the minimum quantity of resources required for achieving utility $\bar{U}^e$ to $\bar{N}$ city residents is $\bar{\Omega}\bar{N} - \left(1-\bar{k}\right)\overline{\text{DLR}}^e$ which leaves no more than $\left(1-\bar{k}\right)\overline{\text{DLR}}^e$ to the absentee landowners as in the equilibrium allocation.

4.5 References

R. J. Arnott and J. G. Riley, 1977, Asymmetrical Production Possibilities, the Social Gains from Inequality and the Optimum Town. *Scandinavian Journal of Economics*, 301–311.

K. J. Arrow, 1951, *Social Choice and Individual Values.* Wiley, New York.

A. K. Dixit, 1973, The Optimum Factory Town. *Bell Journal of Economics and Management Science* 4, 637–651.

J. Hartwick, 1982, Optimal and Monopoly Towns: Degenerate Cases. In J. V. Henderson (ed.), *Research in Urban Economics*, Volume 2. JAI Press, Greenwich Conn., 329–338.

D. Levhari, Y. Oron and D. Pines, 1978, A Note on Unequal Treatment of Equals in an Urban Setting. *Journal of Urban Economics* 5, 278–284.

E. S. Mills and J. MacKinnon, 1973, Notes on the New Urban Economics. *Bell Journal of Economics and Management Science* 4, 593–601.

J. A. Mirrlees, 1972, The Optimum Town. *Swedish Journal of Economics* 74, 114–135.

J. Rawls, 1971, *A Theory of Justice.* Harvard University Press, Cambridge, Mass.

J. G. Riley, 1972, *Optimal Towns.* Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Mass.

J. G. Riley, 1973, Gammaville: An Optimal Town. *Journal of Economic Theory* 6, 471–482.

A. K. Sen, 1977, Social Choice Theory: A Re–Examination. *Econometrica* 45, 53–89.

D. E. Wildasin, 1983, The Welfare Effects of Intergovernmental Grants in an Economy with Independent Jurisdictions. *Journal of Urban Economics* 13, 147–164.

5

Extensions

In this chapter we consider some direct extensions of the basic competitive equilibrium model which was developed in chapter two. We keep our discussion strictly within the competitive framework of that chapter. Each extension further clarifies the fundamental characteristics of the intraurban resource allocation. Since treating these extensions simultaneously complicates the analysis, or even makes it intractable, we treat each extension separately. We focus on the specific implications of the extension, and we point out the results of chapter two which are not robust and the way in which they should be modified under the extension.

5.1 Utility and Location

So far, following Muth (1961), we specified the utility as a function of two commodities only: composite good and land (which represents housing) or housing. However Alonso (1964) has proposed another plausible specification in which distance from the centre appears as a third determinant of utility. In this section we provide two kinds of rationale for such specification, one as a reduced form of utility when the individual consumes *leisure* in addition to the composite good and housing, the other when utility reflects the heterogeneity of locational attributes like pollution, scenic characteristics, etc.

5.1.1 The Value of Time

In chapter two, the value of time spent on the road was not taken explicitly into account. Yet one could maintain that it is implicitly accounted for since, as in

Muth (1969), the transportation cost function incorporates both distance and time cost of transportation. Evaluating time spent in trips by the *wage rate* $\bar{\omega}$, the transportation cost function can be specified as

$$T[x,\bar{\alpha}] = T^{(m)}[x,\bar{\alpha}] + \bar{\omega}T^{(t)}[x,\bar{\alpha}], \tag{5.1}$$

where T now denotes the *total transport cost,* $T^{(m)}$ the *money cost of transportation,* and $\bar{\omega}T^{(t)}$ the *time cost of transportation.* This approach, which hinges on the strong assumption that the length of the working day is flexible, requires some modifications in the results of chapter three.[1].

If we assume that the length of the working day is institutionally fixed, it is no longer necessary to evaluate the time spent in trips by the wage rate. In fact, since the household is constrained by both money and time, the value of time is represented by the shadow price of the time constraint which need not be equal to the wage rate. Under these circumstances a reduced form of the household's choice problem can be written by substituting the time not spent at work and travel for leisure in the utility. With this reformulation, proposed by Alonso (1964), the utility becomes a function of location through its dependence on leisure—even though it is not directly affected by accessibility (through, for example, the inconvenience of commuting).

When utility depends explicitly on location much less can be said about the characteristics of the competitive equilibrium than was said in chapter two. Let us take result 2.3 as an example. Since $\partial u/\partial x < 0$, the first part of this result (about the decline of the equilibrium rent gradient over distance) can be restored. However, in order to guarantee the second part (about the convex

[1]Result 3.12(ii) is not necessarily valid any longer because we have to add a new component, $-\int(\partial\sigma/\partial\bar{\omega})(\partial R^e/\partial x)\,dx$, on the LHS of (3.28). This component is negative because $\sigma \equiv \theta/t$ and $\partial t/\partial\bar{\omega}$ is positive. Therefore, in contrast to result 3.12(ii), we cannot exclude the possibility that the equilibrium rent increases everywhere with increasing income, as represented by an increasing $\bar{\omega}$. For example, consider the case of a circular closed owner city in which, by an appropriate choice of units, $\bar{\Omega} = \bar{\omega}$ and where $T[x] = \bar{\omega}x$. (We omit parameters $\bar{\alpha}, \bar{\beta}$, and $\bar{\gamma}$.) Similarly to the example of section 2.3.5, let $u[Z,H] = \min(Z,H)$. It follows that, in equilibrium, $H^e = \pi\bar{x}^{e2}/\bar{N} = Z^e$. Since the locational costs $R^eH^e + T$ are constant over the city, we have

$$R^e - \bar{R} = \frac{T\langle\bar{x}^e\rangle - T}{H^e} = \frac{\bar{N}\bar{\omega}(\bar{x}^e - x)}{\pi\bar{x}^{e2}}$$

which, upon multiplication by $2\pi x$ and integration over the urban area, implies $\overline{\text{DLR}}^e = \bar{N}\bar{\omega}\bar{x}^e/3$. Since the city is circular and transport cost is linear, $\overline{\text{ATC}}^e = 2\overline{\text{DLR}}^e = 2\bar{N}\bar{\omega}\bar{x}^e/3$ by result 2.6. Taking also into account that total expenditure for the consumption of the composite good is given by $\bar{N}Z^e = \pi\bar{x}^{e2}$, and that $k=1$, the resource constraint in footnote 6 of chapter two becomes

$$\bar{N}\bar{\omega} = \pi\bar{x}^{e2}(1+\bar{R}) + \frac{2\bar{N}\bar{\omega}\bar{x}^e}{3}.$$

Differentiating the above expressions, we conclude that

$$\frac{\partial R^e}{\partial\bar{\omega}} = \frac{\bar{N}\left(\bar{N}(1+\bar{R}) + 2\bar{N}\bar{\omega}(\bar{x}^e - x)/(3\pi\bar{x}^{e2})\right)}{2\pi\bar{x}^e(1+\bar{R}) + 2\bar{N}\bar{\omega}/3}$$

which is positive for all locations in the city contrary to result 3.12(ii).

shape of the rent gradient), we need to restrict $\partial^2 u/\partial x^2$ or to impose some more complicated condition.

5.1.2 Environmental Quality

An alternative way in which utility can be made to depend explicitly on location is to assume that the quality of the environment varies with distance from the centre, and that individual preferences are sensitive to the quality of the environment. In the simplest case, where the quality of the environment is exogenous, we can represent it by allowing $\bar{\gamma}$ to vary with distance from the centre as $\gamma[x]$. Taking into account that the minimum expenditure function is now written as $e\left[R, \bar{U}, \gamma\right]$, result 2.2 can be generalised as

$$H^e \frac{dR^e}{dx} + t - \lambda^e \frac{\partial u}{\partial \gamma}\frac{d\gamma}{dx} = 0, \tag{5.2}$$

where λ is the Lagrangean multiplier of problem (2.1) modified to take into account the variable quality of the environment. Therefore

$$\text{sign}\ \frac{dR^e}{dx} = \text{sign}\left(\lambda^e \frac{\partial u}{\partial \gamma}\frac{d\gamma}{dx} - t\right), \tag{5.3}$$

which implies that equilibrium rent can increase away from the centre—provided that the marginal benefit of a better environment further away dominates over the corresponding marginal cost of transportation.[2]

Newling (1966) and Latham and Yeates (1970) , among others, have provided evidence that residential population density may form a 'crater' around the centre of some large cities. Latham and Yeates argued that the density crater forms because, around the centre, homes are replaced by office buildings. An alternative explanation can be obtained within our framework by taking into account the systematic environmental quality variations observed around such centres. For example, the central area of a large city might suffer from increased levels of pollution, congestion and crime, which can be represented in first approximation by a quality of the environment increasing away from the centre as in figure 5.1(1). Using (5.2), we can graph the components of the equilibrium rent gradient as in figure 5.1(2). Since the marginal benefit of a better environment further away dominates over the corresponding marginal cost of transportation between the centre and A in figure 5.1(2), urban land values form a corresponding 'crater' as in figure 5.1(3). Finally, by (2.21), we know that a rent 'crater' must imply a population density crater if the price elasticity of the compensated demand for housing is constant.

---

[2] If environmental quality variations are strong enough over the city, there maybe an area in which environmental quality is sufficiently low to force the equilibrium rent below the opportunity cost of land. In this case we obtain a spatially discontinuous urban area.

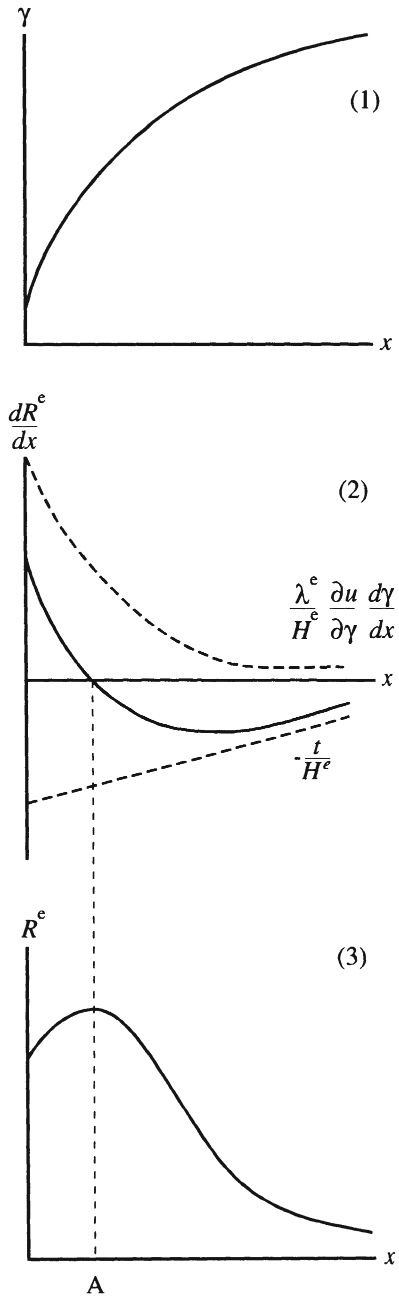

FIGURE 5.1. Formation of an Equilibrium Rent 'Crater'.

5.2 Production

Urban production in chapter two was exogenous, and the initial endowment was defined directly in terms of the composite good. Here urban production is endogenous and the initial endowment is defined in terms of labour. We consider a city which is partitioned into a *central business district* $\mathcal{X}_o$ [CBD] and a surrounding *residential area* $\mathcal{X}_1$. The composite good is produced using labour and land in the CBD, which extends between the centre (zero) and the *inner border* $\bar{x}_o$. All available land within the CBD is used for the production of the composite good. The output of production is concentrated at the centre where it is sold.[3] Workers live in the residential area which surrounds the CBD, and which extends between the inner border and the *city border* $\bar{x}_1$, $0 < \bar{x}_o < \bar{x}_1$. Everyone in the city works to produce the composite good, and no outside labour is used for production.

Suppose that each ring in the CBD is occupied by a single firm which produces the composite good under constant returns to scale. Following the specification proposed by Henderson (1986), the *quantity produced* by the firm at x, $Z_o[x]$, is given by

$$Z_o = S\left[\bar{N}\right] f_Z\left[n_o, \theta\right](1 - T_o) \tag{5.4}$$

where $\bar{S} \equiv S\left[\bar{N}\right]$ is a Hicks neutral shift factor, $F_Z\left[x, \bar{\beta}\right] \equiv f_Z\left[n_o[x], \theta\left[x, \bar{\beta}\right]\right]$ is a linear homogeneous production function, $n_o[x]$ is the *number of workers employed* by the firm at x, and $T_o[x, \bar{\alpha}]$ is the cost of transporting one unit of the composite good from x to the centre where it is sold.[4] Since workers are identical, they must earn the same wage rate irrespectively of who employs them in the CBD. If we now let $\bar{\Omega}$ represent the *wage* (which replaces the initial endowment of chapter two), and $R_o[x]$ the *rent per unit of industrial land* at x, the problem of a firm is to choose the number of workers and the location (hence θ) which maximises profit:

$$\max_{n_o, x}\left(Z_o - \left(\bar{\Omega} n_o + R_o\theta\right)\right). \tag{5.5}$$

Maximising profit implies

$$\begin{aligned} \bar{S}\frac{\partial f_Z^s}{\partial n_o}(1 - T_o) &= \bar{\Omega} \\ \bar{S}\frac{\partial f_Z^s}{\partial \theta}(1 - T_o) &= R_o, \end{aligned} \tag{5.6}$$

where superscript s denotes a solution to problem (5.5).

---

[3] The centre serves as the only marketplace for the composite commodity. This can be justified by imagining that cities specialise in city-specific products, and that the output of a city is exported from the centre for another composite good which arrives at the centre to be sold at a fixed price which, without loss of generality, is assumed to be the unit price.

[4] The same type of extension was also proposed by Mills (1967) and Dixit (1973). However, only in Henderson (1986) can the equilibrium allocation be fully supported by a competitive price system.

Every individual who lives in the residential area and works in the CBD travels to work along the radius passing through his or her residence. The individual's specific place of work is located on that radius. After work, the individual continues travelling toward the centre along the same radius, purchases the composite good at the centre and returns to his or her residence. This implies that the cost of transportation for an individual at x, $\bar{x}_o < x \le \bar{x}_1$, is $T_1[x, \bar{\alpha}] = T[x, \bar{\alpha}]$ as in chapter two, hence that the choice problem of an individual here remains exactly as in that chapter. It also follows that the conditions which define a competitive equilibrium in section 2.2.6 still hold—provided that they are adjusted to apply within the residential area (excepting differential land rent which is defined over the entire urban area) and provided that the initial endowment is replaced by (5.6(a)). Further to those conditions, equilibrium in the residential area requires that everyone who lives there is employed,

$$\int_{\mathcal{X}_o^e} n_o^e dx = \bar{N}, \tag{5.7}$$

and that the allocation of urban land between the CBD and the residential area is determined competitively. Adapting the argument of section 2.2.4, the second requirement implies

$$R_o^e \ge R_1^e \ for\ 0 \le x \le \bar{x}_o^e \ and\ R_o^e \langle \bar{x}_o^e \rangle = R_1^e \langle \bar{x}_o^e \rangle , \tag{5.8}$$

where R_o is given by (5.6(b)) and R_1 is the rent in the residential sector.

5.2.1 Comparative Statics

These modifications can change the comparative statics of chapter three in some important ways. For example, if we keep all parameters except total population size fixed, (3.20) and (3.27) are now respectively written as [5]

$$\sigma \langle \bar{x}_o^e \rangle \frac{dR_1^e \langle \bar{x}_o^e \rangle}{d\bar{N}} + \int_{\mathcal{X}_1^e} \frac{\partial \sigma}{\partial x} \frac{dR_1^e}{d\bar{N}} dx = 1. \tag{5.9}$$

$$(1-\bar{k})\frac{d\overline{\mathrm{DLR}}^e}{d\bar{N}} + \left(\int_{\mathcal{X}_1^e} n_1^e \frac{\partial e^e}{\partial \bar{U}} dx\right) \frac{d\bar{U}^e}{d\bar{N}} = \left(\int_{\mathcal{X}_o^e} Z_o^e dx\right) \frac{dS}{d\bar{N}} - \frac{\bar{k}}{\bar{N}}\overline{\mathrm{DLR}}^e. \tag{5.10}$$

When there are no agglomeration economies ($dS/d\bar{N} = 0$), (5.10) reduces to the corresponding condition (3.27) and result 3.7 remains valid. In fact, the crowding effect is even stronger relative to the case of constant per–capita endowment because land is now required not only for housing but for production

[5] For (5.9), totally differentiate

$$\int_{\mathcal{X}_1^e} n_1^e dx = \bar{N}$$

and follow the procedure which led to (3.27) taking into account that all parameters except total population size are fixed. For (5.10), see section 5.5.1 in the appendix.

as well in order to maintain a fixed per–capita supply of the composite good as population increases. This is reflected by the increased land rent and cannot be observed directly from (5.10). When agglomeration economies are present ($dS/d\bar{N} > 0$), a wide range of possibilities for the behaviour of the equilibrium utility level with respect to total population size is opened. Under such economies, equilibrium utility can increase if the positive effects of further agglomeration outweigh the negative effects of further crowding as represented by $\overline{\mathrm{DLR}}^e/\bar{N}$.[6] If agglomeration economies are sufficiently strong relative to the impact of crowding for small populations and sufficiently weak for large populations, equilibrium utility can first increase and then decrease as total population size increases. For cities that satisfy those conditions there is an optimal city size—a subject further elaborated in chapters nine to twelve.

These modifications of comparative statics in the presence of agglomeration economies can solve the puzzle of the empirically observed negative correlation between urban population size and the density gradient, which we discussed in section 3.3.3. Since, with production and increasing returns, utility can *increase with population size*, the density gradient $1/\bar{\alpha}f[\bar{U}^e, \bar{\gamma}]$ in (2.25) of section 2.3.4 can *decrease with population size* because $\partial f/\partial U > 0$. The importance of this observation is that, in contrast to other explanations of the puzzle which invoked reasons outside of the monocentric domain (see section 3.3.3), we are offering here an explanation *within the monocentric domain*. For one should not identify monocentricity with constant returns to scale in production. The monocentric model is not only about the ways *housing* is distributed around a single centre, but rather about how *urban activities* are located around it. We believe that one of the reasons for the harsh criticism extended by Mieszkowski and Mills (1993) against the monocentric paradigm stems from the failure to recognise this observation.[7]

5.2.2 Sudden Urban Growth

The previous discussion indicates that introducing production enriches the model substantially, and it allows for a deeper understanding of some urban phenomena. As a further example, consider the case in which the shift factor $S[\bar{N}]$ first increases and then decreases as total population size increases. For such cities, the advantages of further agglomeration become stronger as the city grows up to a certain population size, beyond which they begin to weaken progressively. Under these circumstances, using condition (5.10) applied to the case of a closed owner city, the negative effects of further crowding can outweigh the positive effects of further agglomeration for sufficiently small city size. However, as $S[\bar{N}]$ continues to increase with city size, this condition can be reversed over an intermediate range of population. Over that range, where the positive effects of further agglomeration dominate the negative effects of further crowding, the

[6] Comparative statics for a closed renter city of this type, as well as details about possible utility–size profiles can be found in Papageorgiou (1980).

[7] We further discuss and expand on these matters in chapter eight.

equilibrium utility level increases with city size. We thus obtain an equilibrium utility level which changes direction twice. As the city population grows, equilibrium utility first declines, then increases and, finally, declines once again.[8] Such an equilibrium utility profile can account for *sudden urban growth* which, without exception, has characterised the observed evolution of large cities.[9]

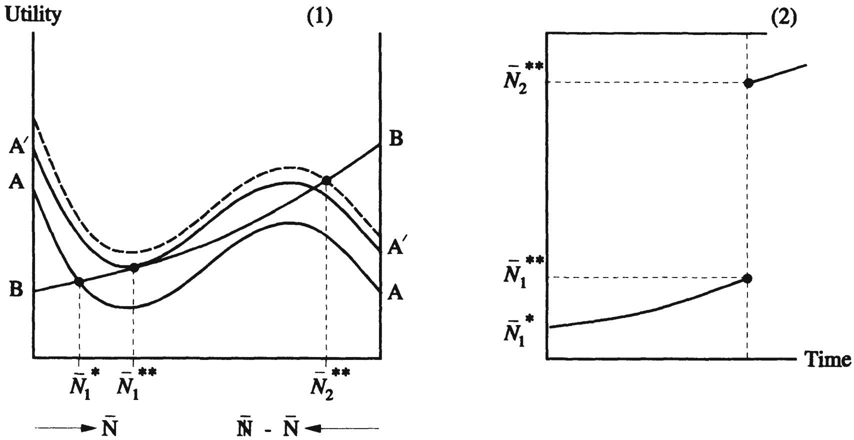

FIGURE 5.2. Discontinuity in the Equilibrium Population Growth Path.

Consider a region of total population size ℕ which is partitioned into urban population $\bar{N}$ and agricultural population $\mathbb{N} - \bar{N}$. The urban population is concentrated in a single city. Figure 5.2(1) presents the utility profiles of the two sectors over the range of partitioning the total population into urban and agricultural. The width of that figure corresponds to ℕ, and urban population is measured from left to right. Line AA represents an urban utility profile, and BB a corresponding agricultural utility profile. The latter is consistent with the idea that a relatively smaller agricultural population, which corresponds to a higher level of urbanisation, enjoys higher utility. If migration between the two sectors is costless, the regional equilibrium partitioning of the total population between urban and agricultural will be found at an intersection of the two utility profiles—where migration stops since the urban and agricultural utility levels

[8] Such configuration can arise under more sophisticated specifications of the production function at the centre. For example, it can be obtained when the scale economies are generated by the concentration of monopolistically competitive firms producing differentiated products (see Hadar (1997), Tabuchi (forthcoming) and, also, our chapter twelve).

[9] The connection between this utility profile and sudden urban growth has been proposed by Casetti (1980) and Papageorgiou (1980). All cities greater than or equal to one million underwent sudden urban growth. In 1860 there were five such cities, namely, Berlin, London, Paris, Peking and Vienna. In 1960 there were one hundred and nine, and in 1975 one hundred and ninety one. As the degree of urbanisation increases around the world, the number of such cities appears to increase at an increasing rate.

are equal. There is just one intersection of AA with BB, which defines the stable equilibrium partition $\bar{N}_1^*$.

The reader should imagine that, over time, the total population increases so that the length of the horizontal segment that measures total population in figure 5.2(1) widens. The significant effect of this change is that the urban utility profile shifts upwards *relative* to the agricultural utility profile as total population increases because agricultural utility declines with total population size. Thus if we measure the partition between urban and agricultural population in terms of percentages, rather than absolute numbers, $\overline{\mathbb{N}}$ will remain fixed through time (say, $\overline{\mathbb{N}} = 1$) while the urban utility profile will shift upwards. Figure 5.2(1) shows these relative shifts. Suppose we begin with a configuration in which the urban utility profile is lower than the agricultural utility profile for any population partition. Then the region is purely agricultural at the beginning. Now let the urban utility profile shift gradually upward until the equilibrium partitioning will be found at $\bar{N}_1^*$. Further gradual rise of the urban utility profile from AA toward A′A′ implies further gradual increase of the equilibrium urban population, until it reaches the critical level $\bar{N}_1^{**}$. Beyond that, we obtain a discontinuity in the equilibrium urban population path from $\bar{N}_1^{**}$ to $\bar{N}_2^{**}$, as also shown in figure 5.2(2). This discontinuity can account for sudden urban growth if the rate of population adjustment is proportional to the difference between the two utility profiles.

5.3 Heterogeneous Population

In this section we keep identical preferences for individuals and introduce heterogeneity with respect to income. The basis for our discussion is provided by von Thünen's (1826) classic framework of concentric agricultural crop rings adapted to the urban land use. The same is true for the general discussions of Alonso (1964) and Muth (1969), as well as for the specific models of Beckmann (1969) and Montesano (1972), among others, in which analytical rent and density functions were derived for cities with preference–homogeneous and income–heterogeneous populations.

5.3.1 The Slope Test

Given any two adjacent rings which accomodate two distinct socioeconomic groups, the group located closer to the centre must have a steeper equilibrium rent. Using figure 5.2, if AA represents the equilibrium rent function of group A and BB the one of group B, households of group A reside to the left of $\bar{x}_A$ and closer to the centre than households of group B, which reside to the right of $\bar{x}_A$. This reflects the principle that, in equilibrium, land is supplied to the highest bidder.[10]

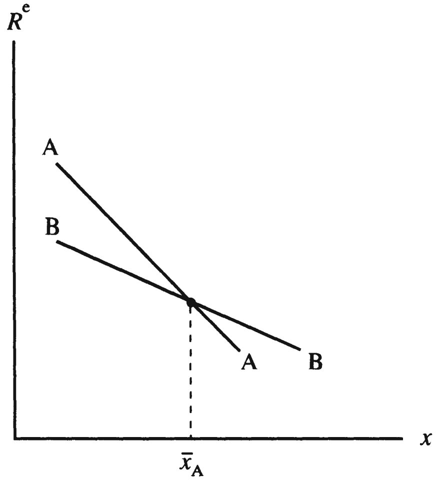

FIGURE 5.3. Competition for Urban Land.

The relationship between the steepness of the rent gradient and centrality provides us with the means for determining the equilibrium land use in a city with heterogeneous population. First, notice that income differences may well imply corresponding transportation cost differences: *ceteris paribus,* individuals with higher income can incur higher transportation cost. Therefore, unlike chapter two, the transportation cost function here is generalised to include income as an argument. Following Pines (1976), recall that the slope of the equilibrium rent is given by

$$\frac{dR^e}{dx} = -\frac{1}{H^e}\frac{\partial T^e}{\partial x}, \tag{5.11}$$

according to result 2.2 adjusted for the generalised transportation cost function. Differentiating (5.11) with respect to income, we obtain

$$\frac{\partial}{\partial \bar{Y}}\left(\frac{dR^e}{dx}\right) = \frac{1}{H^e\bar{Y}^e}\frac{\partial T^e}{\partial x}\left(\eta^e_{H:\bar{Y}} - \eta^e_{t:\bar{Y}}\right) \tag{5.12}$$

where $\eta_{H:\bar{Y}}$ is the *income elasticity of housing,* $\eta_{t:\bar{Y}}$ is the *income elasticity of the transportation rate.* Both elasticities are positive since, by assumption 2.2, housing is normal and since the transportation rate must increase with income. In consequence, the sign on the RHS of (5.12) is ambiguous.

[10] Steeper slope is necessary but not sufficient for being more centrally located. This happens because, even though the equilibrium rent of one group can be steeper than that of another, the flat rent can also be everywhere higher than the steep one. In this case, the group with the steep rent will be outbid everywhere.

Condition (5.12) cannot provide a criterion for determining the spatial equilibrium order unless the difference of the two elasticities on its RHS can be signed independently of the equilibrium allocation. With this in mind, if the income elasticity of housing dominates, socioeconomic status will increase away from the centre in equilibrium because the rent gradient becomes flatter as income increases. If, on the other hand, the income elasticity of the transportation rate dominates, socioeconomic status will decrease away from the centre. Since the income elasticity of housing reflects the relative importance of housing for individuals, while the income elasticity of the transportation rate reflects the relative importance of centrality, we conclude that, in societies where housing is relatively more important than centrality, rich live in the periphery. This might be one of the reasons why socioeconomic status roughly increases with distance from the centre in North American cities. If, on the other hand, centrality is relatively more important than housing, rich live in central locations as in South American cities where socioeconomic status roughly decreases with distance from the centre.

5.3.2 *The Value of Time*

One compelling reason why the transportation cost function must increase with income is that, when the length of the working day is flexible, the direct money cost value of time spent in commuting is evaluated by the individual at the wage rate. Let us consider a simple case in which the transportation cost function includes only the value of time spent in commuting. Given that the length of the working day is flexible, the transportation rate at x must be proportional to the wage rate $\bar{\omega}$ and inversely proportional to the *speed of traffic* $s\,[x]$ at x:

$$t = \frac{\bar{\omega}}{s}. \tag{5.13}$$

If, in addition, income is composed of wage only, (5.12) becomes

$$\frac{\partial}{\partial \bar{\omega}} \left(\frac{dR^e}{dx} \right) = \frac{1}{H^e s} \left(\eta^e_{H:\bar{Y}} - 1 \right). \tag{5.14}$$

It follows that the rich will prefer central (peripheral) locations in equilibrium if the income elasticity of housing is smaller (greater) than unitary—which is the result derived by Becker (1965) and Muth (1969).

We now turn to the case where the length of the working day is fixed institutionally. As we explained in section 5.1.1, the time constraint in this case can be substituted for leisure in the utility function. Utility now depends explicitly on location through *leisure* $\ell\,[x]$, which declines with distance from the centre because more time is spent on the road. Using the procedure of section 2.3.2, the equilibrium rent gradient is given by

$$\frac{dR^e}{dx} = -\frac{1}{H^e} \left(t^e + \frac{\partial e^e}{\partial \ell} \frac{d\ell}{dx} \right). \tag{5.15}$$

Since the time cost of travel has been subsumed in the utility function, the transportation rate on the RHS of (5.15) must represent money cost of travel. If we assume, as in the previous case, that the transportation cost function includes only the value of time spent in commuting, (5.15) is simplified and the effect of income on the slope of the equilibrium rent is given by [11]

$$\frac{\partial}{\partial \bar{Y}}\left(\frac{dR^e}{dx}\right) = \frac{\eta^e_{E:\ell}}{H^e \ell}\frac{d\ell}{dx}\left(\eta^e_{H:\bar{Y}} - \eta^e_{E_\ell:\bar{Y}}\right). \tag{5.16}$$

We know that both $d\ell/dx$ and $\partial e/\partial \ell$ are negative, the latter because leisure is valued by individuals. Furthermore, $\eta_{E_\ell:\bar{Y}}$ is negative (positive) if leisure is normal (inferior). Taking these into account, (5.16) implies that if leisure is inferior then the equilibrium rent gradient flattens with income, and the rich live further away from the centre than the poor. This is intuitively plausible since, with leisure being an inferior good, the higher is the income the less leisure is demanded, which implies living where ℓ is small, that is, at more distant locations. If, on the other hand, leisure is normal then the effect of income on the rent gradient is ambiguous, depending on the relative importance individuals assign to housing and leisure. In particular, if leisure is valued strongly enough relative to housing, the spatial order will be reversed and rich will prefer central locations.[12]

5.3.3 Comparative Statics

The comparative statics of a closed renter city in the case of income heterogeneity have been developed by Hartwick, Schweizer and Varaiya (1976), and they include several interesting results which are analogous to von Thünen's (1826) about concentric agricultural crop rings. These results can be explained intuitively by using the comparative statics of chapter two. We shall discuss the case of three socioeconomic groups, which can be directly extended to any number. Let us assume that $\bar{Y}_1 < \bar{Y}_2 < \bar{Y}_3$ so that $\bar{U}^e_1 < \bar{U}^e_2 < \bar{U}^e_3$. If, further, transportation cost does not depend on income, (5.12) implies that socioeconomic status increases away from the centre in equilibrium. Thus the boundaries of the concentric zone pattern must obey $0 < \bar{x}^e_1 < \bar{x}^e_2 < \bar{x}^e_3$ as in figure 5.3.

[11] In this expression $E_\ell \equiv \partial e/\partial \ell$, and we have taken into account that $E^e = \bar{Y}^e$ since the money cost of travel is zero.

[12] This type of analysis can be adjusted to include cases where the locational preference of households is based on reasons other than leisure, for example, location-specific environmental quality. Some of the existing studies that incorporate locational preference, such as Beckmann (1969) and Montesano (1972) among others, adopt a Cobb–Douglas utility or a monotonic transformation of it. As it turns out, with this specification, including the locational preference becomes redundant. For example, if $U = Z^a H^b \ell$ with $a + b = 1$, we have $\eta_{H:\bar{Y}} = 1$ and $\eta_{E_\ell:\bar{Y}} = -1$ implying that the RHS of (5.16) vanishes. Under these conditions, if the model incorporates the money cost of travel, the effect of income on location is determined by (5.12) irrespectively of the locational preference. When the money cost of travel is ignored, the distribution of households according to income becomes indeterminate under a Cobb–Douglas utility function.

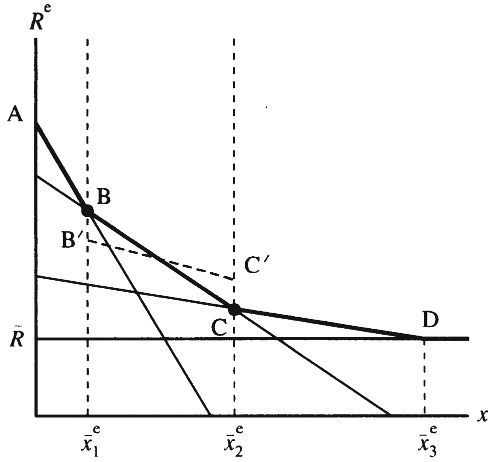

FIGURE 5.4. Effects of Rising Income on the Equilibrium Rent Profile.

Line ABCD in figure 5.3 represents the initial equilibrium rent schedule. Suppose that the income of the middle–income group increases. Holding $\bar{x}_1^e$ fixed, the equilibrium rent gradient must change from BC to B′C′ according to result 3.2. Since the low–income individuals at $\bar{x}_1^e$ can now outbid the middle–income individuals there, the low–income group expands to the right of that border. In consequence, more land becomes available for the low–income group. Since their supply of land increases, equilibrium population density and rent decline everywhere in the low–income area, while the equilibrium level of utility increases. At the outer border of the middle–income group $\bar{x}_2^e$, middle–income individuals can outbid the high–income individuals there. Thus the middle–income group expands to the right of $\bar{x}_2^e$. Since there is now less land available for the high–income group, equilibrium population density and rent increase everywhere in the high–income area, while the equilibrium level of utility decreases. Higher urban rent at the edge of the city $\bar{x}_3^e$ implies that the high–income area also expands outwards.

Result 5.1 (Hartwick, Schweizer and Varaiya (1976)): If the income of a particular socioeconomic group increases, all equilibrium boundaries move further away from the centre, equilibrium population density and rent decreases for those closer to the centre than the group for which income has increased, and increases for those further away. Consequently, the equilibrium level of utility increases for all lower income groups and decreases for all higher income groups.

Suppose now that the population of the middle–income group increases. Holding $\bar{x}_1^e$ and $\bar{x}_2^e$ fixed, the argument of section 3.2.1 implies that the equilibrium

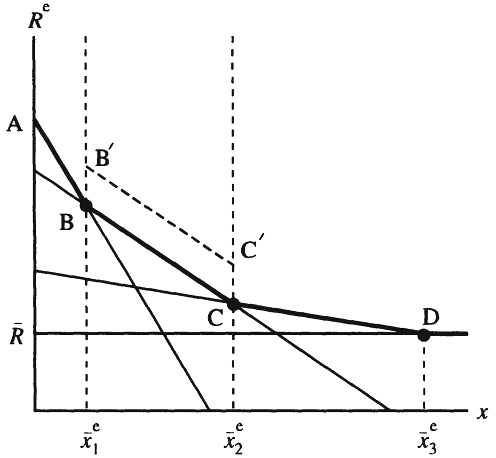

FIGURE 5.5. Effects of Rising Population on the Equilibrium Rent Profile.

rent gradient must change from BC to B′C′ in figure 5.4. It follows that equilibrium population density increases while the corresponding middle–income utility decreases. Since the middle–income equilibrium rent is higher at both borders with adjacent groups, the middle–income group expands in both directions. It follows that less land becomes available for both low– and high–income individuals. Consequently, equilibrium population density and rent increase in low– and high–income areas, while the corresponding levels of utility decrease. Higher urban rent at the edge of the city $\bar{x}_3^e$ implies that the high–income area also expands outwards.

Result 5.2 (Hartwick, Schweizer and Varaiya (1976)): If the population of a particular socioeconomic group increases, equilibrium boundaries between the centre and the group for which population has increased move closer to the centre, while the remaining boundaries move further away from the centre. Equilibrium population density and rent increases for all socioeconomic groups and their equilibrium level of utility decreases.

Notice how perfectly antisymmetric are the effects of disposable income and population. Also notice that the implications of rising income can be disturbing from a social welfare point of view. Namely, when the income of the lowest–income group increases, the utility level of all higher–income groups declines. Thus raising the income of the poorest will be against the interest of all other groups. In contrast, when the income of the highest–income group increases, the utility level of all lower–income groups improves.

Those implications must be qualified in the sense that they have been derived (1) for transportation cost independent of income and (2) for a closed renter city. If either of these assumptions does not apply, the above results are not necessarily valid.

(1) Generally speaking, when the income of a socio–economic group increases there are two spatial effects which, in combination, determine the comparative statics adjustments. Firstly, there is a rise in the demand for land, which creates pressure along the boundaries with the adjacent groups. Secondly, there is a change in the preference for location, which alters the slope of the equilibrium rent. Our conclusions about the effect of an income rise on utilities depend on how strongly the demand for land increases. If the increase in the demand for land is not strong enough, raising the income of the rich improves the utility of the poor. This, in fact, happens when transportation cost is independent of income. However, if transportation cost depends on income, the increase in the demand for land can become strong enough to reverse result 5.1 in the sense that, raising the income of the rich, worsens the utility of the poor. We discuss this possibility following Arnott, MacKinnon and Wheaton (1978).

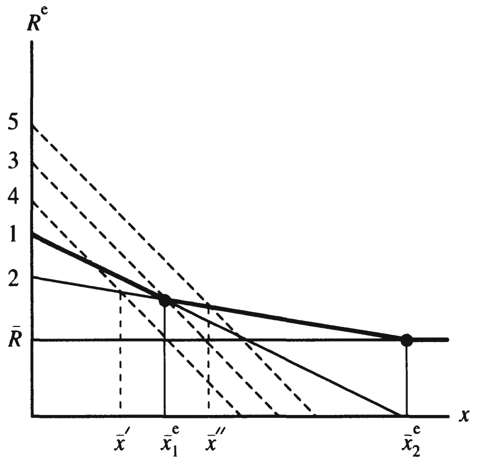

FIGURE 5.6. Welfare Effects of Rising Income.

Consider figure 5.5. There are two equilibrium rent functions, 1 and 2, defining two income groups with common boundary $\bar{x}_1^e$. Suppose that, for some reason, the bid rent of the first group rotates clockwise to 3, thus causing individuals of the first group to pay higher rent everywhere within their income zone. If the reason behind that change does not generate a demand for land sufficiently higher to compensate for the rent increase, the equilibrium rent will shift downward to 4. The new boundary $\bar{x}'$ allows more space for individuals of the second group. In consequence, those individuals will enjoy higher utility. With the first group representing the poor and the second the rich, this describes result 5.1, namely, that decreasing (increasing) the income of the poor will improve (reduce) the utility of the rich. Even if centrality dominates in (5.12) and the spatial order reverses so that the first group in figure 5.5 represents the rich and the second the poor, result 5.1 still holds because increasing the income of the rich will augment the utility of the poor. It is conceivable, nevertheless, that higher income can generate a demand for land sufficiently higher to compensate for the rent increase. Then the equilibrium rent function 3 will be observed for the rich, and the welfare of the poor will remain unaffected. An even stronger effect of rising income upon the demand for land will shift the rent of the rich upward to 5. The new boundary $\bar{x}''$ allows less space for the poor. In consequence, the poor will suffer a loss of utility instead of a gain as before.

(2) When we deal with a closed mixed city or a closed owner city, result 3.4 states that an increase of total population in a group need not cause an equilibrium rent increase near the outer border of that group. More importantly, with redistribution of rent, the very definition of socioeconomic groups becomes endogenous since income itself is endogenous. It may happen, for example, that the first group in figure 5.5 represents the land owners. Now, an increase in the population size of the second group (which is associated with an increase of the differential land rent) implies that the income of individuals in the first group increases. In consequence, the demand for land on the left of $\bar{x}_1^e$ will increase, and if this effect is strong enough then $\bar{x}_1^e$ will move toward $\bar{x}''$—rather than $\bar{x}'$ as predicted by result 5.2. Alternatively, suppose that the second group in figure 5.5 represents the land owners. If the population size of the first group increases, so does the income of individuals in the second group. This increase of income for the second group can more than offset the negative effect predicted by result 5.2, as we show by counter-example in section 5.5.2 of the appendix.

5.4 The City as a Central Place System

Until now we have dealt with a monocentric city where every urban location is characterised by a single distance to the only centre. But as Anas, Arnott and Small (forthcoming, p. 8) observe, monocentricity in the real world is a matter of spatial resolution: although at a coarse level the spatial structure of the city may be described by an homogeneous trend falling away from the CBD, at a more detailed level of resolution it becomes an heterogeneous clustering of economic activity: cities at this level appear to be polycentric rather than monocentric. In this section we develop the model of a polycentric city which is obtained by extending the standard monocentric model of chapter two in the context of central place theory.

5.4.1 The Model

We shall generalise the monocentric city model of sections 2.1 and 2.2 using Papageorgiou (1971). The crucial difference between the two models is that, here, urban residents interact with a *number* of centres instead of a *single* one as in chapter two. The location of centres is exogenously determined as in the case of the monocentric city. These centres represent the only places for the provision of goods and services in the city, and they form an $\bar{n}$–order central place hierarchy where $\bar{n}$ is the *number of levels* in that hierarchy. Order one corresponds to the lowest–order centres in the city, the neighbourhood malls, while order $\bar{n}$ corresponds to the single highest–order centre, the CBD. We assume as Christaller (1933) did that centres of any particular order contain *all goods and services* provided by lower–order centres. Thus a centre of order j operates in effect as a centre of order $1, ..., j$.[13]

The goods and services of any particular order $j = 1, ..., \bar{n}$ form the *composite good of order* j, any amount Z_j of which can be found at the same quality and the same *price* $\bar{P}_j$ in all centres that provide goods and services of that particular order. All individuals consume all types of composite good. Consequently the interaction between any individual and the centres is determined by a set of frequencies $\mathbf{G} \equiv (G_1, ..., G_{\bar{n}})$, where G_j denotes the individual's *number of trips per unit of time* for the purpose of acquiring the composite good of order j. Those frequencies of interaction decrease as the order increases.[14] Thus the highest frequency G_1 corresponds to the acquisition of lowest–order goods and services.

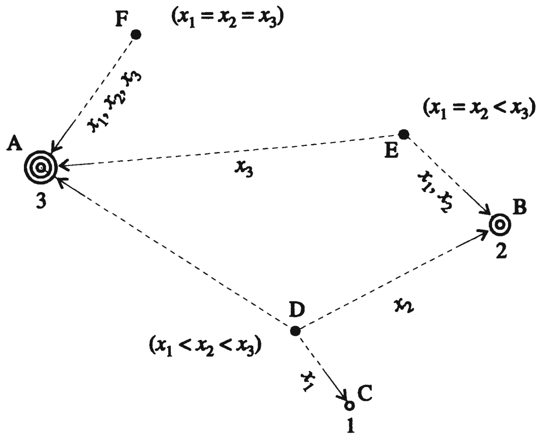

FIGURE 5.7. Individual Trip Patterns.

Since quality and price of the same good is the same anywhere it is sold, individuals will interact with the closest centre that provides the composite good

[13]For example, consider a three order hierarchy in which the CBD of a city represents its highest order centre, large regional malls represent the middle order and small local malls represent the lowest order. Then we expect that, roughly speaking, all goods and services provided locally can also be found in a regional mall but not vice-versa; and that all goods and services provided regionally can also be found in the CBD but not vice–versa.

[14]For example, one is expected to visit more frequently a neigbourhood store for milk (a low-order good) and less frequently a concert hall (a high–order good).

of order $j = 1, ..., \bar{n}$ whenever it is necessary to acquire the goods and services of this order. It follows that every residential location in the city corresponds to a unique set of distances $\mathbf{x} \equiv (x_1, ..., x_{\bar{n}})$, where x_j denotes the *distance between the residential location and the closest centre that provides the composite good of order j*. Examples in the case of a three–order hierarchy are shown in figure 5.7. There are three centres on the map, one third–order at A, one second–order at B and one first–order at C. Three residential locations are also marked as D, E and F, and we have indicated the distance vector $\mathbf{x}$ for each residential location. Notice that since a higher–order centre contains all lower–order goods, the distance travelled to buy a good of particular order cannot be shorter than the distance travelled to buy a good of lower order.[15] In general, locations closer to high–order centres are more advantageous in terms of their accessibility to goods and services.

We have explained how location in the polycentric city can be related to a vector of frequencies and a vector of distances. The frequency vector can be treated either as constant or, more generally, as a function of the distance vector. In any case, since the distance vector is defined over every location once the spatial distribution and the hierarchical properties of the centres are given, any residential location within that system can be characterised by a single vector $\mathbf{x}$ which replaces the single distance x from the centre used in the special case of the monocentric city. Consequently, the transportation cost function is now written as $T[\mathbf{x}, \bar{\alpha}]$ with $t_j \equiv \partial T / \partial x_j$ for $j = 1, ..., \bar{n}$ representing the *transportation rates* associated with trips to various–order centres. Following assumption 2.2, $T[\mathbf{x}, \bar{\alpha}]$ is differentiable, a strictly increasing and unbounded function of x_j and a strictly decreasing function of $\bar{\alpha}$. Because improvements in the quality of transportation infrastructure lower both the cost of transportation and the transportation rates, we also have $\partial t_j / \partial \bar{\alpha} < 0$ for $j = 1, ..., \bar{n}$.

Finally since preferences are defined on the entire set of composite goods provided by the various–order centres, utility is now given as $U = u[\mathbf{Z}, H, \bar{\gamma}]$, where $\mathbf{Z} \equiv (Z_1, ... Z_{\bar{n}})$ denotes the *amounts of composite good consumed* by an individual. Following assumption 2.4, $u[\mathbf{Z}, H, \bar{\gamma}]$ is differentiable, strictly increasing and strictly quasi–concave in its arguments. All commodities are normal. They are also essential for consumption, in the sense that a consumption bundle which includes positive amounts of all commodities is always preferable to a bundle which includes a zero amount for some commodity. With these modifications, the individual choice problem here is a straightforward extension of the corresponding problem (2.3):

$$V \equiv \max_{\mathbf{Z}, H} u[\mathbf{Z}, H, \bar{\gamma}] \text{ subject to } \sum_{j=1}^{\bar{n}} \bar{P}_j Z_j + RH \leq \bar{Y} - T[\mathbf{x}, \bar{\alpha}]. \tag{5.17}$$

[15] This property, together with the observation that lower–order goods and services can be supported by a smaller number of customers, means in the context of central place theory that the number of centres of a given order decreases for higher order. For example, in a three–order hierarchy, we have many small local malls representing order one, a few large regional malls representing order two and a single CBD representing order three.

This close relationship between the specifications of the monocentric and the polycentric city models will allow us to generalise some key results of chapter two.

5.4.2 Some Equilibrium Properties

A fundamental implication of moving from the monocentric to the polycentric urban model is the loss of rotational symmetry. For if we let the single highest–order centre of the polycentric city correspond to the centre of the monocentric city, locations equidistant from the former can no longer be treated as identical. Thus in order to translate the definition of competitive equilibrium (2.13) in the polycentric context, we must consider not only distance from the highest–order centre but orientation as well.[16] Although this complicates the definition of an equilibrium, it does not prevent us from determining the polycentric spatial distribution of land rent and population density to a considerable extent.

The Principle of Zero Marginal Location Costs

If we assume an homogeneous urban population as in chapter two, equilibrium utility is uniform across the urban area. It is straightforward to extend (2.8) into

$$e\left[\bar{\mathbf{P}}, R, \bar{U}, \bar{\gamma}\right] = \bar{Y} - T\left[\mathbf{x}, \bar{\alpha}\right] \tag{5.18}$$

where $\bar{\mathbf{P}} \equiv \left(\bar{P}_1, ..., \bar{P}_{\bar{n}}\right)$. This implies that the equilibrium rent gradient of the polycentric city is given as $R^e\left[\mathbf{x}\right]$. In equilibrium, for every residential location, total expenditure $e\left[\bar{\mathbf{P}}, R, \bar{U}, \bar{\gamma}\right] + T\left[\mathbf{x}, \bar{\alpha}\right]$ must be minimised. Taking the total differential of $e\left[\bar{\mathbf{P}}, R, \bar{U}, \bar{\gamma}\right] + T\left[\mathbf{x}, \bar{\alpha}\right]$ with respect to location, using (2.15) and equating the result to zero gives

$$\sum_{j=1}^{\bar{n}} \left(H^e \frac{\partial R^e}{\partial x_j} + t_j\right) dx_j = 0. \tag{5.19}$$

Since this equality must hold for arbitrary $d\mathbf{x}$, we have

$$H^e \frac{\partial R^e}{\partial x_j} + t_j = 0 \text{ for } j = 1, ..., \bar{n}, \tag{5.20}$$

which extends Muth's (1961) result 2.2 in the context of multiple centres.

---

[16] For example, in order to generalise the land–use constraint (2.13(b)), we must calculate the average population over the corresponding distance band, that is,

$$n^e = \frac{1}{2\pi} \int_0^{2\pi} \hat{n}^e \left[x, \phi\right] d\phi$$

where ϕ is an angle that specifies orientation ($\hat{n}$ is expressed in polar coordinates). Similarly

$$H^e = \frac{1}{2\pi} \int_0^{2\pi} \hat{H}^e \left[x, \phi\right] d\phi.$$

Also notice that the city border in that case must also depend on orientation, that is, $\bar{x}^e = \hat{x}\left[\phi\right]$.

Rent and Density Profiles

We can now use (5.20) and follow the procedure in section 2.3.2 to extend Muth's result 2.3:

$$\text{For } j = 1, ..., \bar{n}, \frac{\partial R^e}{\partial x_j} < 0; \text{ if } \frac{\partial t_j}{\partial x_j} \leq 0 \text{ then } \frac{\partial^2 R^e}{\partial x_j^2} > 0. \tag{5.21}$$

Moreover since $H^e = h\left[\bar{\mathbf{P}}, R, \bar{U}, \bar{\gamma}\right]$ and since the population density is the inverse of land per capita, we have

$$\frac{\partial D^e}{\partial x_j} = -\frac{1}{H^{e2}} \frac{\partial h^e}{\partial R} \frac{\partial R^e}{\partial x_j} \text{ for } j = 1, ..., \bar{n}. \tag{5.22}$$

Using (5.21) and taking into account that the substitution effect is negative, we obtain

$$\frac{\partial D^e}{\partial x_j} < 0 \text{ for } j = 1, ..., \bar{n} \tag{5.23}$$

which is an extension of Muth's (1961) result 2.5. Therefore, roughly speaking, both land rents and population densities fall away from the centres in equilibrium. However, as with the population density of a monocentric city, the population density of a polycentric city may not satisfy $\partial^2 D^e / \partial x_j^2 > 0$. It will if the price elasticity of the compensated demand for housing is constant, which can be established by following the argument of section 2.3.3. Under these circumstances the spatial distribution of population density is qualitatively similar to the distribution of land rent.

We are now ready to specify how equilibrium rent and density surfaces unfold over a polycentric city. We first notice that they attain a global maximum at the location of the highest–order centre. This follows directly from (5.21) and (5.23) together with the observation that $\mathbf{x} = \mathbf{0}$ holds only at the location of the highest–order centre. We also know that rents and densities that correspond to the locations of a given k order of centres decrease as the distances between those centres and higher than k–order centres increase. This follows once again from (5.21) and (5.23) if we take into account that rents and densities at zero distance from a k–order centre correspond to locations $\mathbf{x}$ such that $x_j = 0$ for $j = 1, ..., k$ and $x_j > 0$ for $x_j = k + 1, ..., \bar{n}$. Furthermore, if the equilibrium rent and density functions can be expressed in terms of linear combinations of distances to various–order centres then (i) local maxima correspond to the location of centres; (ii) there may be centres that do not correspond to local maxima; and (iii) maxima correspond only to the location of centres.[17]

Figure 5.8 illustrates the above principles over part of a three–order polycentric city that contains three first–order centres (at B, D and E), two second–order (at C and F) and one third–order (at A). For simplicity of exposition these

[17] See Papageorgiou (1971).

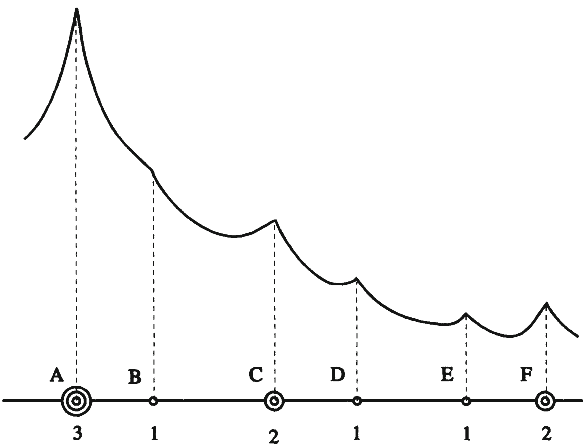

FIGURE 5.8. Polycentric Rent and Density Profile.

centres have been located along a straight line. The highest rent and density is found at the location of the highest–order centre. For the second–order centres C and F, rent and density is higher at C because it is closer to A. The locations B, D and E of the first–order centres have been drawn so that all three are equidistant from a centre that provides second–order goods, but they differ in terms of their distance from the CBD. Consequently rent and density becomes lower as we move from B to D and from D to E. Also notice that local maxima correspond to the location of centres (at C, D, E and F), that the centre at B does not correspond to a local maximum and that maxima correspond only to the location of centres. We can visualise equilibrium rent and density over a polycentric city as a mountain range in which higher peaks roughly correspond to higher–order centres, while the single highest peak corresponds to the single highest–order centre in the system.

5.4.3 Negative Exponential Rent and Density

Suppose as in Papageorgiou and Pines (1989) that the expenditure function yields a unitary price elasticity of the compensated demand for housing. Then (2.22) holds, while (2.23) and (2.24) become respectively

$$e\left[\bar{\mathbf{P}}, R^e, \bar{U}^e, \bar{\gamma}\right] = f\left[\bar{\mathbf{P}}, \bar{U}^e, \bar{\gamma}\right] \ln R^e + F\left[\bar{\mathbf{P}}, \bar{U}^e, \bar{\gamma}\right] \tag{5.24}$$

and

$$H^e = \frac{f\left[\bar{\mathbf{P}}, \bar{U}^e, \bar{\gamma}\right]}{R^e} \, . \tag{5.25}$$

Suppose now that the trip frequencies are fixed, $\bar{\mathbf{G}} \equiv \left(\bar{G}_1, ..., \bar{G}_{\bar{n}}\right)$, and that the transportation cost function is given by

$$T = \sum_{j=1}^{\bar{n}} \frac{\bar{G}_j x_j}{\bar{\alpha}} \, . \tag{5.26}$$

Then it follows from (5.20), (5.25) and (5.26) that

$$\frac{1}{R^e}\frac{\partial R^e}{\partial x_j} = -\frac{\bar{G}_j}{\bar{\alpha} f\left[\bar{\mathbf{P}}, \bar{U}^e, \bar{\gamma}\right]} \equiv -\bar{\delta}_j \text{ for } j = 1, ..., \bar{n}. \tag{5.27}$$

Since $f > 0$ by (5.25), we have $\bar{\delta}_j > 0$.[18] Solving (5.27) yields

$$R^e[\mathbf{x}] = R^e\langle\mathbf{0}\rangle \exp\left(-\sum_{j=1}^{\bar{n}} \bar{\delta}_j x_j\right). \tag{5.28}$$

Finally since the population density is the inverse of land per capita, we can use (5.25) and (5.28) to obtain

$$D^e[\mathbf{x}] = D^e\langle\mathbf{0}\rangle \exp\left(-\sum_{j=1}^{\bar{n}} \bar{\delta}_j x_j\right) \tag{5.29}$$

which represents the generalised Clark (1951) formula. As in section 2.3.4, the generalised rent and density gradients are the same because the price elasticity of the compensated demand for housing is unitary by assumption. Since these functions are expressed in terms of linear combinations of distances to various–order centres, all the spatial properties described in the previous section are valid.[19]

We close this section by pointing out the possibility of a significant relationship between the monocentric and the polycentric negative exponential formulae. It is empirically known that, on average, the spacing of lower–order centres increases with their distance from higher–order centres.[20] If this relationship is

---

[18] Since the observed frequencies of interaction decrease as the order increases, it must also be true that estimates of $\bar{\delta}$ will be ordered as $\bar{\delta}_1 > \bar{\delta}_2 > \cdots > \bar{\delta}_{\bar{n}}$.

[19] To our knowledge, (5.29) is the only polycentric exponential density formulation which is obtained as a result of a model. The existing alternatives are variations of

$$D[\mathbf{x}] = \sum_{j=1}^{\bar{n}} \bar{D}_j \exp\left(-\bar{\delta}_j x_j\right),$$

first introduced by Griffith (1981) and subsequently employed by several authors in the context of *employment* rather than *service* centres as proposed here. For further information about empirical polycentric density specifications see Anas, Arnott and Small (forthcoming).

[20] Traditionally, this spatial arrangement has been attributed to the existence of agglomeration economies operating within hierarchical systems of this type (see for example Isard (1956, pp. 270–274), esp. the diagram on p. 272). To fix ideas, consider a two–level hierarchical system with a single highest–order centre corresponding to the CBD of a city or to the largest city of a region. We know that population densities fall away from the highest–order centre—not only under the urban interpretation but under the regional as well (see Bogue (1949) for an early regional reference). It is also known that the density of subcentres strictly increases with the density of population (see for example Stephan (1988) and Gusein–Zade (1993)). This happens because, as the population density increases, the market area necessary to support a subcentre decreases. Then, since the population density declines with distance from the highest–order centre, the density of subcentres decreases as distance from the highest–order centre increases.

assumed to be linear then the monocentric Clark formula describes the general trend of its polycentric counterpart (see Papageorgiou (1971)). This provides some justification for using the monocentric formula to estimate the density gradient of polycentric cities—a standard practice in the vast literature on the subject.

5.5 Appendices

5.5.1 Proof of (5.10)

Using (2.8), (2.12) and (5.6(a)), we have

$$\begin{aligned} e\left[R_1^e, \bar{U}^e\right] + T_1 &= S\left[\bar{N}\right] \frac{\partial f_Z^e}{\partial n_o} (1 - T_o) + \bar{k}\overline{\mathrm{DLR}}^e \\ &= \frac{1}{\bar{N}} \left(\int_{\mathcal{X}_o^e} \left(S\left[\bar{N}\right] f_Z\left[\cdot\right](1 - T_o) - \theta R_o^e\right) dx + \bar{k}\overline{\mathrm{DLR}}^e \right) \end{aligned} \tag{5.30}$$

by using the zero–profit condition. Totally differentiating (5.30), and keeping all parameters other than total population size fixed, we obtain

$$\begin{aligned} &\frac{\partial e^e}{\partial R_1} \frac{dR_1^e}{d\bar{N}} + \frac{\partial e^e}{\partial \bar{U}} \frac{d\bar{U}^e}{d\bar{N}} \\ = & -\frac{1}{\bar{N}^2} \left(\int_{\mathcal{X}_o^e} \left(\bar{S}F_Z^e (1 - T_o) - \theta R_o^e\right) dx + \bar{k}\overline{\mathrm{DLR}}^e \right) \\ + & \frac{1}{\bar{N}} \left(\int_{\mathcal{X}_o^e} \left(\frac{dS}{d\bar{N}} F_Z^e (1 - T_o) - \theta \frac{dR_o^e}{d\bar{N}} \right) dx + \bar{k} \frac{\overline{\mathrm{DLR}}^e}{d\bar{N}} \right) \\ + & \frac{1}{\bar{N}} \left(\bar{S}F_Z^e (1 - T_o) - \theta R_o^e\right)\Big|_{x = \bar{x}_o^e} \frac{d\bar{x}_o^e}{d\bar{N}} \\ + & \frac{1}{\bar{N}} \int_{\mathcal{X}_o^e} \bar{S} \frac{\partial f_Z^e}{\partial n_o} \frac{dn_o^e}{d\bar{N}} (1 - T_o)\, dx. \end{aligned} \tag{5.31}$$

Recall that $n_1 \partial e / \partial R_1 = n_1 H_1 = \theta$ by (2.12) and by (2.15). Also notice that, keeping all parameters other than total population size fixed, we have

$$\frac{d\overline{\mathrm{DLR}}^e}{d\bar{N}} = \int_{\mathcal{X}_o^e} \theta \frac{dR_o^e}{d\bar{N}} dx + \int_{\mathcal{X}_1^e} \theta \frac{dR_1^e}{d\bar{N}} dx. \tag{5.32}$$

Multiplying both sides of (5.31) by n_1^e and integrating over the residential area, we obtain

$$
\begin{aligned}
(1-\bar{k})\frac{d\overline{\mathrm{DLR}}^e}{d\bar{N}} + \left(\int_{\mathcal{X}_1^e} n_1^e \frac{\partial e^e}{\partial \bar{U}} dx\right) \frac{d\bar{U}^e}{d\bar{N}} &= \\
\left(\int_{\mathcal{X}_o^e} Z_o^e dx\right) \frac{dS}{d\bar{N}} - \frac{\bar{k}}{\bar{N}}\overline{\mathrm{DLR}}^e &- \\
\frac{1}{\bar{N}} \int_{\mathcal{X}_o^e} \left(\bar{S}F_Z^e (1-T_o) - \theta R_o^e\right) dx &+ \qquad (5.33)\\
\left(\bar{S}F_Z^e (1-T_o) - \theta R_o^e\right)\Big|_{x=\bar{x}_o^e} \frac{d\bar{x}_o^e}{d\bar{N}} &+ \\
\int_{\mathcal{X}_o^e} \bar{S}\frac{\partial f_Z^e}{\partial n_o}\frac{dn_o^e}{d\bar{N}} (1-T_o)\, dx. &
\end{aligned}
$$

In order to establish (5.10) we must show that the sum of the three terms below the first line of (5.33) equals zero. Taking into account the zero–profit condition, the first and second terms can be written as

$$
-\frac{1}{\bar{N}} \int_{\mathcal{X}_o^e} \left(\bar{S}F_Z^e (1-T_o) - \theta R_o^e\right) dx = -\bar{\Omega} \qquad (5.34)
$$

and

$$
\left(\bar{S}F_Z^e (1-T_o) - \theta R_o^e\right)\Big|_{x=\bar{x}_o^e} \frac{d\bar{x}_o^e}{d\bar{N}} = n_o^e \left\langle \bar{x}_o^e \right\rangle \bar{\Omega}\frac{d\bar{x}_o^e}{d\bar{N}} \qquad (5.35)
$$

respectively. The third term can be written as

$$
\int_{\mathcal{X}_o^e} \bar{S}\frac{\partial f_Z^e}{\partial n_o}\frac{dn_o^e}{d\bar{N}} (1-T_o)\, dx = \bar{\Omega}\left(1 - n_o^e \left\langle \bar{x}_o^e \right\rangle \frac{d\bar{x}_o^e}{d\bar{N}}\right) \qquad (5.36)
$$

because of (5.6(*a*)) and because

$$
\int_{\mathcal{X}_o^e} \frac{dn_o^e}{d\bar{N}} dx = 1 - n_o^e \left\langle \bar{x}_o^e \right\rangle \frac{d\bar{x}_o^e}{d\bar{N}} \qquad (5.37)
$$

by (5.7). We conclude that (5.10) must hold since the sum of the three terms on the RHS of (5.34)–(5.36) equals zero.

5.5.2 Counter–Example to Result 5.2

In this example, we assume that there are two groups with identical preferences, initial endowment and cost of transportation, and that individuals of the first group rent the land from the second group which owns commonly all the urban land. The first group is located between the centre and $\bar{x}_1$, the second between $\bar{x}_1$ and $\bar{x}_2, 0 < \bar{x}_1 < \bar{x}_2$. Following the example of section 2.3.5, let $u[Z, H] = \min(Z, H)$ for both groups. Also let $\theta = 2x, T_1 = T_2 = x, \bar{R} = 0$ and $\bar{\Omega}_1 = \bar{\Omega}_2 = 10$. In equilibrium, we have

$$\bar{Z}_1^e = \bar{H}_1^e = \frac{1}{\bar{N}_1} \int_{\mathcal{X}_1^e} \theta dx = \frac{\bar{x}_1^{e2}}{\bar{N}_1} = \bar{U}_1^e \tag{5.38}$$

and

$$\bar{Z}_2^e = \bar{H}_2^e = \frac{1}{\bar{N}_2} \int_{\mathcal{X}_2^e} \theta dx = \frac{\bar{x}_2^{e2} - \bar{x}_1^{e2}}{\bar{N}_2} = \bar{U}_2^e. \tag{5.39}$$

We next calculate the aggregate resource constraint for each of the two groups. Since the first group rents from the second, these constraints are given as

$$\int_{\mathcal{X}_1^e} n_1^e \bar{Z}_1^e dx + \overline{\mathrm{ATC}}_1^e + \overline{\mathrm{DLR}}_1^e = 10\bar{N}_1 \tag{5.40}$$

and

$$\int_{\mathcal{X}_2^e} n_2^e \bar{Z}_2^e dx + \overline{\mathrm{ATC}}_2^e = 10\bar{N}_2 + \overline{\mathrm{DLR}}_1^e. \tag{5.41}$$

Taking into account the condition on land $n_i H_i = \theta = 2x$ for $i = 1, 2$, the aggregate expenditures on consumption are respectively

$$\int_{\mathcal{X}_1^e} n_1^e \bar{Z}_1^e dx = \int_{\mathcal{X}_1^e} n_1^e \bar{H}_1^e dx = \bar{x}_1^{e2} \tag{5.42}$$

and

$$\int_{\mathcal{X}_2^e} n_2^e \bar{Z}_2^e dx = \int_{\mathcal{X}_2^e} n_2^e \bar{H}_2^e dx = \bar{x}_2^{e2} - \bar{x}_1^{e2}. \tag{5.43}$$

The aggregate expenditures on transportation can be written as

$$\overline{\mathrm{ATC}}_1^e = \int_{\mathcal{X}_1^e} n_1^e T_1 dx = \int_{\mathcal{X}_1^e} \frac{2x^2}{\bar{H}_1^e} dx \underset{(5.38)}{=} \frac{2}{3} \bar{x}_1^e \bar{N}_1 \tag{5.44}$$

and

$$\overline{\mathrm{ATC}}_2^e = \int_{\mathcal{X}_2^e} n_2^e T_2 dx = \int_{\mathcal{X}_2^e} \frac{2x^2}{\bar{H}_2^e} dx \underset{(5.39)}{=} \frac{2}{3} \frac{\bar{x}_2^{e3} - \bar{x}_1^{e3}}{\bar{x}_2^{e2} - \bar{x}_1^{e2}} \bar{N}_2. \tag{5.45}$$

It remains to calculate the differential land rent for the first group which appears in both resource constraints. Since the equilibrium locational costs $R_1^e \bar{H}_1^e + x$ in this model do not vary with location, and taking into account that the opportunity cost of land is zero, we have

$$R_1^e = \frac{\bar{x}_1^e - x}{\bar{H}_1^e} + R_1^e \langle \bar{x}_1^e \rangle \underset{(5.38)}{=} \frac{\bar{x}_1^e - x}{\bar{x}_1^{e2}} \bar{N}_1 + R_1^e \langle \bar{x}_1^e \rangle \tag{5.46}$$

TABLE 5.1. Numerical Example.

$\bar{N}_1$	$\bar{x}_1^e$	$\bar{x}_2^e$	$\bar{U}_2^e$
.33	1.0	9.02	4.02
1.39	2.0	9.20	4.03
3.39	3.0	9.51	4.07
6.49	4.0	9.97	4.17
11.40	5.0	10.63	4.40
19.25	6.0	11.57	4.89
32.69	7.0	12.98	5.59
59.40	8.0	15.36	8.98
85.42	8.5	17.36	11.46
135.51	9.0	20.72	17.42
275.94	9.5	28.33	35.60

and

$$R_2^e = \frac{\bar{x}_2^e - x}{\bar{H}_2^e} + R_2^e \langle \bar{x}_2^e \rangle \underset{(5.39)}{=} \frac{\bar{x}_2^e - x}{\bar{x}_2^{e2} - \bar{x}_1^{e2}} \bar{N}_2. \tag{5.47}$$

Using the equality of the two rents at the interior border,

$$R_1^e \langle \bar{x}_1^e \rangle = \frac{\bar{x}_2^e - \bar{x}_1^e}{\bar{x}_2^{e2} - \bar{x}_1^{e2}} \bar{N}_2 = \frac{\bar{N}_2}{\bar{x}_2^e + \bar{x}_1^e}. \tag{5.48}$$

Therefore,

$$\overline{\text{DLR}}_1^e = \int_{\mathcal{X}_1^e} \theta R_1^e dx = \frac{1}{3} \bar{x}_1^e \bar{N}_1 + \bar{x}_1^{e2} R_1^e \left[x = \bar{x}_1^e\right] = \frac{1}{3} \bar{x}_1^e \bar{N}_1 + \frac{\bar{x}_1^{e2}}{\bar{x}_2^e + \bar{x}_1^e} \bar{N}_2. \tag{5.49}$$

We can now use (5.42)–(5.45) and (5.49) to derive explicitly the aggregate resource constraints (5.40) and (5.41) as

$$\bar{x}_1^{e2} + \bar{x}_1^e \bar{N}_1 + \frac{\bar{x}_1^{e2}}{\bar{x}_2^e + \bar{x}_1^e} \bar{N}_2 = 10 \bar{N}_1 \tag{5.50}$$

and

$$\bar{x}_2^{e2} - \bar{x}_1^{e2} + \frac{2}{3} \frac{\bar{x}_2^{e3} - \bar{x}_1^{e3}}{\bar{x}_2^{e2} - \bar{x}_1^{e2}} \bar{N}_2 = 10 \bar{N}_2 + \frac{1}{3} \bar{x}_1^e \bar{N}_1 + \frac{\bar{x}_1^{e2}}{\bar{x}_2^e + \bar{x}_1^e} \bar{N}_2. \tag{5.51}$$

Solving (5.50) and (5.51) for $\bar{x}_1^e$ and $\bar{x}_2^e$, we calculate $\bar{U}_2^e$ from (5.39) as a function of $\bar{N}_1$ and $\bar{N}_2$. Some numerical results for variable $\bar{N}_1$ and fixed $\bar{N}_2$ appear in table 5.1. In this example, $\bar{x}_1^e$ and $\bar{x}_2^e$ move further away as predicted by result 5.2, but $\bar{U}_2^e$ increases with $\bar{N}_1$ instead of decreasing.

5.6 References

W. Alonso, 1964, *Location and Land Use.* Harvard University Press, Cambridge, Mass.

A. Anas, R. J. Arnott and K. A. Small, forthcoming, Urban Spatial Structure. *Journal of Economic Literature.*

R. J. Arnott, J. G. MacKinnon and W. C. Wheaton, 1978, The Welfare Implications of Spatial Interdependence. *Journal of Urban Economics* 5, 131–136.

G. Becker, 1965, A Theory of the Allocation of Time. *The Economic Journal* 75, 493–517.

M. Beckmann, 1969, On the Distribution of Urban Rent and Residential Density. *Journal of Economic Theory* 1, 60–67.

D. J. Bogue, 1949, *The Structure of the Metropolitan Community: A Study in Dominance and Subdominance.* University of Michigan, Ann Arbor.

E. Casetti, 1980, Equilibrium Population Partitions Between Urban and Agricultural Occupations. *Geographical Analysis* 12, 47–54.

W. Christaller, 1933, *Die Centralen Orte in Süddeutschland.* Translated in English as *Central Places in Southern Germany* by C. W. Baskin, Prentice–Hall, Englewood Cliffs, 1966.

A. K. Dixit, 1973, The Optimum Factory Town. *Bell Journal of Economics and Management Science* 4, 637–651.

D. A. Griffith, 1981, Evaluating the Transformation from a Monocentric to a Polycentric City. *Professional Geographer* 33, 189–196.

S. M. Gusein–Zade, 1993, Alternative Explanations of the Dependence of the Density of Centres on the Density of Population. *Journal of Regional Science* 33, 547–558.

Y. Hadar, 1997, The Formation of Cities. Tel–Aviv University, (unpublished).

J. Hartwick, U. Schweizer and P. Varaiya, 1976, Comparative Static Analysis of a Residential Economy with Several Classes. *Journal of Economic Theory* 13, 396–413.

J. V. Henderson, 1986, Efficiency of Resources Usage and City Size. *Journal of Urban Economics* 19, 47–70.

W. Isard, 1956, *Location and Space Economy.* MIT Press, Cambridge, Mass.

R. Latham and M. Yeates, 1970, Population Density in Metropolitan Toronto. *Geographical Analysis* 2, 177–185.

E. S. Mills, 1967, An Aggregative Model of Resource Allocation in a Metropolitan Area. *American Economic Review* 57, 197–210.

P. Mieszkowski and E. S. Mills, 1993, The Causes of Metropolitan Suburbanization. *Journal of Economic Perspectives* 7, 135–147.

A. Montesano, 1972, A Restatement of Beckmann's Model on the Distribution of Urban Rent and Population Density. *Journal of Economic Theory* 4, 329–354.

R. F. Muth, 1961, The Spatial Structure of the Housing Market. *Papers and Proceedings of the Regional Science Association* 7, 207–220.

R. F. Muth, 1969, *Cities and Housing.* The University of Chicago Press, Chicago.

B. E. Newling, 1966, Urban Growth and Spatial Structure: Mathematical Models and Empirical Evidence. *The Geographical Review* 56, 213–225.

Y. Y. Papageorgiou, 1971, The Population Density and Rent Distribution Models Within a Multicentre Framework. *Environment and Planning* 3, 267–282.

Y. Y. Papageorgiou, 1980, On Sudden Urban Growth. *Environment and Planning* A 12, 1035–1050.

Y. Y. Papageorgiou and D. Pines, 1989, The Exponential Density Function: First Principles, Comparative Statics and Empirical Evidence. *Journal of Urban Economics* 26, 264–268.

D. Pines, 1976, On the Spatial Distribution of Households According to Income. *Economic Geography* 51, 142–149.

E. G. Stephan, 1988, The Distribution of Service Establishments. *Journal of Regional Science* 28, 29–40.

T. Tabuchi, forthcoming, Urban Agglomeration and Dispersion: A Synthesis of Alonso and Krugman. *Journal of Urban Economics.*

J. H. von Thünen, 1826, *Der Isolierte Staadt in Beziehung auf Landwirtschaft und Nationalöconomie.* Translated in English as *von Thünen's Isolated State* by C. W. Wartenberg. Pergamon Press, Oxford, 1966.

6
Urban Housing

Housing consumption in chapter two was treated as being equivalent to current consumption of services rendered by land occupancy. This treatment seems to oversimplify reality in two respects. First, housing is a service created by land, capital and perhaps labor combined to produce a house. In this sense it appears that the treatment of housing by Muth (1961, 1969), and Mills (1967) is more appropriate than that of Alonso (1964) where housing services are provided by land only. In particular the analysis of chapter two, which is similar to Alonso's treatment, can only be used to examine the spatial variation of population density (number of households per unit of land) because it is not refined enough to make the important distinction between the spatial variation of structural density (density of housing services per unit of land) and population crowding in the housing stock. Second, housing in chapter two has been treated as perfectly malleable, being adjusted costlessly in response to parametric changes. In fact housing is not perfectly malleable and its adjustment is costly. Therefore a more detailed study of housing requires a clear distinction to be made between the housing stock and the flow of housing services. This distinction calls for a dynamic framework.

As we will show, the above two extensions need not make our chapter two conclusions obsolete. On the contrary, chapter five demonstrates the robustness of our simplified model. Furthermore, these extensions will allow us to derive some additional interesting results which cannot be derived otherwise.

6.1 Nondurable Housing Production

In this section we introduce housing production and demonstrate that, under reasonable assumptions, the simplified model in chapter two is robust with respect to this extension. To support our claim we construct a model in which housing as an output is distinguished from land as an input in housing production, and we specify conditions under which the expanded model is equivalent to the basic model of chapter two. The analysis in this section also allows for the derivation of additional results.

6.1.1 A Model with Households and Housing Producers

Let housing be defined as a final good produced by a *composite good input* Z_H and *land* L according to a linear homogeneous production function $f_H[Z_H, L]$. Housing is purchased by the households from competitive housing producers. As in chapter two, prices are given for any location in the city.

Similarly to section 2.2.2, an individual located at distance x from the centre selects a feasible consumption bundle $(Z_C[x], H[x])$ which maximises its utility level $U = u[Z_C, H, \bar{\gamma}]$, where the amount of the composite good consumed is now denoted by Z_C. Formally, the problem can be expressed as

$$V \equiv \max_{Z_C, H} u[Z_C, H, \bar{\gamma}] \text{ subject to } Z_C + PH = \bar{Y} - T, \tag{6.1}$$

where $P[x]$ is the *price per unit of housing* at x. A necessary condition for this problem is given by

$$\left(\frac{\partial u^i}{\partial H}\right) / \left(\frac{\partial u^i}{\partial Z_C}\right) = P, \tag{6.2}$$

where superscript i denotes the solution to problem (6.1).

We now turn to the suppliers of housing. A producer at x selects a combination of inputs $(Z_H[x], L[x])$ which maximises profit:

$$\max_{Z_H, L} (P f_H[Z_H, L] - (Z_H + RL)). \tag{6.3}$$

The necessary conditions for this problem are

$$\begin{aligned} \frac{\partial f_H^s}{\partial Z_H} &= \frac{1}{P} \quad (a) \\ \frac{\partial f_H^s}{\partial L} &= \frac{R}{P} \quad (b) \end{aligned} \tag{6.4}$$

where superscript s denotes a solution to problem (6.3). Conditions (6.2) and (6.4) imply

$$\begin{aligned} \left(\frac{\partial u^i}{\partial H}\right) / \left(\frac{\partial u^i}{\partial Z_C}\right) &= 1/\left(\frac{\partial f_H^s}{\partial Z_H}\right) \quad (a) \\ \left(\frac{\partial f_H^s}{\partial L}\right) / \left(\frac{\partial u^s}{\partial Z_H}\right) &= R. \quad (b) \end{aligned} \tag{6.5}$$

Finally, since housing is produced under constant returns to scale, the size of a housing production unit remains indeterminate. If we allow a single producer per ring, and taking into account the land constraint (2.9), we have $L^s = \theta$ at every distance within the city boundary. This implies

$$nH^i = f_H\left[Z_H^s, \theta\right]. \tag{6.6}$$

6.1.2 A Model with Housing Produced by Households

Let us now consider a model in which housing for each household is produced by the household itself, which purchases land in the market and uses some of the composite good to produce housing. The total amount of the composite good available to the household is partitioned according to

$$Z = Z_C + Z_H, \tag{6.7}$$

and housing is produced with inputs (Z_H, L) as

$$H = f_H\left[Z_H, L\right]. \tag{6.8}$$

We can express the choice problem of a household at x as follows. The household selects a feasible bundle (Z, Z_C, L) which maximises its utility level determined by $U = u\left[Z_C, f_H\left[Z - Z_C, L\right], \bar{\gamma}\right]$. We can express this problem as

$$V \equiv \max_{Z, Z_C, L} u\left[Z_C, f_H\left[Z - Z_C, L\right], \bar{\gamma}\right] \text{ subject to } Z + RL = \bar{Y} - T. \tag{6.9}$$

The necessary conditions are

$$\begin{aligned} \left(\frac{\partial u^h}{\partial H}\right) / \left(\frac{\partial u^h}{\partial Z_C}\right) &= 1/\left(\frac{\partial f_H^h}{\partial Z_H}\right) \quad (a) \\ \left(\frac{\partial f_H^h}{\partial L}\right) / \left(\frac{\partial f_H^h}{\partial Z_H}\right) &= R, \quad (b) \end{aligned} \tag{6.10}$$

where superscript h denotes a solution to problem (6.9). These have the same structure as the necessary conditions (6.5) of section 6.1.1.

We shall argue that, under constant returns to scale for housing production, the two models not only have the same necessary conditions, but also the same solution. For, in this case, (6.5(b)) and (6.10(b)) uniquely determine the same factor proportion in $f_H[\cdot]$. Now taking into account the land constraint (2.9), we have $nL^h = \theta$ and, therefore, $L^h = \theta/n$. Since the two models have the same factor proportion in the production of housing, it must also be true that $Z_H^h = Z_H^s/n$. In consequence, multiplying both sides of (6.8) by n and taking into account linear homogeneity we get

$$nH^h = f_H\left[Z_H^s, \theta\right]. \tag{6.11}$$

Upon comparison with (6.6), we conclude that $H^h = H^i$.

It remains to show that the consumption of the composite good is also the same in the two models, *i.e.* $Z_C^h = Z_C^i$. This follows from the budget constraints in (6.1) and (6.9), after taking into account the zero profit condition for the housing producers in section 6.1.1.[1] Thus, under constant returns to scale, the single housing producer who produces housing on a ring can be replaced by all the households on the ring which act independently as their own housing producers.

A household in this section can be thought of as making decisions in two stages. During the first stage, given Z and L, the household maximises utility by allocating the total amount of the composite good available between consumption and housing. The maximised utility becomes a function of Z and L as

$$\tilde{u}[Z, L, \bar{\gamma}] = \max_{Z_C} \; u[Z_C, f_H[Z - Z_C, L], \bar{\gamma}]. \tag{6.12}$$

In the second stage, the household selects a feasible combination (Z^h, L^h) that maximises its utility level $U = \tilde{u}[Z, L, \bar{\gamma}]$. Thus, in the second stage, the problem of a household is given by

$$V \equiv \max_{Z,L} \tilde{u}[Z, L, \bar{\gamma}] \text{ subject to } Z + RL = \bar{Y} - T. \tag{6.13}$$

This is precisely the standard choice problem of section 2.2.2 in which L is substituted by H to represent housing. Since $f_H[\cdot]$ is concave, $\tilde{u}[\cdot]$ is quasi–concave as required by assumption 2.2. Hence the problem with housing production under linear homogeneity is reduced to the basic model of chapter two.

Result 6.1: If housing production exhibits constant returns to scale, the basic model of chapter two is a reduced form of a model with housing production. Therefore all results derived in chapters two and three also apply to a competitive equilibrium with housing production under constant returns to scale.

The equivalence described by result 6.1 does not imply that extending the model to incorporate housing production under constant returns is redundant. On the contrary, including production in our model allows for the derivation of

---

[1] Starting with the budget constraint in (6.9), we have

$$\begin{aligned} Z^h &= \bar{Y} - T - RL^h \\ &= \bar{Y} - T - R\frac{L^s}{n} \\ &= \bar{Y} - T - PH^i + Z_H^h \qquad \text{(by zero profit and (6.6))} \\ &= Z_C^i + Z_H^h \qquad \text{(by (6.1))} \end{aligned}$$

which, upon comparison with (6.7), establishes that $Z_C^h = Z_C^i$ as required.

additional results. For example, using result 2.2, equilibrium land rent declines with distance when housing production is under constant returns. This result, together with (6.5) and linear homogeneity, implies that the proportion of the composite good relative to land in the production of housing, Z_H^e/θ, declines with distance from the centre. Declining Z_H^e/θ, in turn, implies that both the average product of land, $f_H\left[Z_H^e,\theta\right]/\theta$, as well as the price per unit of housing, also decrease with distance.[2] The same must be true regarding population density

$$D^e = \frac{n^e}{\theta} \underset{(6.6)}{=} \frac{f_H\left[Z_H^e,\theta\right]/\theta}{H^e} \tag{6.14}$$

because the average product of land declines while housing consumption increases by (2.17). Finally, the value of housing production per unit of land, $P^e f_H\left[Z_H^e,\theta\right]/\theta$ must also fall because the average product of land and the price of housing decrease with distance.[3]

In summary,

Result 6.2: If housing production exhibits constant returns to scale then, for $x \in \mathcal{X}$,

(i) $\frac{d}{dx}\left(\frac{Z_H^e}{L^e}\right) < 0.$

(ii) $\frac{d}{dx}\left(\frac{f_H\left[Z_H^e,\theta\right]}{L^e}\right) < 0.$

(iii) $\frac{d}{dx}\left(P^e f_H\left[Z_H^e,\theta\right]\right) < 0.$

(iv) $\frac{dD^e}{dx} < 0.$

[2] Alternatively, one could begin with the extended model and derive from maximising utility with respect to distance that P^e declines with distance. Using the first condition in (6.4), the marginal product of the composite good must increase with distance. Under linear homogeneity, this can be true only if Z_H^e/θ decreases with distance. But then the marginal product of land decreases with distance, and so does $\left(\partial f_H^e/\partial L\right)P^e$ which, by the second condition in (6.4), equals land rent. It follows that land rent declines with distance. Furthermore, since Z_H^e/θ decreases with distance, the average product of land must also decrease with distance in equilibrium.

[3] Muth (1969, p. 51) investigates thoroughly the configuration of this decline and relates it to both the elasticity of substitution between L^e and Z_H^e, and to its rate of change with distance. For example, if the elasticity of substitution is constant and less than unitary then the value of housing production per unit of land declines at a decreasing rate with distance from the centre.

6.2 Durable Housing Production

Our analysis so far assumed that buildings are perfectly malleable thus disregarding a fundamental characteristic of housing structure—its durability. In order to account for durability, we must not only distinguish between housing structure and the land it occupies but also between the structure itself and the housing services rendered by the structure. Furthermore, we have to account for the evolution of the housing stock through time as it moves from construction, through maintenance, to demolition and replacement. This process, *filtering*, is essential if one wants to examine urban housing issues that go beyond the purely spatial aspects of the housing market previously examined.

Early models of durable housing, such as Smith (1972) and Muth (1973), have restricted filtering to a given rate of quality deterioration through time—thus suppressing the impact of maintenance on the rate of filtering and, therefore, on the supply of housing. A time–invariant amount of maintenance expenditure was introduced by Muth (1976) and Brueckner (1981), which affects the rate of depreciation but cannot reverse it. The same is true in Sweeney (1974*a,b*), and Ohls (1975), who generalised to allow for maintenance expenditure fluctuations determined in response to market price fluctuations: in all these models, *downward* filtering is inevitable. By relaxing this restriction, Henderson (1977) was led to conclude that housing quality in a city at equilibrium must remain constant through time, and that the only way downward filtering can occur is through changes in demand or technology. However, as Arnott, Davidson and Pines (1983) (hereafter ADP) have shown, these conclusions hinge upon a very specific assumption concerning the technology of maintenance.. One then is bound to ask about what can happen in a more general framework. "Will [the developer] keep [his] building at a constant quality as Henderson argues? Will he continually downgrade as Sweeney and Ohls assume? Or are both of these behaviours, and perhaps others which include upgrading and rehabilitation possible?" (ADP (1983).) These questions have been examined in detail by ADP (1983, 1986) who provide the basis for our discussion.

6.2.1 Supply of Housing

Conceptual Framework

Urban housing is built by the same development corporations which were described in section 2.2.1 of chapter two. Therefore development corporations here behave both as land owners and as builders, who construct and maintain their housing stock while renting it to the individuals in the city. We consider a particular development corporation at distance x from the centre. The development corporation (hereafter developer) builds on its site a particular quantity of housing which remains fixed over the entire life of the building. An important attribute of the residential structure is quality, which can be understood as a composite index of various building characteristics at the time of construction and which is affected by maintenance expenditures over the life of the building.

Higher quality implies higher rental price and higher construction cost per unit of housing. In particular, the developer at x faces a *price per unit of housing* $\hat{P}[q[\bar{\tau}], x] \equiv P[q[\bar{\tau}]]$, where $q[\bar{\tau}]$ represents the *quality of housing at time* $\bar{\tau}$ *after construction*, and a *construction cost per unit area* $K[q\langle 0\rangle, \rho]$, where ρ is the *structural density of housing units per unit of land*. Since, for a particular level of quality, developers at different distances from the centre face different prices per unit of housing, they behave differently. This implies that housing characteristics change over distance from the centre, and so does construction cost and housing price.

We next specify those three variables in more detail. We assume that there is no demand for housing below a minimum standard:

Assumption 6.1: The housing price function $P[q[\bar{\tau}]]$ is differentiable and strictly increasing at a decreasing rate for $q[\bar{\tau}] \in (q^{min}, q^{max})$, where $q^{min} \geq 0$ and $q^{max} > q^{min}$ are given quality levels, and where q^{max} is not necessarily finite. For $q[\bar{\tau}] \leq q^{min}$, we have $P[q[\bar{\tau}]] = 0$.

For construction cost we assume that producing zero–quality dwellings, as well as remaining idle, is costless. Construction cost increases at an increasing rate as the quality of construction or structural density become higher. In addition, as structural density becomes higher, the marginal cost of quality increases relatively faster than structural density does:

Assumption 6.2: The construction cost function $K[q\langle 0\rangle, \rho]$ is differentiable and such that $K[0, \rho] = 0$ and $K[q\langle 0\rangle, 0] = 0$. Furthermore

$$\text{(i)} \qquad \frac{\partial K}{\partial q} > 0 \text{ and } \frac{\partial^2 K}{\partial q^2} > 0$$

$$\text{(ii)} \qquad \frac{\partial K}{\partial \rho} > 0 \text{ and } \frac{\partial^2 K}{\partial \rho^2} > 0$$

$$\text{(iii)} \qquad \frac{\partial^2 K}{\partial q \partial \rho} > \frac{1}{\rho}\frac{\partial K}{\partial q}.$$

Both the housing price, which contains all information needed about the demand side of the housing market, and the construction cost are taken by the developer as given functions.

We now turn to housing quality and the way it evolves through time. Housing quality depends on the level of *maintenance expenditure per unit of housing*

$m[\bar{\tau}]$ through a *maintenance technology* which associates filtering to quality and the flow of maintenance expenditure as

$$\dot{q}[\bar{\tau}] = F[q[\bar{\tau}], m[\bar{\tau}]], \tag{6.15}$$

where the dot represents a time derivative. In general, upward filtering cannot occur without maintenance expenditure. Moreover, higher quality implies a stronger need for maintenance. As quality increases, the need for maintenance increases faster. On the other hand, higher maintenance expenditure implies a slower rate of filtering downward (or a faster rate of filtering upward). For higher levels of maintenance, additional expenditure becomes less effective. Finally, the improvement from an additional dollar spent on maintenance decreases for higher quality because maintaining still higher–quality standards becomes increasingly more difficult.

Assumption 6.3: The maintenance technology function $F[q[\bar{\tau}], m[\bar{\tau}]]$ is differentiable, bounded and concave. In addition we have $F[q[\bar{\tau}], 0] \leq 0$ and

(i) $\dfrac{\partial F}{\partial q} < 0$ and $\dfrac{\partial^2 F}{\partial q^2} < 0$

(ii) $\left.\dfrac{\partial F}{\partial m}\right|_{m=0} = \infty$, $\dfrac{\partial F}{\partial m} > 0$ and $\dfrac{\partial^2 F}{\partial m^2} < 0$

(iii) $\dfrac{\partial^2 F}{\partial q \partial m} < 0.$

Given the above housing price and construction cost schedules, and given the maintenance technology, the profit–maximising developer must choose both the quality of housing and the associated level of maintenance over time, as well as the structural density and the *demolition age of housing* A. Formally, the developer must determine the functions $q[\bar{\tau}]$ and $m[\bar{\tau}]$, as well as the level of the parameters $q_0 = q\langle 0\rangle, q_A = q\langle A\rangle, \rho$ and A. Notice that the demolition age can be infinite. If however A is finite, we assume that when the developer demolishes a building, he or she replaces it immediately with an exact copy of the original one. Thus the profit of the developer is calculated over an infinite number of demolition cycles.

In what follows we assume a demolition cycle of finite length A, leaving the case of an infinitely long A for later on. We specify the building decisions of the developer in three steps. In the first step we determine $q^s[\bar{\tau}]$ and $m^s[\bar{\tau}]$ which maximise the present value of net income received per housing unit over a particular demolition cycle, given the housing quality at the beginning and the end of the cycle, as well as the length of the cycle. In particular, we have

$$\mathcal{J}^s[q_0, q_A, A] \equiv \max_{q[\bar{\tau}], m[\bar{\tau}]} \int_0^A \xi[\bar{\tau}] (P[q[\bar{\tau}]] - m[\bar{\tau}]) \, d\bar{\tau}$$

$$\text{subject to} \begin{cases} \text{(i)} & \dot{q}[\bar{\tau}] &= F[q[\bar{\tau}], m[\bar{\tau}]] \\ \text{(ii)} & q\langle 0\rangle &= q_0 \\ \text{(iii)} & q\langle A\rangle &= q_A \end{cases} \tag{6.16}$$

where $P[\cdot]$ and $F[\cdot]$ are given, $\xi[\bar{\tau}]$ is a *discounting factor at time* $\bar{\tau}$, $\xi[\bar{\tau}] \equiv \exp[-\bar{r}\bar{\tau}]$, and $\bar{r}$ is the *interest rate*. In the second step we determine q_0^*, q_A^* and A^* which maximise the total profit per unit of land given $\mathcal{J}^s[q_0, q_A, A]$ from the first step:

$$\Pi[\rho] \equiv \max_{q_0, q_A, A} \frac{1}{1-\xi\langle A\rangle}\left(\rho\mathcal{J}^s[q_0, q_A, A] - K[q_0, \rho]\right), \tag{6.17}$$

where $K[\cdot]$ is given.[4] Finally, in the third step we determine the profit–maximising structural density ρ.

Profit–Maximising Construction and Maintenance Schedules

First Step

In section 6.3.1 of the appendix we show that the necessary conditions of problem (6.16) for housing quality and maintenance expenditure are given by

$$\begin{aligned} q^s[\bar{\tau} \mid q_0, q_A, A]: &\quad \frac{dP^s}{dq} + \dot{\mu}^s &= \mu^s\left(\bar{r} - \frac{\partial F^s}{\partial q}\right) &\quad (a) \\ m^s[\bar{\tau} \mid q_0, q_A, A]: &\quad \mu^s\frac{\partial F^s}{\partial m} &= 1. &\quad (b) \end{aligned} \tag{6.18}$$

where $\mu[\bar{\tau}]$ can be interpreted as the *current marginal shadow price of housing.*

Notice that the solutions of problem (6.16) in the first step, including μ^s, are conditional upon the parameters to be determined in the second step. Moreover, notice that the LHS of the conditions in (6.18) represents marginal benefits and the RHS marginal costs—both applying to any particular $\bar{\tau} \in [0, A]$. For housing quality, at a particular time, the marginal benefit includes the increase in current revenue and the capital gain associated with one more unit of quality, while the marginal cost includes the corresponding additional interest and the marginal depreciation. For maintenance expenditure, the marginal benefit is represented by the value of the decrease in depreciation associated with one more unit of expenditure, while the marginal cost is the unit of expenditure itself.

---

[4]The factor $(1-\xi\langle A\rangle)^{-1}$ on the RHS of (6.17) appears because of the cyclical nature of housing construction. At the beginning of a new cycle the profit component is further discounted by $\xi\langle A\rangle$, so that

$$\Pi \equiv \max(\rho J - K)\left(1 + \xi\langle A\rangle + \xi\langle A\rangle^2 + \cdots\right)$$

which, at the limit where the number of cycles becomes infinite, is given by (6.17).

Using (6.15) and (6.18(b)) we can express $\dot{q}^s$ and m^s as functions of q^s and μ^s. We can also write $\dot{\mu}^s$ as a function of q^s and μ^s by (6.18(a)). Thus the housing–value maximising maintenance schedules of the developer over a demolition cycle (given the housing quality at the beginning and the end of the cycle, as well as the length of the cycle) can be described by the dynamic interactions between housing quality $q^s[\bar{\tau}]$ and its corresponding shadow price $\mu^s[\bar{\tau}]$ alone. Using (6.15) and (6.18(a)), the quality–shadow price dynamics are given by the system

$$\begin{aligned} \dot{q}^s[q^s,\mu^s] &= F[q^s, m[q^s,\mu^s]] && (a) \\ \dot{\mu}^s[q^s,\mu^s] &= \mu^s\left(\bar{r} - \frac{\partial F}{\partial q^s}\right) - \frac{dP}{dq^s}\,. && (b) \end{aligned} \tag{6.19}$$

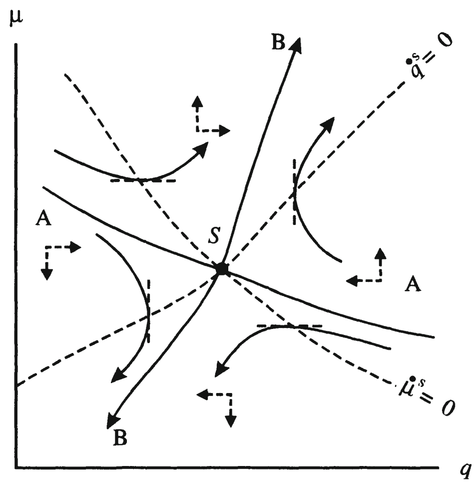

FIGURE 6.1. Value–Maximising Maintainance Schedules.

We first consider the loci $\dot{q}^s = 0$ and $\dot{\mu}^s = 0$ on the q–μ plane. In section 6.3.2 of the appendix, we prove that

$$\left.\frac{d\mu^s}{dq^s}\right|_{\dot{q}^s\,=\,0} > 0 \text{ and } \left.\frac{d\mu^s}{dq^s}\right|_{\dot{\mu}^s\,=\,0} < 0. \tag{6.20}$$

Using this information, the graphs of $\dot{q}^s = 0$ and $\dot{\mu}^s = 0$ are shown in figure 6.1 by the two dashed lines intersecting at point S, which represents the unique stationary solution of the system (6.19).[5] These two lines partition the q–μ plane into four sectors, each one characterised by different general directions of movement obeyed by the solution trajectories of the system (6.19). In order

[5] We assume that the intersection occurs at a point where both quality and shadow price are positive.

to determine the general directions of movement for each sector, we need some information about how $\dot{q}^s$ and $\dot{\mu}^s$ vary over the q–μ plane. This is provided by

$$\frac{\partial \dot{q}^s}{\partial \mu^s} > 0 \text{ and } \frac{\partial \dot{\mu}^s}{\partial q^s} > 0, \tag{6.21}$$

the proof of which appears in section 6.3.3 of the appendix. Using (6.21) we can specify the general movement of the solution trajectories as indicated by the four directional angles corresponding to the four sectors of the q–μ plane in figure 6.1. Since the direction of trajectories changes sign as they cross the graphs of $\dot{q}^s = 0$ and $\dot{\mu}^s = 0$, they are locally vertical on $\dot{q}^s = 0$ and locally horizontal on $\dot{\mu}^s = 0$. Moreover, since trajectories cannot cross each other, the unique stationary solution S can be attained along the single line segment AA specifying the two *stable–arm* trajectories of the solution.[6] For the same reason there are only two trajectories moving away from S along BB, the *unstable–arm* trajectories of the solution. Lines AA and BB (the *separatrices*) determine four new sectors, each containing trajectories with similar characteristics of movement as shown in figure 6.1.

Second Step

In section 6.3.4 of the appendix we show that the necessary first-order conditions for problem (6.16) with respect to $[q_0, q_A, A]$ are given by

$$\begin{array}{lrcl}
q_0^* : & \dfrac{\rho}{1-\xi\langle A^*\rangle}\left(\mu^*\langle 0\rangle - \dfrac{1}{\rho}\dfrac{\partial K^*}{\partial q_0}\right) & = & 0 \quad (a) \\[2ex]
q_A^* : & \dfrac{\xi\langle A^*\rangle\rho}{1-\xi\langle A^*\rangle}\mu^*\langle A^*\rangle & = & 0 \quad (b) \\[2ex]
A^* : & \left(\begin{array}{l} -\dfrac{\xi\langle A^*\rangle\bar{r}\rho}{(1-\xi\langle A^*\rangle)^2}\left(\mathcal{J}^* - \dfrac{K[q_0^*,\rho]}{\rho}\right) \\[2ex] +\dfrac{\xi\langle A^*\rangle\rho}{1-\xi\langle A^*\rangle}\left(P[q_A^*] - m^*\langle A^*\rangle\right)\end{array}\right) & = & 0 \quad (c)
\end{array} \tag{6.22}$$

where we have $m^*[\bar{\tau}] \equiv m^s[\bar{\tau} \mid q_0^*, q_A^*, A^*]$, $\mu^*[\bar{\tau}] \equiv \mu^s[\bar{\tau} \mid q_0^*, q_A^*, A^*]$ and $\mathcal{J}^* \equiv \mathcal{J}^s[q_0^*, q_A^*, A^*]$, with asterisk denoting an optimal value given ρ.

Condition (6.22(a)) says that the marginal cost of the optimal initial quality of a housing unit is equal to the marginal revenue of the optimal quality of a housing unit. Condition (6.22(b)) implies that the marginal value of terminal

[6] Since we have

$$\frac{d\mu^s}{dq^s} = \frac{\dot{\mu}^s[q^s, \mu^s]}{\dot{q}^s[q^s, \mu^s]},$$

each point on the q–μ plane is associated with only one slope dm^s/dq^s. Crossing trajectories require two different slopes at the same point.

quality vanishes at the optimum. We now turn to (6.22(c)). For its interpretation, it is useful to introduce the *current–valued Hamiltonian* function

$$\mathcal{H}[\bar{\tau}] \equiv P[q[\bar{\tau}]] - m[\bar{\tau}] + \mu[\bar{\tau}] F[q[\bar{\tau}], m[\bar{\tau}]] \qquad (6.23)$$

which is defined in terms of the variables determined at the first step.[7] Since the maximised variables at the first step are conditional upon the parameters determined at the second step, we can write the maximised Hamiltonian as $\mathcal{H}^s[\bar{\tau} \mid q_0, q_A, A]$. In section 6.3.5 of the appendix we show that

$$\mathcal{J}^s[q_0, q_A, A] = \frac{1}{\bar{r}}(\mathcal{H}^s\langle 0\rangle - \xi\langle A\rangle \mathcal{H}^s\langle A\rangle). \qquad (6.24)$$

If we introduce (6.24) to (6.22(c)) we obtain

$$\frac{\xi\langle A^*\rangle \rho}{(1-\xi\langle A^*\rangle)^2}\left(\mathcal{H}^*\langle A^*\rangle - \mathcal{H}^*\langle 0\rangle + \frac{\bar{r}}{\rho}K^*\right) = 0 \qquad (6.25)$$

where $\mathcal{H}^*[\bar{\tau}] \equiv \mathcal{H}^s[\bar{\tau} \mid q_0^*, q_A^*, A^*]$. By definition, the current–valued Hamiltonian represents the temporal revenue per unit of housing minus the corresponding direct cost. The latter includes both maintenance expenditure and depreciation.[8] Taking this into account, we can interpret (6.22(c)) using (6.25) as follows. When the developer decides on the profit–maximising length of the demolition cycle, he or she compares the possibility of keeping the old structure for the next time period against the possibility of replacing it with a new structure in the current time period. If the developer decides to adopt the first alternative, he or she receives the net revenue per unit of the existing structure ($\mathcal{H}\langle A\rangle$) in the current period and, at the same time, saves the interest to be paid in the next period on capital required to build a new housing unit ($\bar{r}K/\rho$) in the current period. If, on the other hand, the developer decides to adopt the second alternative, he or she receives the net revenue per unit of the new structure ($\mathcal{H}\langle 0\rangle$). The profit–maximising length of the demolition cycle is then determined by the period in which the developer becomes indifferent between the two alternatives.

One would expect that, at the time demolition, there is no maintenance expenditure. This additional implication for the end of the demolition cycle can be obtained by (6.22(b)). Introducing it in (6.18(b)), we must necessarily have $\partial F[q_A^*, m\langle A^*\rangle]/\partial m = \infty$ which, by assumption 6.3(ii), implies

$$m\langle A^*\rangle = 0. \qquad (6.26)$$

[7] Upon substitution of (6.23) in the Lagrangean of problem (6.16) (see (6.44) in the appendix) we can define the current valued Hamiltonian as a component of the Lagrangean function, namely,

$$\mathcal{L} = \int_0^A \xi \mathcal{H} d\bar{\tau} - \int_0^A \xi \mu \dot{q} d\bar{\tau}.$$

See, for example, Intrilligator (1971).

[8] Observe that $F[\cdot]$ is the physical depreciation in terms of quality, while μ is the marginal value of quality. Thus $\mu F[\cdot]$ represents the temporal rate of the asset–value decrease, that is, the depreciation cost per unit of housing.

We now turn to a diagrammatic representation of trajectories on the q–μ plane that satisfy the conditions of the first two steps for a given level of structural density. We know by (6.22(b)) that the marginal shadow price of housing quality must be zero at the end of a demolition cycle. Using this fact we can separate the solution trajectories of figure 6.1 into *finite–duration* trajectories (corresponding to a demolition cycle of finite length A^*) and *infinite–duration* trajectories (corresponding to an infinite A^*). Since finite–duration trajectories must reach the q^s–axis by (6.22(b)), they are found below the separatrix AA in figure 6.1. All remaining trajectories are of infinite duration—including the two stable–arm trajectories on AA which lead to the unique stationary state S. Problem (6.17), hence all conditions in (6.22), pertain to finite–duration trajectories. Infinite–duration trajectories must satisfy (6.22(a)) about initial quality of housing, as well as $\lim_{A\to\infty} \xi \langle A\rangle \mathcal{H} \langle A\rangle = 0$.[9] Thus (6.22($a$)) is satisfied by both finite–duration and infinite–duration trajectories which are consistent with profit maximisation (hereafter, optimal trajectories).

Using (6.22(a)) we know that, given a particular structural density level, the shadow price of quality at construction time $\mu^* \langle 0\rangle$ must equal $(\partial K/\partial q_0^*)/\rho$ which is a known function of quality. We can therefore define the *initial locus*

$$\mu^{\circ} [q \mid \rho] = \frac{1}{\rho} \frac{\partial K}{\partial q_0}\bigg|_{q_0 = q_0^*} \tag{6.27}$$

on the q–μ plane, knowing that all optimal trajectories must begin on this locus.

> **Assumption 6.4:** The initial locus lies everywhere below the locus $\dot{q}^s = 0$ on the q–μ plane.

Since $\partial^2 K/\partial q^2 > 0$ by assumption 6.2(i), the initial locus is increasing with q. Also since K is strictly increasing on quality by the same assumption, $\mu^{\circ} \langle 0\rangle > 0$.

An initial locus consistent with the above requirements is indicated in figure 6.2. We consider first infinite–duration optimal trajectories. Since they must begin on the initial locus, only one of the two stable–arm trajectories can be optimal, namely, that which begins at point A on the initial locus and converges toward the stationary state S. All other infinite–duration optimal trajectories must begin on the initial locus east of A and diverge. The question then arises about which one of these optimal trajectories is the best for the developer, in other words, which one yields the highest total profit per unit of housing.

> **Result 6.3** (Arnott, Davidson and Pines (1983)): The best feasible infinite–duration trajectory that starts at a given quality is that which follows one of the two stable arms to the stationary state.[10]

---

[9] The second requirement is a transversality condition applying to infinitely long optimal trajectories. For some further details and references to technical literature see ADP (1983).

[10] For a proof see ADP (1983).

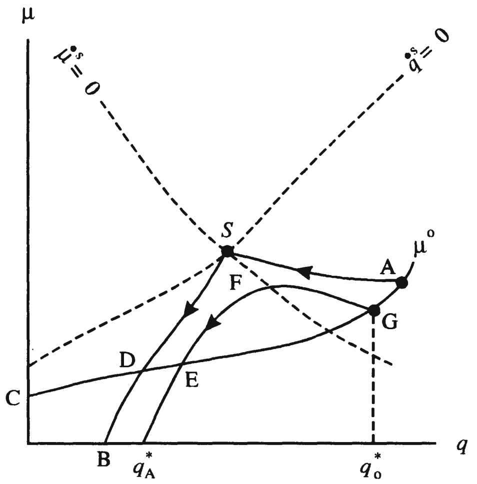

FIGURE 6.2. Profit–Maximising Construction and Maintainance Schedules.

This result, together with assumption 6.4, allows us to concentrate on a single infinite–duration optimal trajectory AS which prescribes a steadily declining quality until the building attains its stationary state, beyond which it is maintained by the developer at the same quality level for all time.

We now turn to finite–duration optimal trajectories, which must satisfy all three conditions in (6.22). By (6.22(*a*)), we know that these trajectories must begin at the initial locus as their infinite–duration counterpart—but at a lower initial quality. By (6.22(*b*)), we know that they must follow a demolition path which ends on the q–axis of figure 6.2 at a terminal quality level that satisfies $\mu^* \langle A^* \rangle = 0$. Once again these two conditions admit an infinite number of trajectories. However, using (6.25), we can determine at most one finite–duration trajectory that satisfies all three conditions. Toward this end, notice that since we can express m^s as a function of q^s and μ^s, we can also write the maximised Hamiltonian as

$$\mathcal{H}[q^s, \mu^s] \equiv P[q^s] - m[q^s, \mu^s] + \mu^s F[q^s, m[q^s, \mu^s]]. \tag{6.28}$$

Taking the total differential of (6.28) yields

$$\begin{aligned} d\mathcal{H}^s &= \frac{\partial \mathcal{H}}{\partial q^s} dq^s + \frac{\partial \mathcal{H}}{\partial \mu^s} d\mu^s \\ &= \left(\frac{dP}{dq^s} + \left(\mu^s \frac{\partial F}{\partial m^s} - 1 \right) \frac{\partial m}{\partial q^s} + \mu^s \frac{\partial F}{\partial q^s} \right) dq^s \\ &+ \left(F^s + \left(\mu^s \frac{\partial F}{\partial m^s} - 1 \right) \frac{\partial m}{\partial q^s} \right) d\mu^s \end{aligned}$$

$$\underset{(6.18)}{=} \quad (\bar{r}\mu^s - \dot{\mu}^s)\,dq^s + F^s d\mu^s$$

$$= \quad \bar{r}\mu^s dq^s - (\dot{q}^s - F^s)\,d\mu^s \tag{6.29}$$

$$\underset{(6.15)}{=} \quad \bar{r}\mu^s dq^s.$$

If we integrate both sides of (6.29) over $[0, A]$ we get

$$\mathcal{H}^s \langle A \rangle - \mathcal{H}^s \langle 0 \rangle = \bar{r} \int_{q_0}^{q_A} \mu^s dq. \tag{6.30}$$

Observe that

$$\begin{aligned} \mathcal{H}^* \langle A^* \rangle - \mathcal{H}^* \langle 0 \rangle \quad &= \quad -\frac{\bar{r}}{\rho} K^* \\ &\underset{\text{(Assumption 6.2)}}{=} \quad -\int_0^{q_0^*} \frac{\bar{r}}{\rho} \frac{\partial K}{\partial q_0} dq \end{aligned} \tag{6.31}$$

holds if and only if (6.25) holds. It follows that

$$\int_0^{q_0^*} \frac{1}{\rho} \frac{\partial K}{\partial q_0} dq = \int_{q_A^*}^{q_0^*} \mu^* dq \tag{6.32}$$

is satisfied if and only if (6.22(c)) is satisfied.

We can now apply this condition to figure 6.2. On the one hand the LHS of (6.32) is represented by the area under the initial locus indicated by 0CGq_0^* in figure 6.2. On the other hand the RHS of (6.32) is represented by the area under the demolition–cycle trajectory q_A^*FGq_0^*. Equality between the LHS and the RHS of (6.32) implies equality between areas 0CEq_A^* and EFG. We conclude that, for a given level of structural density, there is at most a single finite–duration optimal trajectory GFq_A^* that satisfies this *equal areas* condition.[11] Such an optimal trajectory, if it exists, defines a demolition cycle of finite length A^* at the end of which the building is demolished and a new building is immediately constructed in its place with the same structural density ρ^* and the same initial quality q_0^*. In consequence, a new demolition cycle identical to the previous one begins once again at G.

The next question is under what conditions a demolition cycle can exist. What determines whether or not the equal areas condition is satisfied? Since trajectories do not cross, we know that the demolition–cycle trajectory that begins at G must be contained within the sector bordered by the two separatrices AS and SB. This implies that the area represented by the LHS of (6.32) cannot be smaller than 0CDB in figure 6.2, while the area represented by the RHS

[11] Notice that our assumptions on the initial locus, together with the directional constraints of figure 6.1, ensure that the demolition path which begins at G crosses the initial locus only once again at E.

of (6.32) cannot be larger than ASD in the same figure. Therefore a profit–maximising demolition cycle exists if and only if area 0CDB is smaller than area ASD. It follows that the existence of a profit–maximising demolition cycle hinges upon the level of the initial locus $\mu^o[q \mid \rho]$ in figure 6.2. By assumption 6.2(iii), we know that this level rises with increasing structural density:

$$\frac{\partial \mu^o}{\partial \rho} = \frac{1}{\rho}\left(\frac{\partial^2 K}{\partial q_0 \partial \rho} - \frac{1}{\rho}\frac{\partial K}{\partial q_0}\right)\Bigg|_{q_0=q_0^*} > 0. \tag{6.33}$$

If the level of the initial locus varies sufficiently over the feasible range of structural density, this last condition implies that an optimal demolition cycle exists if structural density is low enough. If, on the other hand, structural density is high enough then demolition does not occur and the only optimal trajectory is AS along the stable–arm path of figure 6.2. This implication seems intuitively plausible since, for sufficiently large initial investment per unit of land, one would expect that demolition becomes too costly. If, however, structural density is low enough then there are two optimal trajectories—one that leads to a stationary state, the other to a demolition cycle. Under these circumstances we need to know which one of the two alternatives is preferred by the developer.

Result 6.4 (Arnott, Davidson and Pines (1983)): If an optimal trajectory of finite duration does exist, it yields a higher total profit per unit of housing than the associated infinite–duration optimal trajectory.[12]

We conclude that if structural density is low enough to allow for the possibility of demolition, the demolition–cycle alternative will be preferred by the developer over a construction and maintenance programme that leads to the stationary state.

Third Step

Using the profit definition (6.17), we can determine the optimal level of structural density ρ^* through

$$\begin{aligned} \lim_{A\to\infty} \mathcal{J}^s\left[q_0^*, q^{(S)}, A\right] &= \frac{\partial K^*}{\partial \rho} \quad (a) \\ \mathcal{J}^s\left[q_0^*, q_A^*, A^*\right] &= \frac{\partial K^*}{\partial \rho} \quad (b) \end{aligned} \tag{6.34}$$

for infinite–duration and finite–duration trajectories respectively, where $q^{(S)}$ denotes housing quality at stationary state S in figure 6.2. Sufficiency requires

[12] For a proof see ADP (1983).

that construction cost increases at an increasing rate with increasing structural density. Since this requirement is satisfied by assumption 6.2(ii), we conclude that ρ^* indeed maximises total profit per unit of land.

Once we determine the optimal density, the corresponding total profit per unit of land is readily obtainable. The solution of (6.17) for a given density ρ yields $m[\rho]$ for $\overline{\tau} = 0$ and $q_0[\rho]$. Therefore, using (6.18), it also yields $\mu[\rho] = 1/(\partial F/\partial m)$ for $\overline{\tau} = 0$. Substitution of (6.25) and (6.24) into (6.17) yields the optimal profit conditional on ρ as

$$\Pi[\rho] = \rho\tilde{\Pi}[\rho] - \tilde{K}[\rho], \qquad (6.35)$$

where

$$\begin{aligned} \tilde{\Pi}[\rho] &\equiv \frac{\mathcal{H}\langle 0\rangle}{\overline{r}} \\ &= \frac{1}{\overline{r}}\left(P[q_0[\rho]] - m[\rho] + \mu[\rho]F[m[\rho], q_0[\rho]]\right) \quad (a) \\ \tilde{K}[\rho] &\equiv K[q_0[\rho], \rho]. \quad (b) \end{aligned} \qquad (6.36)$$

Correspondingly, the optimal profit is

$$\Pi[\rho^*] = \rho^*\tilde{\Pi}[\rho^*] - \tilde{K}[\rho^*]. \qquad (6.37)$$

The Spatial Characteristics of Structural Density

In this section we use (6.35) and (6.36) to investigate the effect of housing price on structural density. We confine our comparative statics arguments to changes in demand for housing that affect the level, but not the slope, of the price function. We therefore decompose the price into $P[q] = \hat{P}[q] + \varsigma$, where ς is a shift parameter. Notice that this decomposition has no effect on the dynamics of (6.19), which implies that all trajectories on the q–μ plane, including the stable–arm trajectories, are not altered by a change in ς. The same is true for the initial locus in (6.27). Consequently, for given structural density, profit–maximising construction and maintenance schedules (either of the stable–arm or of the demolition–cycle variety) are not affected by a change in the shift parameter.

Using (6.35) and (6.36) we have

$$\begin{aligned} \Pi[\rho,\varsigma] &= \frac{\rho}{\overline{r}}\left(\hat{P}[q_0[\rho]] + \varsigma - m[\rho] + \mu[\rho]F[m[\rho], q_0[\rho]]\right) \\ &- K[q_0[\rho], \rho]. \end{aligned} \qquad (6.38)$$

and

$$\frac{\partial^2\Pi}{\partial\rho\partial\varsigma} = \frac{1}{\overline{r}} > 0. \qquad (6.39)$$

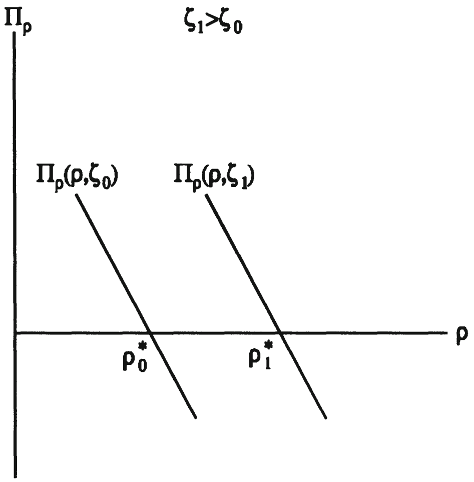

FIGURE 6.3. The Impact of Housing Price on Marginal Profit.

We know from (6.39) that the marginal profit as a function of structural density increases with the price schedule (reflected by an increase of the shift parameter ς). Moreover third–step profit maximisation, when evaluated at the optimal structural density ρ^*, requires $\partial^2\Pi/\partial\rho^2 < 0$. Thus the effect of the increase in the shift parameter ς is represented in figure 6.3 by the rightward shift of $\partial\Pi/\partial\rho$, which increases the optimal structural density from ρ_0^* to ρ_1^*. It follows that $d\rho^*/d\varsigma > 0$.

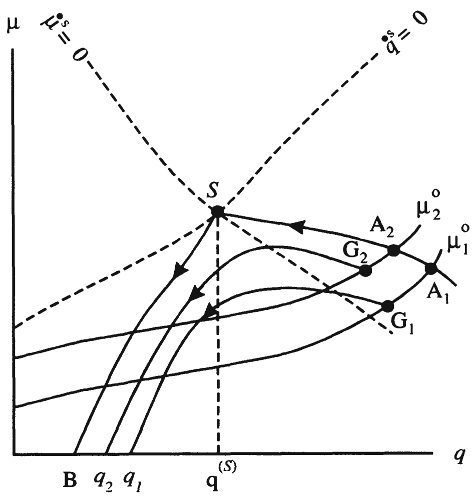

FIGURE 6.4. Effects of an Increased Housing Price.

We now use figure 6.4 to examine the effect of an increased structural density on optimal stable–arm and demolition–cycle trajectories. Since $d\rho^*/d\varsigma > 0$, this is equivalent to examining the effect of an increased housing price on the optimal maintenance and demolition policy. We know from (6.33) that the initial locus shifts from μ_1^o upwards to μ_2^o as the shift parameter, hence as the structural density, increases. Consider first the impact of this change on the stable–arm construction and maintenance schedule. Since trajectories are not affected by that change, both the stable–arm path and the stationary state remain as before. However, since the origin of the stable–arm schedule moves from A_1 to A_2 as housing price increases, the quality of construction in that case becomes lower and urban land is used more intensively than before. Since everything occurs along the same path A_1S, we also infer that higher price means a shorter time of attaining the stationary state at S.

Consider next the impact of an increased housing price on the demolition–cycle construction and maintenance schedule. Contrary to the previous case, we do not know the way construction quality changes as the origin of the profit–maximising demolition cycle moves from G_1 to G_2 in response to the housing price increase—only that structural density must increase. However, using the equal areas condition (6.32), we know that the demolition quality declines from q_1 to q_2 as the housing price increases. The intuition behind this result is straightforward: with higher structural density resulting from higher housing price the alternative cost of demolition increases. Therefore housing quality is left to deteriorate further before it becomes profitable to demolish the structure.

We also know that the lower bound for demolition quality is given by the endpoint of the unstable–arm path SB at B—provided that demolition quality at B is no lower than q^{min}: as we have already mentioned, a demolition cycle schedule ending at a demolition quality lower than q^{min} cannot be supported by the developer because it entails a loss.

Suppose that there is an initial locus level, say $\tilde{\mu}^o$, such that the equal areas condition holds for any initial locus level below $\tilde{\mu}^o$ and it does not hold for any initial locus level at or above $\tilde{\mu}^o$. According to (6.33) this implies that there is a structural density level $\tilde{\rho}$ such that the equal areas condition holds if and only if $\rho^* < \tilde{\rho}$. Let $\tilde{\varsigma}$ satisfy the requirements that $\tilde{P} \equiv \hat{P} + \tilde{\varsigma}$ is a feasible housing price level and that the profit–maximising structural density at that price equals $\tilde{\rho}$. Then it follows from $d\rho^*/d\varsigma > 0$ and result 6.2 that the developer adopts a profit–maximising demolition cycle schedule if and only if housing price is smaller than $\tilde{P}$, and a profit–maximising schedule leading to the stationary state if and only if housing price is equal to or greater than $\tilde{P}$. Finally, let ς_1 denote the initial value of the shift parameter and $\varsigma_2 > \varsigma_1$ denote its value after the change. Then if $P_1 \equiv \hat{P} + \varsigma_1 \geq \tilde{P}$ we have the stable–arm comparative statics described above, and if $P_2 \equiv \hat{P} + \varsigma_2 < \tilde{P}$ we have the corresponding demolition–cycle comparative statics. If, however, $P_1 < \tilde{P}$ and $P_2 \geq \tilde{P}$ then the profit–maximising construction and maintenance schedule changes from a demolition cycle to the stable arm as the housing price increases. Under these circumstances the developer prefers a higher structural density after the increase and adopts a maintenance schedule that eventually brings the dwelling to its stationary state. According to figure 6.2, the quality at stationary state is higher than the quality at the end of the previous demolition cycle. In summary,

Result 6.5 (Arnott, Davidson and Pines (1983)): (i) If the housing price before the increase is no smaller than a given *critical value* $\tilde{P}$ then the profit–maximising schedule is along the same stable–arm trajectory (which leads to the same stationary state) both before and after the increase in the housing price. As housing price increases, the quality of construction decreases, structural density increases and the stationary state is attained more rapidly.

(ii) If the housing price is smaller than its critical value before the increase and no smaller than its critical value after the increase then the profit–maximising schedule changes from a demolition–cycle trajectory to the stable–arm trajectory as the housing price increases. After the price increase, the developer builds at a higher structural density and adopts a maintenance schedule that brings the dwelling to a stationary state of higher quality than the quality at the end of the demolition cycle before the price increase.

(iii) If the housing price after the increase is smaller than its critical value then the profit–maximising schedule is along a demolition cycle both before and after the price increase. The structural density increases and the demolition quality declines as the price increases. In any case, the demolition quality cannot be lower than a minimum level at which the demand becomes zero.

6.2.2 Demand for Housing

In this section we specify utility and transportation cost functions which generate a monotone relationship between the shift parameter ς and distance from the city center. To that end we have to modify the basic framework of chapters two and three as follows. (1) Instead of allowing individuals to choose an amount of land (or an amount of housing as in section 6.1.1), here we allow individuals to choose the quality of the dwelling unit they rent. Thus the utility of any individual is given by $U = u[Z,q]$. Since individuals now occupy a single unit of housing, population density and structural density must coincide in equilibrium at any location. (2) Instead of allowing only one type of individual as in chapters two and three, here we allow for income heterogeneity. In equilibrium, we know that individuals with the same income Y must attain the same level of utility $\bar{U}[Y]$ anywhere they locate, where $\bar{U}$ is a known function and $d\bar{U}/dY > 0$.

Consider a particular individual at a particular distance from the centre x. The maximum amount of *housing rent* p that the individual is willing to pay for a dwelling unit of a particular quality q at this location is determined by the constrained optimisation problem

$$\max_{Z,p}\ p \equiv Y - T - Z \text{ subject to } u[Z,q] \geq \bar{U}[Y]. \tag{6.40}$$

The necessary condition for this problem is given by

$$\nu^i = 1/\frac{\partial u}{\partial Z^i} \tag{6.41}$$

where ν is Lagrangean multipliers of the constraint and where superscript i denotes the solution to problem (6.40). This condition and the constraint (which applies with equality) give p^i and Z^i as functions of x, q and Y.

We next examine in more detail the structure of the *bid rent* p^i. Applying the envelope theorem, and taking into account (6.41), we obtain

$$\begin{aligned} \frac{\partial p^i}{\partial x} &= -t & (a) \\ \frac{\partial p^i}{\partial q} &= \frac{\partial u}{\partial q}/\frac{\partial u}{\partial Z^i}. & (b) \end{aligned} \tag{6.42}$$

Since all cross derivatives of the bid rent with respect to distance are zero, we conclude that the bid–rent function must have the form

$$p^i[x, q, Y] = p[q, Y] - t[x] \tag{6.43}$$

which implies that, for dwelling units of the same quality, the bid rent of identical individuals declines with increasing distance from the centre at the same rate as transportation cost does. Using the quasi–concavity of the utility function, we also know from (6.42(b)) that the bid rent increases at a decreasing rate as quality increases.

In order to specify further the relationship between the bid–rent function and its arguments, we impose that housing quality is a normal good, that is, as utility increases while the price of quality is held constant relative to the composite good, quality also increases. This implies that, as we increase the composite good keeping quality constant, the MRS$[q, Z]$ increases:

Assumption 6.5: $\frac{\partial}{\partial Z}\left(\frac{\partial u}{\partial q}/\frac{\partial u}{\partial Z}\right) > 0.$

If we differentiate (6.42(b)) totally with respect to income, while holding location and quality fixed at (x', q'), we have

$$\frac{d}{dY}\left(\frac{\partial p^i\langle x', q'\rangle}{\partial q}\right) = \frac{\partial}{\partial Z^i}\left(\frac{\partial u}{\partial q}/\frac{\partial u}{\partial Z^i}\right)\frac{d\bar{U}}{dY}/\frac{\partial u}{\partial Z^i} \underset{\text{Assumption 6.5}}{>} 0.$$

Applying the above information to the ring at distance x from the centre, the bid–rent functions for three different incomes $Y_1 < Y_2 < Y_3$ are shown in figure 6.5. Since dwelling units are rented to the highest bidder by the profit–maximising developers, we know that individuals of the low–income group Y_1 can rent housing of a quality level below q_1 at x, those of the middle–income group Y_2 between q_1 and q_2 at x, and those of the high–income group Y_3 above q_2 at x.

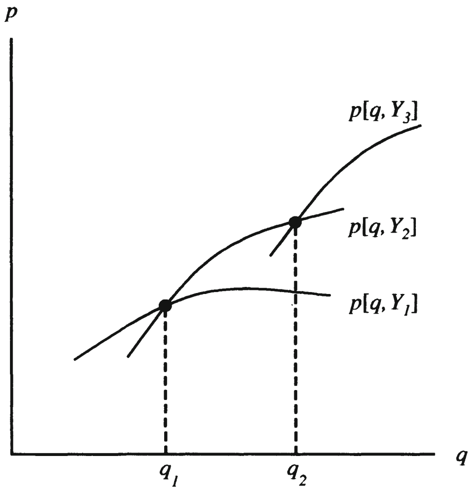

FIGURE 6.5. A Family of Bid Rents at a Given Distance from the Centre.

6.2.3 Spatial Structure of a Durable Housing Market

One interpretation of a demolition cycle at a particular distance from the centre is that the developer builds its entire stock at the beginning of the cycle, so that all those dwelling units age together and all are demolished at the end of the cycle to be replaced by new identical units. With this interpretation, at any particular time, the developer supplies stock of a single quality somewhere between q_0^* and q_A^* in figure 6.2. Recall, however, that each development corporation in our basic model of chapter two, which is represented by the developer of this chapter, develops the land on an entire ring at a particular distance from the centre. Since there is a large number of dwelling units on each ring and since the demolition cycle can be applied to each dwelling unit separately, it is plausible that different units on the ring are associated with distinct demolition cycles of the same type. Under this interpretation, we can imagine that the developer continuously demolishes and rebuilds different units of its stock on different parts of the ring so that, at any particular time, there is a whole range of qualities between q_0^* and q_A^* on the ring.

In what follows, we restrict our discussion to a city in which the rate of demolition and rebuilding is fixed throughout the urban area, so that the age distribution of dwellings on any demolition–cycle ring is uniform. If initial construction began simultaneously in every urban ring, the uniformly distributed range of qualities available on any demolition–cycle ring will eventually be described by $[q_0^*, q_A^*]$ in figure 6.2. Under these circumstances the entire city is in stationary equilibrium, and we can represent the price per unit of housing $P[q]$ introduced in section 6.2.1 as the upper envelope of the bid–rent curves in figure 6.5. Using this observation, together with $d\rho^*/d\varsigma > 0$, we can state

Result 6.6 (Arnott, Davidson and Pines (1986)): Both the demand price for housing quality and the structural density of housing decrease away from the centre.

Recall that the bid–rent function of every group shifts downward parallel and at the same rate $-t$ by (6.43) as distance from the centre increases. Furthermore, that the quality range of a demolition cycle also changes over distance. However, as suggested by figure 6.4, such quality ranges overlap. This implies that dwelling units of the same quality can be found at different distances from the centre, and that they will be occupied by individuals of the same income group irrespectively of where they are found in the city. Given that dwelling units of different quality can also be found at the same distance from the centre, we conclude that durable housing allows for the possibility of income–heterogeneous areas within the city.

We can now use result 6.6, in conjunction with the remarks preceding result 6.5, to associate the critical level of structural density $\tilde{\rho}$ (below which the developer adopts a demolition cycle and above which it adopts a stationary state path) with a corresponding critical value of distance $\tilde{x}$ from the centre. If structural density at the border of the city $\rho\langle\bar{x}\rangle$ is higher than $\tilde{\rho}$ then $\tilde{x} = \bar{x}$ and all developers adopt a stationary state path everywhere within the urban area. By contrast if structural density at the centre of the city $\rho\langle 0\rangle$ is lower than $\tilde{\rho}$ then $\tilde{x} = 0$ and all developers adopt a demolition cycle everywhere within the urban area. Suppose now that $0 < \tilde{x} < \bar{x}$. In that case, the critical value $\tilde{x}$ partitions the city into an *inner city area* and an *outer city area* with significantly different characteristics described by the following

Result 6.7 (Arnott, Davidson and Pines (1986)): If $0 < \tilde{x} < \bar{x}$ then

(i) In the inner city area all buildings follow a stable arm path which leads to the same quality at stationary state.[13] The building quality in the inner city area increases away from the centre.[14]

(ii) In the outer city area all buildings follow a demolition cycle path. The demolition quality in the outer city area increases away from the centre.[15]

Assume that the city in stationary equilibrium has both an inner city area and an outer city area, and that it accommodates an heterogeneous population.

---

[13] All inner–city dwelling units at any distance from the centre attain the same final quality because all trajectories in the q–m plane of figures 6.2 and 6.4, including the stable–arm trajectories, are not affected by a parallel shift of the demand price for housing quality.

[14] This holds because, according to result 6.5(i), construction quality decreases with increasing demand price.

[15] The second part follows from result 6.5(iii).

Consider two locations in the outer city area, x_1 and x_2, and one inner–city location x_3 such that $x_1 > x_2 > \tilde{x} > x_3$. Using figure 6.3, we observe that, at any time, both outer–city rings contain houses of quality both below and above the inner–city quality. This means that the inner city area is occupied by the middle class, while the outer city area accommodates the entire range of incomes. Since demolition quality increases away from the centre, we know that the poorest people locate just outside the inner city area. However, we do not know whether or not the richest people locate just inside the urban fringe.

6.3 Appendices

6.3.1 Proof of (6.18)

If we form the Lagrangean that corresponds to problem (6.16) and integrate it by parts, we obtain

$$\begin{aligned}
\mathcal{L}[q,m,\mu,\lambda_0,\lambda_A] &= \int_0^A \xi\left(P[q] - m - \mu\left(\dot{q} - F[q,m]\right)\right) d\bar{\tau} \\
&- \lambda_0\left(q\langle 0\rangle - q_0\right) - \lambda_A\left(q\langle A\rangle - q_A\right) \\
&= \int_0^A \xi\left(P[q] - m + \mu F[q,m]\right) d\bar{\tau} - \int_0^A \xi\mu\dot{q}\,d\bar{\tau} \\
&- \lambda_0\left(q\langle 0\rangle - q_0\right) - \lambda_A\left(q\langle A\rangle - q_A\right) \\
&= \int_0^A \xi\left(P[q] - m + \mu F[q,m] + \dot{\mu}q - \bar{r}\mu q\right) d\bar{\tau} \\
&+ \mu\langle 0\rangle q\langle 0\rangle - \xi\langle A\rangle \mu\langle A\rangle q\langle A\rangle \\
&- \lambda_0\left(q\langle 0\rangle - q_0\right) - \lambda_A\left(q\langle A\rangle - q_A\right),
\end{aligned} \tag{6.44}$$

where $\mu[\bar{\tau}]$, λ_0 and λ_A are Lagrangean multipliers. To obtain (6.18) differentiate the Lagrangean partially with respect to $[q,m]$ and equate the result to zero.

6.3.2 Proof of (6.20)

(1) For the first part of (6.20), totally differentiate (6.19(a)) evaluated at $\dot{q}^s = 0$ to derive

$$\left.\frac{d\mu^s}{dq^s}\right|_{\dot{q}^s = 0} = -\left(\frac{\partial F}{\partial q^s} + \frac{\partial F}{\partial m^s}\frac{\partial m}{\partial q^s}\right) / \left(\frac{\partial F}{\partial m^s}\frac{\partial m}{\partial \mu^s}\right). \tag{6.45}$$

In order to sign this expression, we need to determine the sign of the partial derivatives of $m\left[q^s, \mu^s\right]$. Toward this end, totally differentiate (6.18(b)) to obtain

$$\mu^s \frac{\partial^2 F}{\partial m^s \partial q^s} dq^s + \mu^s \frac{\partial^2 F}{\partial m^{s2}} dm^s + \frac{\partial F}{\partial m^s} d\mu^s = 0, \tag{6.46}$$

which implies

$$\begin{aligned} \frac{\partial m}{\partial q^s} &= -\frac{\partial^2 F}{\partial m^s \partial q^s} / \frac{\partial^2 F}{\partial m^{s2}} \underset{\text{(Assumption 6.3)}}{<} 0 \quad (a) \\ \frac{\partial m}{\partial \mu^s} &= -\frac{1}{\mu^s} \frac{\partial F}{\partial m^s} / \frac{\partial^2 F}{\partial m^{s2}} \underset{\text{(Assumption 6.3)}}{>} 0. \quad (b) \end{aligned} \tag{6.47}$$

Our claim then follows by using assumption 6.3 and (6.47) on (6.45).

(2) For the second part of (6.20), totally differentiate (6.19(b)) evaluated at $\dot{\mu}^s = 0$ to obtain

$$\left(-\mu^s \left(\frac{\partial^2 F}{\partial q^{s2}} + \frac{d^2 P}{dq^{s2}} \right) dq^s - \mu^s \frac{\partial^2 F}{\partial m^s \partial q^s} dm^s + \left(\bar{r} - \frac{\partial F}{\partial q^s} \right) d\mu^s \right) \Bigg|_{\dot{\mu}^s = 0} = 0. \tag{6.48}$$

Substituting dm^s from (6.46) into (6.48) yields

$$\begin{aligned} \left. \frac{d\mu^s}{dq^s} \right|_{\dot{\mu}^s = 0} &= \left(\mu^s \frac{\partial^2 F}{\partial q^{s2}} + \frac{d^2 P}{dq^{s2}} - \mu^s \left(\frac{\partial^2 F}{\partial m^s \partial q^s} \right)^2 / \frac{\partial^2 F}{\partial m^{s2}} \right) \\ &\div \left(\bar{r} - \frac{\partial F}{\partial q^s} + \frac{\partial F}{\partial m^s} \frac{\partial^2 F}{\partial m^s \partial q^s} / \frac{\partial^2 F}{\partial m^{s2}} \right). \end{aligned} \tag{6.49}$$

which is negative by assumptions 6.1 and 6.3.

6.3.3 *Proof of (6.21)*

(1) For the first part of (6.21), differentiate (6.19(a)) to derive

$$\frac{\partial \dot{q}^s}{\partial \mu^s} = \frac{\partial F}{\partial m^s} \frac{\partial m}{\partial \mu^s} \tag{6.50}$$

because q^s is held constant. Condition (6.46) yields

$$\frac{\partial m}{\partial \mu^s} = -\frac{1}{\mu^s} \frac{\partial F}{\partial m^s} / \frac{\partial^2 F}{\partial m^{s2}} \underset{\text{(Assumption 6.3)}}{>} 0. \tag{6.51}$$

Our claim then follows by using assumption 6.3 and (6.51) on (6.50).

(2) For the second part of (6.21), differentiate (6.19(b)) to derive

$$\frac{\partial \dot{\mu}^s}{\partial q^s} = -\mu^s \left(\frac{\partial^2 F}{\partial q^{s2}} + \frac{\partial^2 F}{\partial m^s \partial q^s} \frac{dm}{dq^s} \right) - \frac{d^2 P}{dq^{s2}} \tag{6.52}$$

because μ^s is held constant. Using once again (6.46), we obtain

$$\frac{\partial m}{\partial q^s} = -\frac{\partial^2 F}{\partial m^s \partial q^s} / \frac{\partial^2 F}{\partial m^{s2}} . \tag{6.53}$$

If we combine the two last equalities, we arrive at

$$\frac{\partial \dot{\mu}^s}{\partial q^s} = -\mu^s \left(\frac{\partial^2 F}{\partial q^{s2}} \frac{\partial^2 F}{\partial m^{s2}} - \left(\frac{\partial^2 F}{\partial m^s \partial q^s} \right)^2 \right) / \frac{\partial^2 F}{\partial m^{s2}} - \frac{d^2 P}{d q^{s2}} \tag{6.54}$$

which is positive by assumption 6.1 and by the concavity of F in assumption 6.3.[16]

6.3.4 Proof of (6.22)

Use the envelope theorem on problem (6.16) (by differentiating the Lagrangean (6.44) with respect to the parameters $[q_0, q_A, A]$) to obtain

$$\begin{aligned} \frac{\partial \mathcal{J}^s}{\partial q_0} &= \mu^s \langle 0 \rangle & (a) \\ \frac{\partial \mathcal{J}^s}{\partial q_A} &= -\xi \langle A \rangle \mu^s \langle A \rangle & (b) \\ \frac{\partial \mathcal{J}^s}{\partial A} &= \xi \langle A \rangle \left(P \langle q_A \rangle - m^s \langle A \rangle \right). & (c) \end{aligned} \tag{6.55}$$

Then differentiate (6.17) with respect to the above parameters and substitute (6.55) into the result.

6.3.5 Proof of (6.24)

We begin with

$$\begin{aligned} \mathcal{J}^s [q_0, q_A, A] &\underset{(6.16)}{=} \int_0^A \xi \left(P[q^s] - m^s \right) d\bar{\tau} \\ &\underset{(6.23)}{=} \int_0^A \xi \mathcal{H}^s d\bar{\tau} - \int_0^A \xi \mu^s F^s d\bar{\tau} \\ &\underset{(6.15)}{=} \int_0^A \xi \mathcal{H}^s d\bar{\tau} - \int_0^A \xi \mu^s \dot{q}^s d\bar{\tau}. \end{aligned} \tag{6.56}$$

[16] Recall that concavity in our case means

$$\frac{\partial^2 F}{\partial m^{s2}} < 0 \text{ and } \frac{\partial^2 F}{\partial q^{s2}} \frac{\partial^2 F}{\partial m^{s2}} - \left(\frac{\partial^2 F}{\partial m^s \partial q^s} \right)^2 > 0.$$

Integrating by parts the first integral on the RHS of (6.56), we obtain

$$\int_0^A \xi \mathcal{H}^s d\bar{\tau} = \frac{1}{\bar{r}} \left(\mathcal{H}^s \langle 0 \rangle - \xi \langle A \rangle \mathcal{H}^s \langle A \rangle + \int_0^A \dot{\xi} \mathcal{H}^s d\bar{\tau} \right). \tag{6.57}$$

Substitution of (6.57) in (6.56) gives

$$\begin{aligned} \mathcal{J}^s [q_0, q_A, A] &= \frac{1}{\bar{r}} \left(\mathcal{H}^s \langle 0 \rangle - \xi \langle A \rangle \mathcal{H}^s \langle A \rangle \right) + \int_0^A \frac{\xi}{\bar{r}} \left(\dot{\mathcal{H}}^s - \bar{r} \mu^s \dot{q}^s \right) d\bar{\tau} \\ &= \frac{1}{\bar{r}} \left(\mathcal{H}^s \langle 0 \rangle - \xi \langle A \rangle \mathcal{H}^s \langle A \rangle \right) \\ &+ \int_0^A \frac{\xi \dot{q}^s}{\bar{r}} \left(\frac{dP}{dq^s} + \dot{\mu}^s \frac{F^s}{\dot{q}^s} - \mu^s \left(\bar{r} - \frac{\partial F}{\partial q^s} \right) \right) d\bar{\tau} \\ &+ \int_0^A \frac{\xi \dot{q}^s}{\bar{r}} \left(\mu^s \frac{\partial F}{\partial m^s} - 1 \right) d\bar{\tau} \\ &= \frac{1}{\bar{r}} \left(\mathcal{H}^s \langle 0 \rangle - \xi \langle A \rangle \mathcal{H}^s \langle A \rangle \right) \end{aligned} \tag{6.58}$$

because the two integrals vanish by $\dot{q} = F[q, m]$ and (6.18).

6.4 References

W. Alonso, 1964, *Location and Land Use.* Harvard University Press, Cambridge, Mass.

R. J. Arnott, R. Davidson and D. Pines, 1983, Housing Quality, Maintenance and Rehabilitation. *Review of Economic Studies* 50, 467–494.

R. J. Arnott, R. Davidson and D. Pines, 1986, Spatial Aspects of Housing Quality, Density and Maintenance. *Journal of Urban Economics* 19, 190–217.

J. K. Brueckner, 1981, A Dynamic Model of Housing Production. *Journal of Urban Economics* 10, 1–14.

J. V. Henderson, 1977, *Economic Theory and the Cities.* Academic Press, New York. Second edition in 1986.

M. D. Intrilligator, 1971, *Mathematical Optimization and Economic Theory.* Prentice–Hall, Englewood Cliffs, N.J.

E. S. Mills, 1967, An Aggregative Model of Resource Allocation in a Metropolitan Area. *American Economic Review* 57, 197–210.

R. F. Muth, 1961, The Spatial Structure of the Housing Market. *Papers and Proceedings of the Regional Science Association* 7, 207–220.

R. F. Muth, 1969, *Cities and Housing.* The University of Chicago Press, Chicago.

R. F. Muth, 1973, A Vintage Model of the Housing Stock. *Papers and Proceedings of the Regional Science Association* 30, 141–146.

R. F. Muth, 1976, A Vintage Model with Housing Production. In Y. Y. Papageorgiou (ed.), *Mathematical Land Use Theory.* Lexington Books, Toronto, 245–259.

J. C. Ohls, 1975, Public Policy Towards Low Income Housing and Filtering in Housing Markets. *Journal of Urban Economics* 2, 144–171.

W. F. Smith, 1972, A Theory of Filtering. In M. Edel and J. Rosenberg (eds.), *Readings in Urban Economics.* Macmillan, New York, 193–204.

J. L. Sweeney, 1974a, Quality, Commodity Hierarchies and Housing Markets. *Econometrica* 42, 147–167.

J. L. Sweeney, 1974b, A Commodity Hierarchy Model of the Rental Housing Market. *Journal of Urban Economics* 1, 288–323.

7

Transportation Infrastructure

Urban infrastructure, which in our model includes transportation, land reclamation and a public good, has been represented in previous chapters by $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$ respectively. These exogenous shift parameters have characterised in the most general terms the quality of urban infrastructure. However, even though such general treatment has provided us with some information about how cities adjust to changes in the overall quality of urban infrastructure (see chapters two and three), it is not good enough for policy analysis. In order to examine the provision of urban infrastructure in sufficient detail we need to abandon $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$. In this chapter we concentrate on the implications of abandoning $\bar{\alpha}$. The more detailed discussion of what $\bar{\alpha}$ represents, which we adopt in this chapter, allows us to elaborate on specific reasons why the transportation infrastructure is publicly provided rather than supplied by competitive markets as, for example, housing. The level of detail we use makes it possible to delineate more precisely the economic reasons for the involvement of urban governments in transportation. We argue that. in contrast to what is common belief in the literature, the interdependencies among the net benefits to those agents who commonly use the transportation infrastructure (often referred to as externalities) need not cause market failure. These interdependencies can be internalised in the market by profit–maximising firms. Only when (a) it is too costly to exclude potential users or (b) there are increasing returns to scale, will competitive firms be unable to survive. In those two cases public intervention becomes indispensable.

7.1 Competitive Market for Urban Transportation

We have seen in previous chapters how does the level of service in transportation affect the equilibrium urban structure. In reality, however, the urban structure also affects the level of service through the demand for and the supply of transportation. On the one hand the demand for transportation is directly affected by the population distribution. On the other, the corresponding supply of transportation is affected by the cost of land used in the production of transportation infrastructure. Thus the level of service in transportation both affects and is affected by the equilibrium urban structure. This fundamental interdependence between urban transportation and equilibrium urban structure was first modelled in a comprehensive way by Strotz (1965), who introduced many of the basic ideas in the economic literature on the relationship between transportation and land use, and to whom we can trace many arguments developed in this chapter.

With few exceptions, the transportation infrastructure in urban areas is publicly supplied. Is is therefore natural to study the issue of transportation in the context of a normative public good theory.[1] However, in this section, we portray the urban transportation system as a standard industry producing an excludable private good. We then discuss why this description is unrealistic and the associated reasons for market failure.

Studies in which individuals purchase transportation as if it were an excludable good do exist in the literature. One of them is Mills (1967), who assumed that every individual occupies an excludable land strip of constant width extending between his or her residential location and the centre. Thus, for a given location, the derived demand for land in transportation is perfectly inelastic with respect to its price. Later on Pines (1971) introduced other inputs along with land in the production of roads, while Hochman and Pines (1971) and Oron, Pines and Sheshinski (1973) introduced a substitution between transportation infrastructure and all other goods. These extensions added more flexibility to the original specification of Mills, thus allowing for the derivation of additional results.

In the following discussion we present urban transportation as an excludable private good using the approach of Mohring and Harwitz (1962). We define a transportation market for each location x within the urban area. Denote the *trip demand* at x as $Q[x]$, which is given by

$$Q[x] = \int_{x}^{\bar{x}} n dx' \tag{7.1}$$

because every individual located beyond x' must cross the circle with radius x' on his or her way to the centre. Technology is represented by

$$\Psi[x] = \psi[z_t, l, M], \tag{7.2}$$

[1] See, for example, Dixit (1973), Legey, Ripper and Varaiya (1973), and Livesey (1973).

where Ψ the *marginal resource cost of travel* (with respect to distance) which is expressed in terms of the composite good; z_t is the *amount of the composite good* used in the construction of the road, l is the *width of the land strip used for roads,* and M is the *number of commuters using the specific road.* We assume that the above cost is an increasing function of M, and that the marginal productivity of infrastructure is diminishing.

Assumption 7.1:

(i) $\dfrac{\partial \psi}{\partial z_t} < 0$ and $\dfrac{\partial^2 \psi}{\partial z_t^2} > 0$

(ii) $\dfrac{\partial \psi}{\partial l} < 0$ and $\dfrac{\partial^2 \psi}{\partial l^2} > 0$

(iii) $\dfrac{\partial \psi}{\partial M} > 0$ and $\dfrac{\partial^2 \psi}{\partial M^2} > 0.$

We next characterise the transportation industry in terms of a standard production function. Toward this end multiply both sides of (7.2) by M to obtain $M\Psi = M\psi\left[z_t, l, M\right] \equiv a$, which represents the *aggregate marginal resource cost of travel* on a particular road at a particular distance from the centre. Then invert this relationship to obtain

$$M = f_t\left[z_t, l, a\right]. \tag{7.3}$$

In section 7.4.1 of the appendix we show that all partial derivatives of f_t are positive. Therefore (7.3) represents a standard production function, where M is the output (volume of traffic), z_t and l are fixed inputs required for the infrastructure, and a is the variable resource input. We also show in the appendix that if ψ exhibits zero returns to scale then f_t exhibits constant returns to scale. In this case an infinite number of viable transportation firms can prevail. Thus when ψ exhibits zero returns to scale, we have the standard conditions for perfect competition with marginal cost pricing. A similar case conducive to competitive equilibrium is one where the *average cost* per commuter

$$\mathrm{AC}\left[M\right] = \min_{\Psi, z_t, l}\left(\Psi + \frac{z_t + lR}{M}\right) \text{ subject to } M = f_t\left[z_t, l, a\right] \tag{7.4}$$

has a U–shape which is minimised at M^* such that $\bar{N}/M^*$ is a very large integer number.

In both these cases, under free entry, the supplier of transportation charges commuters the minimum average (aggregate) commuting cost which includes the marginal resource cost of travel. Therefore the commuting rate t is equal to AC$[M]$. Since, in equilibrium, result 2.2 implies that $dR^e/dx = -t^e/H^e = -\mathrm{AC}[M^e]/H^e < 0$, it follows by the application of the envelope theorem on (7.4) that

$$\frac{dt^e}{dx} = \frac{l^e}{M^e}\frac{dR^e}{dx} < 0. \tag{7.5}$$

In consequence we have

Result 7.1: The transportation cost increases at a decreasing rate in equilibrium.

Result 7.1 provides an explanation for the spatial structure of transportation cost as imposed by assumption 2.2. By considering explicitly urban transportation, we can justify this 'black box' assumption in the context of competitive market equilibria.

7.2 Market Failure

So far, we have proposed a model for the competitive provision of urban transportation, including the necessary infrastructure. Notice that our arguments in section 7.1 are based on two fundamental premises: (a) transportation is an excludable good, and (b) there is a large enough number of competitively sustainable roads. If either condition fails, the competitive market structure is not viable. Based on these observations, we can employ the artificial framework of section 7.1 to explain the supply of urban infrastructure by local governments in terms of market failure caused by non–excludability and increasing returns to scale.

7.2.1 Monitoring the Use of Transportation Infrastructure and Non–Excludability

The implementation of a market for urban transportation services requires monitoring individual trips in detail. Although the technology is now feasible in principle, and although partial systems have been tested with good results, there is still no instance of the full marginal cost pricing required by our framework for the competitive supply of transportation.[2] Thus transportation is a non–excludable good at present, and this is a major reason for market failure.

What about the future? All experts agree that marginal cost pricing is both feasible and desirable. "The theory is now refined and standard; implementation has been widely explored; numerous empirical studies have predicted its effects; and the whole package has made its way into standard textbooks in urban and transportation economics."[3] Yet it seems doubtful that the technique

[2] Small, Winston and Evans (1989) discuss an 'automatic vehicle identification' experiment in Hong Kong which involved twenty–six hundred vehicles over a period of twelve months. This experiment demonstrated that the available technology for determining marginal cost prices can satisfy very demanding goals for reliability and economy of application. In Canada one section of Ontario highway 407, the first fully automated toll freeway, has been opened in 1997. Sensors at the entry and exit points of this highway register vehicles with transponders, while those without have their plates photographed automatically and drivers are billed through the licence database.

[3] Small, Winston and Evans. (1989, p. 87) For a review of theory and associated empirical studies see Winston (1985).

will be fully applied to improve urban traffic conditions. The reasons are political rather than economic. Using the road is perceived as a fundamental right, and the idea that one can be excluded from the entire network altogether seems inconceivable. Individuals accept the idea of exclusion for parts of the network, such as certain highways, bridges and tunnels, for as long as they retain the right of using the network itself. Furthermore, the need for a detailed collection of data about individual trip–making raises important issues of privacy. Such records, where available, could be used for other purposes by governments. This is not to say that practical improvements cannot be realised using means which stop short of the stringent information requirements of comprehensive marginal cost pricing required by the theory of section 7.1. But anything short of these requirements will not satisfy the conditions of the theory. These difficulties indicate that the market failure arising from the presently non–excludable character of transportation may persist in the future.

7.2.2 Returns to Scale

Our arguments in section 7.1 were based on the assumption of constant returns to scale. As in the standard case of a private good, so in the case of transportation the competitive equilibrium is inconsistent with increasing returns to scale. The reason for this fundamental inconsistency is straightforward. By definition, increasing returns are associated with decreasing average cost. Therefore marginal cost is lower than average cost. Under these circumstances, marginal cost pricing, which characterises a competitive market structure, implies that the price is lower than average cost and hence it results in a loss.

The Meaning of Returns to Scale in Transportation

In section 7.4.1 of the appendix, we show that

$$\varepsilon_{\psi} \leq (\geq)\ 0 \Leftrightarrow \varepsilon_f \geq (\leq)\ 1 \tag{7.6}$$

where ε_{ψ} and ε_f denote the degree of homogeneity for ψ and f_t respectively. Thus a negative (zero, positive) degree of homogeneity of ψ implies increasing (constant, decreasing) returns to scale in traffic volume.

Since Mohring and Harwitz (1962), it is often assumed that

$$\Psi = \tilde{\psi}\left[M/b\left[z_t, l\right]\right] \text{ and } \tilde{\psi}' > 0 \tag{7.7}$$

where $B = b\left[z_t, l\right]$ is the *transportation infrastructure* corresponding to a particular road. Under this specification, we have

$$\begin{aligned} \Psi\varepsilon_{\tilde{\psi}} &= \tilde{\psi}'\frac{M}{B} - \tilde{\psi}'\frac{M}{B^2}\left(\frac{\partial b}{\partial z_t}z_t + \frac{\partial b}{\partial l}l\right) \\ &= \tilde{\psi}'\frac{M}{B}\left(1 - \varepsilon_b\right). \end{aligned} \tag{7.8}$$

Therefore

$$\varepsilon_b \geq (\leq)\ 1 \Leftrightarrow \varepsilon_{\tilde{\psi}} \leq (\geq)\ 0 \Leftrightarrow \varepsilon_f \geq (\leq)\ 1. \tag{7.9}$$

In words, under this restriction on technology, increasing (constant, decreasing) returns to scale in the production of infrastructure implies increasing (constant, decreasing) returns to scale in traffic volume. This approach, which was adopted by, among others, Keeler and Small (1977), Kraus (1981) and Small, Winston and Evans,(1989) implies that the only relevant scale economies are those associated with the transportation infrastructure.

There still remains the question of how to measure infrastructure scale economies. On the one hand both Strotz (1965) and Small, Winston, and Evans (1989) define returns to scale as a property of the cost function, so that "...often we are interested in expanding the urban road network by adding new roads, interchanges, skyways, etc. In these circumstances we are apt to encounter both more expensive construction and more expensive land acquisition costs. For this reason, my conjecture is that ... [$\Psi\varepsilon\psi$] is positive, or that, in improving the urban road network, we encounter *adverse* economies of scale. If so, road expenditure should be less than toll receipt, much as in an industry of decreasing returns, cost net of rents should be less than sales receipts (rents positive)." (Strotz (1965).) On the other hand both Kraus (1981) and Berechman and Pines (1991) define it as a property of the production function. In both cases the relationship between long–run marginal and average cost of travel can be used for evaluating scale economies. However, in evaluating scale economies, those two definitions differ from one another in the treatment of factor prices: whereas the former treats factor prices as variable, the latter treats them as constant.

We believe that prices should be kept constant. For, otherwise, any industry which is characterized by technical global increasing returns to scale but faces a factor supply which is not perfectly elastic, may survive in a competitive environment—a false implication under competitive equilibrium which, by definition, is based on price–taking behaviour.

Empirical Evidence

The empirical evidence regarding scale economies in transportation is not clear–cut. Keeler and Small (1977) estimate a degree of homogeneity for the transportation production function equal to 1.03, which implies that there are no statistically significant scale economies. Kraus (1981) on the other hand, based on engineering considerations, concludes that the average cost exceeds the marginal cost by 19%, which is equivalent to saying that the degree of homogeneity equals 1.19. More recently, Small, Winston and Evans (1989), support the original estimates of Keeler and Small, although they believe that they should be even lower taking into account the rising supply price of land.[4] If Keeler and

[4] Accounting not only for congestion but also for road-wear by truck traffic, they analyse the financing problem from a broader view than that adopted in our chapter. Formally they view the industry as supplying a joint product: volume of traffic (measured as the number

Small are right, returns to scale are not a cause for market failure. If, on the other hand, we rely on Kraus, his evidence implies that the competitive provision of transportation fails because producers must operate under a loss.

All this empirical work by Keeler and Small and by Kraus has been based on models that assume a constant degree of homogeneity. Casual observation, however, may suggest that global scale economies or diseconomies are not consistent with what we know about transportation. On the one hand global scale economies would suggest a single radial road, which defies any intuition we have about real transportation networks. On the other hand global scale diseconomies would suggest that all radial roads are as narrow as possible, which again seems implausible. What we actually observe is a number of relatively wide arterial roads in the radial transportation network of a city. This leads us to the conjecture that the typical road is neither characterized by global scale economies nor by global scale diseconomies, but rather by a U–shape cost function with an optimal (from the viewpoint of a supplier) number of commuters.[5] Consequently, in equilibrium, there is a finite number of radial roads. If this conjecture is true, the crucial question is whether or not the number of roads is sufficiently large to sustain competition. If not, firms can collude or behave strategically. We conclude that, whether it follows from global scale economies as claimed by Kraus, or from an insufficient number of efficient transportation firms, it seems that competitive market structure in transportation fails.

7.3 Public Supply of Transportation Infrastructure

In section 7.2, we have seen how the market for the transportation infrastructure can fail. If this is true, the public sector is required to step in to guarantee the efficient provision of urban transportation infrastructure. In this section we examine this possibility within the framework of a mixed economy in which the government supplies transportation infrastructure financed by taxes, while all remaining economic activities are performed in the market as before. We shall show that, if the government has sufficient instruments at its disposal, it

of passenger car equivalent that pass over the road during peak periods over an entire year) and traffic loadings (measured as the number of equivalent standard axle loads that pass over the road during the year). They conclude that there are scale economies in the two products taken separately, but since there are diseconomies of scope as well, the warranted user charge may cover up to 80% of the long-term capital and maintenance costs. In view of their finding about the presence of scope diseconomies, these authors recommend separating the roads for trucks and cars. If this were done, the scale diseconomies would imply higher deficits in the two separate systems than what is implied in the integrated one.

[5] Small, Winston and Evans (1989) report that, for the traffic volumes and loadings they use in their calculations, optimal highways must include two to four lanes in each direction. However, we should be careful in using this conclusion to support our conjecture about the U–shape relation because it relies on the diseconomies of scope in the joint production of traffic volumes and loadings. In the absence of such diseconomies, there would be a single road serving all passengers because the optimal width is as large as possible.

can achieve an efficient allocation. Furthermore, if the competitive equilibrium of section 7.1 does exist (in the sense that transportation is excludable and returns to scale are constant), then the outcome of the competitive allocation is the best that the government can achieve.

We consider a closed owner's city as in section 7.1. Without loss of generality, we simplify the transportation setup by suppressing the composite good as an input in the production of the infrastructure. We also assume that the partition of land available for transportation into roads, as well as the allocation of commuters among them, are efficient. Thus, if L_t denotes the *total width of land used in transportation* on the circle with radius x, only the aggregate quantities Q and L_t determine the transportation conditions so that $\Psi = \psi[Q, L_t]$.

In this mixed economy, the private sector takes the policy instruments of the government as parametric. Those instruments are the amount of land allocated to roads L_t, a *congestion toll* Φ_t at a particular distance from the centre and a *land conversion tax* $\bar{\Phi}_c$. The equilibrium of the private sector $\left(n^e, R^e, Q^e, \bar{Y}^e, \bar{U}^e, \bar{x}^e\right)$ satisfies

$$
\begin{aligned}
e\left[R^e, \bar{U}^e\right] &= \bar{Y}^e - \int_0^x \left(\psi\left[Q^e, L_t\right] + \Phi_t\right) dx' && (a) \\
n^e h\left[R^e, \bar{U}^e\right] + L_t &= \theta && (b) \\
\int_x^{\bar{x}^e} n^e dx &= Q^e && (c) \\
\int_{\mathcal{X}^e} n^e dx &= \bar{N} && (d) \\
R^e\langle\bar{x}^e\rangle &= \bar{R} + \bar{\Phi}_c && (e) \\
\bar{N}\bar{\Omega} &= \int_{\mathcal{X}^e} \left(n^e z\left[R^e, \bar{U}^e\right] + Q^e \psi\left[Q^e, L_t\right]\right) dx \\
&\quad + \Theta\langle\bar{x}\rangle \bar{R}. && (f)
\end{aligned}
\qquad (7.10)
$$

In this definition, z, h, θ, ψ, L_t and Φ_t are known functions, and $\bar{\Omega}, \bar{N}, \bar{R}$ and $\bar{\Phi}_c$ are given parameters. The first equilibrium condition differs from the corresponding (2.8) in two respects. Firstly, the cost of transportation is now determined by the aggregate demand for transportation and the public supply of transportation infrastructure. Secondly, the cost of transportation is also modified to take into account that, under public supply of transportation, a congestion toll Φ_t may be collected. The fifth equilibrium condition differs from the corresponding (2.11) in that conversion of land from agricultural to urban can now be taxed. The last equilibrium condition is the resource constraint modified to reflect the transport cost as a function of crowding and infrastructure. By the Walras law, the budget and resource constrains determine income as

$$\begin{aligned}\bar{Y}^e &= \bar{\Omega} + \frac{1}{\bar{N}} \int_{\mathcal{X}^e} \theta\left(R^e - \bar{R}\right) dx \\ &+ \frac{1}{\bar{N}} \left(\int_{\mathcal{X}^e} n^e \int_0^x \Phi_t dx' dx - \int_{\mathcal{X}^e} L_t R^e dx \right).\end{aligned} \tag{7.11}$$

7.3.1 First–Best Allocation

We define that the problem of the government is to choose $\left(L_t^{\text{o}}, \Phi_t^{\text{o}}, \bar{\Phi}_c^{\text{o}}\right)$ that maximises the equilibrium utility level subject to the equilibrium conditions (7.10). In this problem, $\left(L_t, \Phi_t, \bar{\Phi}_c\right)$ are control variables and $(n^e, R^e, Q^e, \bar{Y}^e, \bar{U}^e, \bar{x}^e)$ are endogenous variables.[6] Section 7.4.2 of the appendix contains the corresponding Lagrangean function modified to suit our analysis and the derivation of the first–order conditions. As we show in this appendix, a first–best policy imposes the following requirements. Firstly, the optimal congestion toll must be given by

$$\Phi_t^{\text{o}} = Q^e \frac{\partial \psi}{\partial Q^e}. \tag{7.12}$$

This is the well–known marginal social cost pricing rule in the presence of congestion effects (Pigouvian taxation of an externality). Secondly, in the first–best allocation there is no need for a land conversion tax:

$$\bar{\Phi}_c^{\text{o}} = 0. \tag{7.13}$$

Finally, the optimal rent must satisfy

$$R^e = -Q^e \frac{\partial \psi}{\partial L_t^{\text{o}}}. \tag{7.14}$$

This represents the Samuelson (1954) rule for the optimal provision of a public good. The LHS of (7.14) is the social opportunity cost of provision given by the alternative cost of land used for transportation, that is, the value of residential land foregone. The RHS represents the corresponding social benefit given by the reduction in transport cost associated with the land increment.

The competitive allocation of section 7.1, if it exists, is equivalent to the corresponding optimal allocation of the mixed economy.[7] this happens because the first welfare theorem applies. However, as we noted above, the market may fail

---

[6] The reader should note the following special notations which apply to the rest of this chapter. We denote $\psi\left[Q^e, L_t^{\text{o}}\right]$ simply as Ψ. We also write

$$\frac{\partial \psi}{\partial Q^e} \equiv \frac{\partial \psi}{\partial Q} \langle Q^e, L_t^{\text{o}} \rangle \text{ and } \frac{\partial \psi}{\partial L_t^{\text{o}}} \equiv \frac{\partial \psi}{\partial L_t} \langle Q^e, L_t^{\text{o}} \rangle.$$

[7] Since, in the mixed economy, we suppressed z_t in $\psi[\cdot]$, we make the comparison with a simplified competitive equilibrium in which the composite good is not a fixed input in the production of transportation.

to exist because of increasing returns to scale. In that case the mixed–economy allocation is necessary for achieving optimality. Such optimality, however, becomes unfeasible when monitoring the use of roads is too expensive. In that case the market cannot provide the infrastructure altogether, while the local government, which is able to finance its provision, must adopt a second–best policy.

7.3.2 Second–Best Allocation

We have already discussed in section 7.2 why imposing optimal congestion tolls is unlikely to happen because of the strong political opposition it creates. It is therefore relevant to examine mixed economies in which the optimal congestion toll is not a policy instrument for the government. These give rise to 'second–best' allocations, in which the main issue is about the nature of the appropriate cost–benefit calculation to apply for allocating land to roads. The first one to raise this issue was Solow (1973), who suggested that the shadow land rent to be used in calculating the cost of the infrastructure must exceed the corresponding market rent. Consequently, Solow believed, too much land is used for the transportation infrastructure relative to the first–best rule (7.14). This implies that, on the margin, the saving in transportation cost associated with an infrastructure project should be strictly higher than its marginal cost in order to justify implementation. This conclusion was later challenged by several authors, mainly Kanemoto (1975, 76, 77, 80) and Arnott (1979).[8] In response to Solow, these authors investigated the appropriate second–best criterion to be used, as well as the implications of using a wrong criterion based on the market rent and on the direct saving in transport cost associated with allocating land to roads (often referred to as 'naive' or 'market' social benefit). Although the latter issue is interesting historically, we focus our discussion on the appropriate second–best criterion in allocating land to roads.

Formally, a second–best allocation requires to solve the first–best problem with the additional constraint

$$\Phi_t = 0. \tag{7.15}$$

In section 7.4.3 of the appendix we show that

$$\lambda_2^o = -\left(Q^e - \Lambda_1^o\right)\frac{\partial \psi}{\partial L_t^o} \tag{7.16}$$

where λ_2 and Λ_1 are Lagrangean multipliers, the former associated with the land–clearing constraint (7.10(b)) and the latter with the new constraint (7.15). The LHS of this equation, the shadow price of land, represents the *marginal social cost* MSC of allocating land to roads, while the RHS represents the corresponding true *marginal social benefit* MSB.

---

[8] Arnott raises a wider issue, about the distortion in land rents according to land–use type (residential and transportation) and for any arbitrary system of roads. This analysis has been extended by Pines and Sadka (1985).

Lemma 7.1:

$$\Lambda_1^o \langle 0 \rangle = -\Lambda_1^o \langle \bar{x}^e \rangle = 0 \qquad (a)$$

$$\Lambda_1^o [x] > 0 \qquad \text{for} \quad x \in (0, \bar{x}^e)\,. \qquad (b)$$

PROOF: See section 7.4.4 of the appendix. □

The first part of lemma 7.1 implies that, both at the centre and at the border of the city, the MSB in (7.16) is equal to the *direct* saving in transportation cost, $-Q^e \partial\psi/\partial L_t^o$, associated with allocating land to roads.[9] This direct effect is often referred to as the 'naive' or 'market' benefit. However, in the rest of the city, Λ_1^o does not vanish as stated in the second part of lemma 7.1. This implies that, in the interior of the city, the true MSB is smaller than the direct saving in transportation cost. The explanation is straightforward. Notice that Λ_1^o is the shadow price of constraint (7.15), thus reflecting the MSB of increasing the congestion toll. Since, under the first–best, the optimal toll is positive by (7.12), increasing the congestion toll marginally from zero is expected to improve welfare, that is, $\Lambda_1^o > 0$ for $\Phi_t = 0$. In fact, any increase in the marginal rate is a substitute for the toll, thus contributing to improve welfare. However, the increase in road width reduces the marginal rate by $-\partial\psi/\partial L_t^o$, thus reducing the direct saving effect by $-\Lambda_1^o \partial\psi/\partial L_t^o$. In summary, we have

Result 7.2 (Kanemoto (1977), Arnott (1979): Under second–best conditions, the marginal social benefit of allocating land to roads is given by

$$\text{MSB} = -Q^e \frac{\partial\psi}{\partial L_t^o}$$

for $x = 0$ and $x = \bar{x}^e$, and by

$$\text{MSB} = -\left(Q^e - \Lambda_1^o\right) \frac{\partial\psi}{\partial L_t^o} < -Q^e \frac{\partial\psi}{\partial L_t^o}$$

for $x \in (0, \bar{x}^e)$.

The relationship between the shadow price of land and the market rent is more subtle. It is described in the following

Result 7.3 (Kanemoto (1977), Arnott (1979):[10] Under second–best conditions, there exists a unique $x^* \in (0, \bar{x}^e)$, such that the

---

[9] At the end points, Λ_1^o vanishes because the transportation demand is infinitely inelastic: at $x = 0$ we have $Q[0] = \bar{N}$, and at $x = \bar{x}^e$ we have $Q[\bar{x}^e] = 0$.

[10] Although we state result 7.3 in the context of a second-best analysis as in Kanemoto (1980), the proof does not depend on the optimal (second–best) allocation of land to roads. Thus result 7.3 holds for any given allocation of land to roads. This point was emphasized in Arnott (1979). Furthermore, as in previous chapters, we assume that

$$\frac{\partial^2 e^e}{\partial R^2} = \frac{\partial h^e}{\partial R} < 0.$$

Thus we disregard the special case, discussed extensively in the literature, where the elasticity of the compensated demand is infinitely inelastic, i.e., when $\partial h^e/\partial R = 0$ (see, for example, chapter five in Kanemoto (1980)).

marginal social cost of allocating land to roads is given by

$$\text{MSC} = \lambda_2^o > R^e$$

for $0 \leq x < x^*$, and by

$$\text{MSC} = \lambda_2^o < R^e.$$

for $x^* \leq x < \bar{x}^e$. For $x = \bar{x}^e$, we have

$$\text{MSC} = \lambda_2^o = \bar{R}.$$

PROOF: See section 7.4.5 of the appendix. □

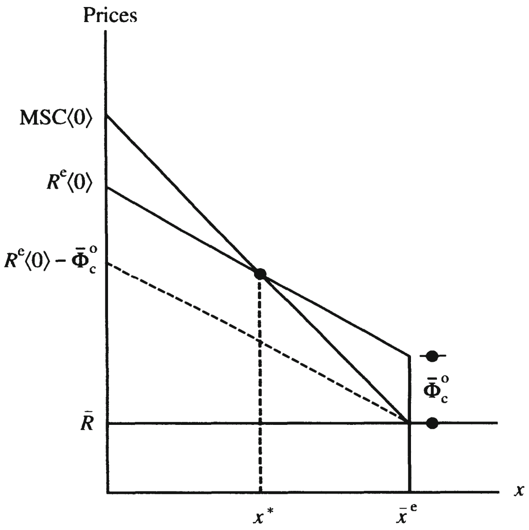

FIGURE 7.1. Market and Shadow Prices of Urban Land.

Result 7.3 is illustrated in figure 7.1. The explanation is, once again, straightforward. Any increase in land supply at any given location attracts residents from other parts of the city. If the increment in land supply occurs close to the centre, it contributes not only to an increase of housing consumption and to a decrease of individual transportation cost, but also to a decrease of the external effects by reducing the total demand on the congested roads. Since these external effects are not internalised by a congestion toll, the shadow price of urban land is higher than its market value close to the centre. The opposite is true if the increment occurs close to the border. There, the increased land supply attracts residents from more central locations, thus contributing to an increase of the external effects by augmenting the total demand on the congested roads. Consequently the shadow price of urban land is lower than its market value close to the border. Since the alternative cost of land at the border is $\bar{R}$, we obtain the last part of result 7.4. But then it must be the case that $R^e \langle \bar{x}^e \rangle > \bar{R}$. This difference is absorbed by the land conversion tax

$$\begin{aligned} \bar{\Phi}_c^{\text{o}} &= R^e \langle \bar{x}^e \rangle - \lambda_2^{\text{o}} \langle \bar{x}^e \rangle \\ &= R^e \langle \bar{x}^e \rangle - \bar{R} > 0. \end{aligned} \tag{7.17}$$

It follows that, contrary to the first–best case, where there was no land conversion tax, this tax is necessary under second–best conditions as a wedge between the equilibrium and agricultural rents at the border. The analysis above can also rationalise planning regulations designed to discourage the conversion of agricultural land to urban land use at the fringe of a city.

Results 7.2 and 7.3 imply that, in general, under second–best conditions, neither the MSB equals the direct reduction in transport cost, nor the equilibrium land rent is represented by the MSC. Formally, this happens because, in contrast to first–best, the simple envelope property does not apply under second–best conditions.

7.4 Appendices

7.4.1 Proof of (7.6)

Differentiate (7.2) and (7.3), and evaluate the homogeneity degrees of ψ and f_t to obtain

$$\begin{aligned} \Psi \varepsilon_\psi &= \frac{\partial \psi}{\partial z_t} z_t + \frac{\partial \psi}{\partial l} l + \frac{\partial \psi}{\partial M} M \quad (a) \\ M \varepsilon_f &= \frac{\partial f_t}{\partial z_t} z_t + \frac{\partial f_t}{\partial l} l + \frac{\partial f_t}{\partial a} a. \quad (b) \end{aligned} \tag{7.18}$$

We can also write, using the same equations,

$$M = f_t \left[z_t, l, M\psi \left[M, l, z_t \right] \right]. \tag{7.19}$$

The derivatives of (7.19) are given by

$$\begin{aligned} 0 &= M \frac{\partial f_t}{\partial a} \frac{\partial \psi}{\partial z_t} + \frac{\partial f_t}{\partial z_t} \quad (a) \\ 0 &= M \frac{\partial f_t}{\partial a} \frac{\partial \psi}{\partial l} + \frac{\partial f_t}{\partial l} \quad (b) \\ 1 &= \frac{\partial f_t}{\partial a} \left(\Psi + \frac{\partial \psi}{\partial M} M \right). \quad (c) \end{aligned} \tag{7.20}$$

Using assumption 7.1 on (7.20), we have

$$\frac{\partial f_t}{\partial z_t} > 0, \frac{\partial f_t}{\partial l} > 0 \text{ and } \frac{\partial f_t}{\partial a} > 0. \tag{7.21}$$

Upon substitution of (7.20) in (7.18(*b*)), we obtain

$$\begin{aligned} M\varepsilon_f &= -M\frac{\partial f_t}{\partial a}\left(\frac{\partial \psi}{\partial z_t}z_t + \frac{\partial \psi}{\partial l}l + \frac{\partial \psi}{\partial M}M\right) + M \\ &\underset{(7.18)}{=} M\left(1 - \frac{\partial f_t}{\partial a}\left(\frac{\partial \psi}{\partial z_t}z_t + \frac{\partial \psi}{\partial l}l + \frac{\partial \psi}{\partial M}M\right)\right) \\ &\underset{(7.20)}{=} M\left(1 - \Psi\varepsilon_\psi\frac{\partial f_t}{\partial a}\right). \end{aligned} \tag{7.22}$$

Taking into account that $\partial f_t/\partial a > 0$, we conclude that $\varepsilon_\psi \leq (\geq)\ 0 \Leftrightarrow \varepsilon_f \geq (\leq)\ 1$.

7.4.2 First–Best Allocation Problem

It is convenient to replace some equilibrium conditions in (7.10) with other, equivalent, conditions. Taking into account that (7.10(*c*)) is equivalent to

$$\begin{aligned} \frac{dQ^e}{dx} &= -n^e \quad (a) \\ Q^e\langle\bar{x}\rangle &= 0, \quad (b) \end{aligned} \tag{7.23}$$

and that (7.10(*d*)) is equivalent to

$$Q^e\langle 0\rangle = \bar{N}, \tag{7.24}$$

we can write the Lagrangean function of the government's problem as

$$
\begin{aligned}
\frac{\mathcal{L}\left[L_t, \boldsymbol{\Phi}; \boldsymbol{\lambda}\right]}{\bar{\lambda}_7} &= \frac{\bar{U}^e}{\bar{\lambda}_7} \\
&- \int_{\mathcal{X}^e} \lambda_1 \left(e\left[R^e, \bar{U}^e\right] + \int_0^x \left(\psi\left[Q^e, L_t\right] + \Phi_t\right) dx' - \bar{Y}^e\right) dx \\
&- \int_{\mathcal{X}^e} \lambda_2 \left(n^e H^e + L_t - \theta\right) dx \\
&- \int_{\mathcal{X}^e} \lambda_3 n^e dx + \int_{\mathcal{X}^e} \frac{d\lambda_3}{dx} Q^e dx - \lambda_3 \langle \bar{x}^e \rangle Q^e \langle \bar{x}^e \rangle \\
&+ \lambda_3 \langle 0 \rangle Q^e \langle 0 \rangle - \bar{\lambda}_4 Q^e \langle \bar{x}^e \rangle - \bar{\lambda}_5 \left(Q^e \langle 0 \rangle - \bar{N}\right) \\
&- \bar{\lambda}_6 \left(R^e \langle \bar{x}^e \rangle - \bar{R} - \bar{\Phi}_c\right) \\
&- \left(\int_{\mathcal{X}^e} \left(n^e Z^e + \int_0^{x'} Q^e \Psi^e dx'\right) dx + \Theta \langle \bar{x}^e \rangle \bar{R} - \bar{N} \bar{\Omega}\right),
\end{aligned} \tag{7.25}
$$

where $\boldsymbol{\Phi} \equiv \left(\Phi_t, \bar{\Phi}_c\right)$ and $\boldsymbol{\lambda} \equiv \left(\lambda_1, ..., \bar{\lambda}_7\right)$, and where integration by parts has been used on the constraint associated with $(7.23(a))$.[11] Notice that we have normalised (7.25) in terms of $\bar{\lambda}_7$, which is the shadow price associated with the resource constraint. Thus all shadow prices in (7.25) are defined relative to $\bar{\lambda}_7$. By inspection we observe that, since Φ_t appears only in the first constraint of the problem, we have $\partial \mathcal{L}/\partial \Phi_t = \lambda_1^{\circ} = 0$. Thus, we can disregard the second term on the RHS of the Lagrangean. For the same reason $\partial \mathcal{L}/\partial \bar{\Phi}_c = \bar{\lambda}_6^{\circ} = 0$, and we can also disregard the sixth term.

We begin by taking the first–order condition of the Lagrangean with respect to land rent:

$$
-n^e \left(\frac{\partial z^e}{\partial R} + \lambda_2^{\circ} \frac{\partial h^e}{\partial R}\right) = 0, \tag{7.26}
$$

where λ_2° denotes the optimal value of the shadow price associated with the urban land constraint $(7.10(c))$. From the derivative property of the expenditure function we have

$$
\frac{\partial z^e}{\partial R} + R^e \frac{\partial h^e}{\partial R} = 0. \tag{7.27}
$$

[11] This constraint is represented in the original Lagrangean function as

$$
\int_{\mathcal{X}^e} \lambda_3 \left(n^e + \frac{dQ^e}{dx}\right) dx = \int_{\mathcal{X}^e} \lambda_3 n^e dx + \int_{\mathcal{X}^e} \lambda_3 \frac{dQ^e}{dx} dx.
$$

Integrating by parts the second term on the RHS modifies the constraint as in (7.25).

Comparing (7.26) and (7.27), and taking into account that $n^e > 0$ and $\partial h/R^e < 0$ (the latter holding because of the substitution effect), we conclude that

$$R^e = \lambda_2^o. \tag{7.28}$$

The first–order condition of the Lagrangean with respect to population on the ring at a particular distance from the centre is

$$-\lambda_2^o H^e - \lambda_3^o - Z^e = 0 \tag{7.29}$$

where λ_3 corresponds to the population constraint (7.10(b)). Using (7.28) on this last equation we derive

$$e\left[R^e, \bar{U}^e\right] = -\lambda_3^o. \tag{7.30}$$

We next turn to the volume of traffic at a particular distance from the centre. The corresponding first–order condition is

$$\frac{d\lambda_3^o}{dx} - \Psi - Q^e \frac{\partial \psi}{\partial Q^e} = 0, \tag{7.31}$$

where $\Psi^o = \psi\left[Q^e, L_t^o\right]$ denotes the optimum marginal resource cost of travel at a particular distance from the centre. Integrating (7.31), we obtain

$$\lambda_3^o|_0^x - \int_0^x \left(\Psi^o + Q^e \frac{\partial \psi}{\partial Q^e}\right) dx' = 0. \tag{7.32}$$

Upon substitution of (7.30) in (7.32), we have

$$e\left[R^e\langle x\rangle, \bar{U}^e\right] - e\left[R^e\langle 0\rangle, \bar{U}^e\right] = -\int_0^x \left(\Psi + Q^e \frac{\partial \psi}{\partial Q^e}\right) dx'. \tag{7.33}$$

Taking also into account that (7.10) implies

$$e\left[R^e\langle x\rangle, \bar{U}^e\right] - e\left[R^e\langle 0\rangle, \bar{U}^e\right] = -\int_0^x \left(\Psi + \Phi_t^o\right) dx' \tag{7.34}$$

Condition (7.12) follows from (7.33) and (7.34).

Taking the first–order condition with respect to the border of the city we have

$$-\left(n^e \lambda_3^o + n^e Z^e + \theta \bar{R}\right)\Big|_{x=\bar{x}^e} = 0. \tag{7.35}$$

Using (7.10(b)) and (7.30), this becomes

$$\left(n^e H^e R^e - \theta \bar{R}\right)\Big|_{x=\bar{x}^e} = 0. \tag{7.36}$$

At the border of the city, we know that no land is allocated to roads because the volume of traffic is zero there.[12] Hence all land available at the border must be residential which, in conjunction with (7.36), leads to

$$R^e\langle \bar{x}^e\rangle = \bar{R}. \tag{7.37}$$

[12] For a detailed discussion of this issue see Kanemoto (1980, chapter four) and especially footnote 1 on p.135.

This, and (7.10(e)), imply condition (7.13).

Finally, taking the first–order condition with respect to land allocated to roads at any $x < \bar{x}^e$, we obtain

$$-\lambda_2^{\circ} - Q^e \frac{\partial \psi}{\partial L_t^{\circ}} = 0. \tag{7.38}$$

Substitution of (7.28) in (7.38) gives (7.14).

7.4.3 Second–Best Allocation Problem

We use the same Lagrangean multipliers as for the first–best case, with the addition of Λ_1 which corresponds to (7.15). We therefore augment the Lagrangean (7.25) by $\Lambda_1 \Phi_t$. In order to derive the following first–order conditions we define the *congestion toll* paid by someone at distance x from the centre as

$$\mathbb{T}[x] \equiv \int_0^x \Phi_t [x'] \, dx'. \tag{7.39}$$

Thus, after integration by parts, the two components of the Lagrangean associated with the congestion toll are given by

$$\begin{aligned} -\lambda_1 \mathbb{T} - \Lambda_1 \frac{d\mathbb{T}}{dx} &= -\lambda_1 \mathbb{T} + \int_0^x \mathbb{T}[x'] \frac{d\Lambda_1}{dx'} dx' \\ &\quad -\Lambda_1 \langle \bar{x}^e \rangle \mathbb{T} \langle \bar{x}^e \rangle + \Lambda_1 \langle 0 \rangle \mathbb{T} \langle 0 \rangle . \end{aligned} \tag{7.40}$$

With these modifications of the Lagrangean (7.25) we get

$$\begin{aligned} -\lambda_1^{\circ} H^e - n^e \left(\frac{\partial z^e}{\partial R} + \lambda_2^{\circ} \frac{\partial h^e}{\partial R} \right) &= 0 && (a) \\ -\lambda_2^{\circ} H^e - \lambda_3^{\circ} - Z^e &= 0 && (b) \\ \Lambda_1^{\circ} \frac{\partial \psi}{\partial Q^e} + \frac{d\lambda_3^{\circ}}{dx} - \Psi - Q^e \frac{\partial \psi}{\partial Q^e} &= 0 && (c) \\ -\left(n^e \lambda_3^{\circ} + n^e Z^e + \theta \bar{R} \right)\Big|_{x=\bar{x}^e} &= 0 && (d) \\ \Lambda_1^{\circ} \frac{\partial \psi}{\partial L_t^{\circ}} - \lambda_2^{\circ} - Q^e \frac{\partial \psi}{\partial L_t^{\circ}} &= 0 && (e) \\ \Lambda_1^{\circ} \langle 0 \rangle &= 0 && (f) \\ -\Lambda_1^{\circ} \langle \bar{x}^e \rangle &= 0 && (g) \\ \lambda_1^{\circ} &= \frac{d\Lambda_1^{\circ}}{dx} && (h) \end{aligned} \tag{7.41}$$

where the last three conditions have been obtained by taking the derivative of the Lagrangean in terms of $\mathbb{T}\langle 0\rangle$, $\mathbb{T}\langle \bar{x}^e\rangle$ and $\mathbb{T}$ respectively. Condition (7.16) can be obtained by rewriting (7.41(e)).

7.4.4 Proof of Lemma 7.1

The proof is based on Pines and Sadka (1985).

The first part of the lemma follows from (7.41(f)) and (7.41(g)).

We start the proof of the second part by differentiating (7.41(b)) with respect to x to obtain

$$-\frac{\partial z^e}{\partial R}\frac{dR^e}{dx} - \frac{d\lambda_2^{\circ}}{dx}H^e - \lambda_2^{\circ}\frac{\partial h^e}{\partial R}\frac{dR^e}{dx} - \frac{d\lambda_3^{\circ}}{dx} = 0. \tag{7.42}$$

Using the derivative property of the expenditure function and (7.41(c)), (7.42) becomes

$$\frac{\partial h^e}{\partial R}\frac{dR^e}{dx}R^e - \frac{d\lambda_2^{\circ}}{dx}H^e - \lambda_2^{\circ}\frac{\partial h^e}{\partial R}\frac{dR^e}{dx} - \Psi - Q^e\frac{\partial\psi}{\partial Q^e} + \Lambda_1^{\circ}\frac{\partial\psi}{\partial Q^e} = 0. \tag{7.43}$$

Taking into account that Ψ° represents the transportation rate in the second–best, and replacing it through result 2.2, which still holds, we have

$$\begin{aligned}\frac{d}{dx}\left(H^e\left(R^e - \lambda_2^{\circ}\right)\right) &= \frac{\partial h^e}{\partial R}\frac{dR^e}{dx}\left(R^e - \lambda_2^{\circ}\right) + H^e\left(\frac{dR^e}{dx} - \frac{d\lambda_2^{\circ}}{dx}\right) \\ &= \left(Q^e - \Lambda_1^{\circ}\right)\frac{\partial\psi}{\partial Q^e}.\end{aligned} \tag{7.44}$$

If we define

$$\Xi \equiv H^e\left(R^e - \lambda_2^{\circ}\right), \tag{7.45}$$

(7.44) implies

$$\frac{d\Xi}{dx} = \left(Q^e - \Lambda_1^{\circ}\right)\frac{\partial\psi}{\partial Q^e}. \tag{7.46}$$

We now employ (7.41(a)), (7.41(h)) and the derivative property of the expenditure function to obtain

$$\begin{aligned}\frac{d\Lambda_1^{\circ}}{dx} &= -\frac{n^e}{H^e}\left(\frac{\partial z^e}{\partial R} + \lambda_2^{\circ}\frac{\partial h^e}{\partial R}\right) \\ &= \frac{n^e}{H^e}\frac{\partial h^e}{\partial R}\left(R^e - \lambda_2^{\circ}\right).\end{aligned} \tag{7.47}$$

Recalling that the substitution effect $\partial h/\partial R^e$ is negative, comparison of (7.45) and (7.47) gives

$$\text{sign } \Xi = -\text{sign } \frac{d\Lambda_1^{\circ}}{dx}. \tag{7.48}$$

Now suppose that the second part of the lemma is false. Then there exists $x_1 \in (0, \bar{x}^e)$ where $\Lambda_1^o \langle x_1 \rangle \leq 0$ and, in view of the first part (conditions (7.41(f)) and (7.41(g)), Λ_1^o attains a global minimum. This implies $\Lambda_1^o \langle x_1 \rangle \leq 0$ and, by (7.48), $\text{sign}\, \Xi \langle x_1 \rangle = -\text{sign}\, d\Lambda_1^o \langle x_1 \rangle / dx = 0$. Also, since Λ_1^o achieves a minimum at x_1, there must exist $x_2 \in [x_1, x_1+\epsilon)$ where $\Xi \langle x_2 \rangle$ is as close to zero as we wish while $\text{sign}(d\Lambda_1^o \langle x_2 \rangle / dx) \geq 0$. Therefore, by (7.48), $\text{sign}\, \Xi \langle x_2 \rangle = -\text{sign}\, d\Lambda_1^o \langle x_3 \rangle / dx \leq 0$. It follows that there exists $x_3 \in (x_1, x_2)$ such that

$$\frac{d\Xi \langle x3 \rangle}{dx} \leq 0 \tag{7.49}$$

while $\Lambda_1^o \langle x_3 \rangle$ is either non–positive or sufficiently close to zero to ascertain that, in any case, $\Lambda_1^o \langle x_3 \rangle \leq Q^e \langle x_3 \rangle$. But, then, it follows from (7.46) that

$$\frac{d\Xi \langle x_3 \rangle}{dx} = \left(Q^e \langle x_3 \rangle - \Lambda_1^o \langle x_3 \rangle\right) \frac{\partial \psi}{\partial Q^e} > 0, \tag{7.50}$$

since $\partial \psi / \partial N^e > 0$. This contradiction establishes the lemma.

7.4.5 Proof of Result 7.3

The proof is based on Pines and Sadka (1985). We proceed in steps.

(1) Conditions (7.41(f)) and (7.41(g)), in conjunction with lemma 7.1, imply that $\Lambda_1^o \langle 0 \rangle = 0 < \Lambda_1^o \langle \epsilon \rangle$ and $\Lambda_1^o \langle \bar{x}^e - \epsilon \rangle > \Lambda_1^o \langle \bar{x}^e \rangle = 0$. It follows that $d\Lambda_1^o / dx > 0$ for $x = 0$ and $d\Lambda_1^o / dx < 0$ for $x = \bar{x}^e$. Therefore, from (7.45) and (7.48), we have

$$\text{sign}\left(R^e \langle 0 \rangle - \lambda_2^o \langle 0 \rangle\right) = \text{sign}\, \Xi \langle 0 \rangle = -\text{sign}\, \frac{d\Lambda_1^o}{dx} \langle 0 \rangle < 0 \tag{7.51}$$

and

$$\text{sign}\left(R^e \langle \bar{x}^e \rangle - \lambda_2^o \langle \bar{x}^e \rangle\right) = \text{sign}\, \Xi \langle \bar{x}^e \rangle = -\text{sign}\, \frac{d\Lambda_1^o}{dx} \langle \bar{x}^e \rangle > 0. \tag{7.52}$$

(2) We will show that for any $x_1 \in (0, \bar{x}^e)$ satisfying $\Xi \langle x_1 \rangle \leq 0$, it must be the case that $d\Xi \langle x_1 \rangle / dx > 0$. Suppose not. Then we have $\text{sign}\, \Xi \langle x_1 \rangle = \text{sign}\left(R^e \langle x_1 \rangle - \lambda_2^o \langle x_1 \rangle\right) \leq 0$ and $d\Xi \langle x_1 \rangle / dx \leq 0$. It follows from (7.46) that $\left(Q^e \langle x_1 \rangle - \Lambda_1^o \langle x_1 \rangle\right) \leq 0$. Furthermore, since $\Xi \langle x_1 \rangle \leq 0$, it follows from (7.48) that $\text{sign}\, d\Lambda_1^o / dx \geq 0$. Then, since $\left(Q^e \langle x_1 \rangle - \Lambda_1^o \langle x_1 \rangle\right) \leq 0$, $d\Lambda_1^o / dx \geq 0$, and Q^e is always decreasing with x, (7.46) implies that $\text{sign}\, \Xi \langle x \rangle = \text{sign}(R^e \langle x \rangle - \lambda_2^o \langle x \rangle)$ must remain nonpositive at any $x_1 + \epsilon$ where ϵ is sufficiently small. But then it must become *negative* over the entire interval $[x_1, \bar{x}^e]$. This, however contradicts (7.52).

(3) It follows from steps (1) and (2) that there must exist some $x^* \in (0, \bar{x}^e)$ such that $\Xi \langle x^* \rangle = R^e \langle x^* \rangle - \lambda_2^o \langle x^* \rangle = 0$, sign $\Xi \langle x \rangle = \text{sign}(R^e \langle x \rangle - \lambda_2^o \langle x \rangle) < 0$ for $x \in [0, x^*)$, and $\text{sign}\, \Xi \langle x \rangle = \text{sign}(R^e \langle x \rangle - \lambda_2^o \langle x \rangle) > 0$ for $x \in (x^*, \bar{x}^e]$.

7.5 References

R. J. Arnott, 1979, Unpriced Transportation Congestion. *Journal of Economic Theory* 21, 294–314.

J. Berechman and D. Pines, 1991, Financing Road Capacity and Returns to Scale Under Marginal Cost Pricing: A Note. *Journal of Transport Economics and Policy*, 177–181.

A. K. Dixit, 1973, The Optimum Factory Town. *Bell Journal of Economics and Management Science* 4, 637–651.

O. Hochman and D. Pines, 1971, Competitive Equilibrium of Transportation and Housing in the Residential Ring of an Urban Area. *Environment and Planning* 3, 51–62.

Y. Kanemoto, 1975, Congestion and Cost–Benefit Analysis in Cities. *Journal of Urban Economics* 2, 246–264.

Y. Kanemoto, 1976, Optimum, Market, and Second–Best Land Use Patterns in a von Thünen City with Congestion. *Regional Science and Urban Economics* 6, 23–32.

Y. Kanemoto, 1977, Cost–Benefit Analysis and the Second–Best Land Use in Transportation. *Journal of Urban Economics* 4, 483–507.

Y. Kanemoto, 1980, *Theories of Urban Externalities*. North–Holland, Amsterdam.

E. Keeler and K. A. Small, 1977, Optimal Peak–Load Pricing, Investment, and Service Levels on Urban Expressways. *Journal of Political Economy* 85, 1–25.

M. Kraus, 1981, Scale Economies Analysis of Urban Highway Networks. *Journal of Urban Economics* 9, 1–22.

L. Legey, M. Ripper and P. Varaiya, 1973, Effects of Congestion on the Shape of the City. *Journal of Economic Theory* 6, 162–179.

D. A. Livesey, 1973, Optimum City Size: A Minimum Congestion Cost Approach. *Journal of Economic Theory* 6, 144–161.

E. S. Mills, 1967, An Aggregative Model of Resource Allocation in a Metropolitan Area. *American Economic Review* 57, 197–210.

H. Mohring and M. Harwitz, 1962, *Highway Benefits*. Northwestern University Press, Evanston.

Y. Oron, D. Pines and E. Sheshinski, 1973, Optimum Versus Equilibrium Land Use Pattern and Congestion Toll. *Bell Journal of Economics and Management Science* 4, 619–636.

D. Pines, 1971, The Exponential Density Function. *Journal of Regional Science* 10, 107–110.

D. Pines and E. Sadka, 1985, Zoning, First–Best, Second–Best, and Third–Best Criteria for Allocating Land to Roads. *Journal of Urban Economics* 17, 167–183.

K. A. Small, C. Winston and C. A. Evans, 1989, *Road Work: A New Highway Pricing and Investment Policy*. The Brookings Institution, Washington D.C.

R. M. Solow, 1973, Congestion Cost and the Use of Land for Streets. *Bell Journal of Economics and Management Science* 4, 602–618.

R. H. Strotz, 1965, Urban Transportation Parables. In J. Margolis (ed.) *The Public Economy of Urban Communities.* Resources for the Future, Washington D.C., 127–169.

C. Winston, 1985, Conceptual Developments in the Economics of Transportation: An Interpretive Survey. *Journal of Economic Literature* 23, 57–94.

8

From Monocentricity to Polycentricity

This chapter provides a bridge between the first and the second parts of our book. It defends the utility of the monocentric paradigm which has been the subject of growing criticism in recent years, and it provides background to support the need for a robust polycentric city model. We explore the relationship between the *exponential density function* [EDF] and the monocentric city model, and we examine to what extent a failure of the EDF invalidates the monocentric paradigm itself. We explain why a negative correlation between the density gradient and city size does not imply a failure of the monocentric paradigm—not even a failure of the EDF. We also discuss the declining rôle of the CBD, reasons why such decline does not invalidate the monocentric city model and specific ways in which this paradigm helps, and will continue to help, understanding important urban phenomena. We conclude that the monocentric paradigm is far from obsolete, and we expect that it will provide the stepping stone for the construction of a robust polycentric city model.

8.1 The Alleged Failure of the EDF

The entire first part of our book deals with a monocentric city model in which every spatial characteristic can be expressed as a function of a single variable—the distance to a unique geographical point referred to as *the centre*. This simple specification has proven to be very useful in clarifying concepts vaguely perceived in the past, and in deriving many results and testable hypotheses. Even more importantly, it has become the theoretical cornerstone upon which more realistic models were built. Yet, from the very beginning, Alonso (1960) and Muth (1961) were criticised for the unrealistic assumptions underlying the monocentric paradigm. In fact, an alternative model which was supposed to do without these unrealistic assumptions, was proposed at about the same time by Herbert and Stevens (1960) and was considered by many regional scientists as the appropriate answer to the unrealistic approach of the monocentric city model. The efforts to modify the monocentric paradigm continued throughout the 1970s, and were culminated in the breakthrough papers of Ogawa and Fujita (1980), Fujita and Ogawa (1982) and Imai (1982).[1] Thus while some urban economists were refining the monocentric city model and deriving new robust implications, some others made important progress in developing new models for the polycentric city.[2]

Criticisms about the monocentric city model have grown in recent years, and they actually involve some of the people who have used the model extensively in the past. Such growing dissatisfaction stems to a significant extent from the feeling that the EDF, which was believed to be a robust implication of the monocentric paradigm, does no longer perform well in representing the density pattern of many North American cities. Similar reservations extend to various intertemporal and intercity comparisons around the world. Mieszkowski and Smith (1991) argue that, in the case of Houston, decreasing population density may reflect an increasing proportion of vacant land away from the CBD, rather than a true density decline over developed residential areas. This corroborates an earlier finding of Harrison and Kain (1974) on Denver, who also criticise the EDF. The intertemporal and intercity comparisons look even more disturbing in some respects. Edmonston (1975) found that the density gradient for small urban areas is considerably larger on average than the gradient for large urban areas. Similar findings were reported in Mills and Tan (1980). This empirically observed *negative* correlation between urban population size and the density gradient is at odds with the comparative statics of section 3.2.1, which imply a *positive* correlation.

The failure of the EDF to support some theoretical predictions can be attributed in part to the crude manner in which the density gradient and the explanatory variables are calculated in some studies. Another reason may arise from the complex types of interdependence among key elements of the monocentric city model. For example, we know from chapter seven that the transportation cost depends on traffic density which, in turn, depends on the infrastructure and on the number of commuters in each location. The number of commuters itself depends on both total population and the way housing is spatially distributed. Such extensive interdependencies among the explanatory variables (income, transportation, population) can create significant estimation

[1] Early studies on the polycentric city include Papageorgiou (1970), Hartwick and Hartwick (1974), von Boventer (1976), Romanos (1977), White (1977), Landsberger and Lidgi (1978) and Odland (1978).

[2] In light of this evidence, the indiscriminate criticism of Krugman (1995, p. 58) against the 'new urban economists', namely, that their monocentric paradigm "... became increasingly inadequate because the real world decided to play a nasty trick on the modelers, by abolishing the monocentric city as a reasonable approximation" seems unjustified.

and interpretation problems. However, in what follows, we skip over these difficulties and concentrate on more theoretical issues.

8.2 The Monocentric City and the EDF

Since the EDF is only a small part of the monocentric paradigm, and one which is derived from the paradigm only after using several restrictive assumptions, the reader may ask why should a failure of the EDF to support some theoretical predictions be considered to imply a failure of the monocentric paradigm itself—as Mieszkowski and Mills (1993) claim. This point is important and it needs to be further clarified.

A theoretical justification of the exponential rent and density functions was originally provided by Muth (1961), and later elaborated upon by Brueckner (1982) and Papageorgiou and Pines (1989). As we have seen in section 2.3.4, this explanation requires unitary price elasticity of the compensated demand for housing, linear transportation costs and an homogeneous population. Recently Anas, Arnott and Small (forthcoming) have derived the EDF using a Cobb–Douglas utility function. Although their approach avoids the assumption of unitary price elasticity, it replaces linear transport cost with a peculiar alternative. Under their specification, the transportation rate tends to zero as distance increases beyond all bounds. Thus the transportation cost approaches a finite upper bound asymptotically, which is problematic. But so is the assumption of linear transportation cost since, in reality, it increases at a decreasing rate as the traffic density declines with distance from the CBD.[3] Furthermore, in both derivations of the EDF, population homogeneity is both unrealistic and consequential for EDF estimations. For if the population is heterogeneous and spatially segregated, the density gradient should change with location. If, further, income and/or family size increase with distance from the centre, the gradient itself decreases with distance. As we have seen in chapter five, the assumption of an homogeneous population is not necessary for the monocentric city model—although it has proven necessary for the available derivations of the EDF. We conclude that, in any case, restrictive assumptions are indeed required for deriving the EDF from the monocentric city model so that, in principle, a failure of the EDF cannot be considered to imply a failure of the monocentric paradigm. We shall strengthen this conclusion subsequently.

The failure of the EDF can also be attributed to specific characteristics of the monocentric city paradigm. The model requires homogeneous land distinguishable only by distance from the centre. In particular, the commuting cost associated with any given residential location is uniquely determined by the

[3]Mieszkowski and Mills (1993, p. 138) associate the monocentric paradigm with linear transport cost, although neither Alonso (1960) nor Muth (1961) used exclusively this specification (even Mills (1967) himself did not). Most of the results derived in the first part of our book depend on a decreasing transportation rate, which also appears as an implication of the monocentric city model in result 7.1.

distance between that location and the centre. This implies that the individual can move radially from any point in space at a cost which depends only on distance from the center. This, in turn, requires either that all the urban area is one homogeneous road or that circumferential transportation is costless. Both assumptions are highly unrealistic. Since transportation infrastructure in the real world is composed by an heterogeneous layout of roads, the resulting loci of iso–transportation cost are not circular. The radially shaped developments along the commuter lines in Chicago offer a good example for this observation. Another example is provided by a beltway used by some of the traffic that originates near it and is destined for the CBD. Consequently some locations beyond the beltway may be more accessible to the CBD than some more central locations. It follows that the loci of iso–transportation cost do not coincide with the corresponding iso–distance loci. If indeed accessibility matters, this distortion should be reflected by a corresponding distortion in land use and population density. Such distortion may disappear only under a sufficiently coarse resolution.

One way to overcome this difficulty and still maintain the basic properties of the monocentric city model is to adopt the transformation suggested by Arnott and Stiglitz (1981), which is based on the *transportation cost shape of the city* $\widetilde{\Theta}\left[T,\bar{\alpha},\bar{\beta}\right]$ and described in section 2.1.1. This transformation allows us to translate the problem of spatially irregular iso–transportation cost to one of spatially regular iso–transportation cost in a non–circular city. Since under this transformation T enters as a variable, all spatial distributions are expressed in terms of transportation cost—rather than distance from the centre. Moreover the transportation cost, as a variable, satisfies the linearity requirement imposed in section 2.3.4 for the negative exponential specification. Thus if the price elasticity of the compensated demand for housing is unitary, we can follow the argument in section 2.3.4 to arrive at the new exponential rent and density functions

$$\begin{aligned} \tilde{R}^e[T] &= \tilde{R}^e\langle 0\rangle \exp\left(-\bar{\delta}T\right) \quad (a) \\ \tilde{D}^e[T] &= \tilde{D}^e\langle 0\rangle \exp\left(-\bar{\delta}T\right) \quad (b) \end{aligned} \tag{8.1}$$

which are expressed in terms of transportation cost rather than distance from the centre. Most other properties of the monocentric city model, now defined in terms of transportation cost, are preserved. The estimation results can be translated back on the Euclidean plane to produce three–dimensional, irregular rent and density surfaces which express reality in a much better way than the rotationally symmetric surfaces of the standard model.

In summary, the monocentric city paradigm should not be identified with the EDF. The derivation of the latter requires restrictive assumptions beyond those inherent to the monocentric city model itself. Some of the difficulties arising from the latter can be overcome by an appropriate transformation—provided that the transportation cost shape $\widetilde{\Theta}$ is empirically determined for the city under consideration.

8.3 The Density Gradient and City Size

It is widely believed that the most important problem with the EDF is its failure to predict that larger cities are, on average, more suburbanised (see Edmonston (1975) and Mills and Tan (1980)). As we know from sections 3.2.1 and 3.4.1, the effect of city size on the density gradient is positive: a population increase causes both a larger urban area and a higher population density everywhere. Generally speaking, this information is not sufficient to decide whether or not the model predicts a negative correlation between city size and the density gradient. However, for the EDF, the comparative statics of chapter three imply that an increased population size reduces equilibrium utility which, in turn, steepens the spatial distribution of population density.[4] This unequivocal implication has led both Mieszkowski and Mills (1993) and Anas, Arnott and Small (forthcoming) to conclude in their literature reviews that the EDF clearly fails. We believe that this conclusion is wrong, and that it stems from an unnecessarily narrow interpretation of the monocentric city model and the resulting EDF. We shall argue that there is nothing wrong with the EDF in this respect, and that the problem rather arises from failing to determine correctly the expected sign of the population effect on the density gradient *within the monocentric city model itself*. When this model is properly specified, the density gradient need no longer exhibit a positive correlation with city size.

The comparative statics of chapter three are based on constant returns to scale in production, and so is the specification of Anas, Arnott and Small. Thus the crucial property of declining utility with increasing city size, which is the source of the implied positive correlation between city size and the density gradient, applies to cities under constant returns to scale in production. What about cities under increasing returns? In section 5.2 we discussed the case where production exhibits scale economies which are external to the individual firm (see Dixit (1973) and Henderson (1986)). As we have explained in that section, the equilibrium utility level in those cities can well be an increasing function of population size.[5] But then it follows immediately that the gradient can *decrease* with increasing population size. Furthermore big cities differ from small cities in the richer variety of products they offer, and this allows for a higher utility level in big cities which is not reflected by income. One could say that this represents an appeal outside of the monocentric model. But it is not so. For the monocentric model admits any specification of productive activities that take place at the centre. An example is provided by Tabuchi (forthcoming) who discusses a monocentric city model with differentiated goods produced at the centre under scale economies, and where utility can increase with city size.

---

[4] As we know from section 2.3.4, the density gradient is equal to $(\bar{\alpha} f\,[U^e, \bar{\gamma}])^{-1}$, where $f\,[\cdot] > 0$ and $\partial f / \partial \dot{U} > 0$. Thus a utility decrease implies an increased gradient. Anas, Arnott and Small (forthcoming, p. 20) report that their specification too implies a 'mild positive correlation' between city size and the density gradient.

[5] See also a more thorough elaboration of this issue in chapters ten through twelve.

8.4 Is the Monocentric Paradigm Obsolete?

Empirical evidence indicates that the CBD is not yet obsolete. Even in Los Angeles, the extreme case of a polycentric city, a dominant centre still exists and it affects variables in the region more than any other centre. Proximity to the main centre in Los Angeles still exerts the strongest effect on office–commercial property values relative to all other subcentres in the Los Angeles region (Sivitanidou (1996)). Still, land values and employment densities peak at the main centre. Under both coarse and fine resolutions, these two variables produce spikes well above the corresponding spikes produced by any other subcentres in the Los Angeles region (see Anas, Arnott and Small (forthcoming)). Using data from Giuliano and Small (1991), we know that 10 percent of total employment is concentrated in the main centre of Los Angeles which occupies only .6 percent of the regional area. We also calculate that nearly one third of the aggregate regional employment occurs in the cluster that includes the main centre and the three closest–by subcentres which, together, occupy only 1.1 percent of the regional area. If we use a coarse resolution and combine the four centres along with their surrounding land into a single entity, we obtain more than one third of the aggregate regional employment in less than 3 percent of the regional area. And if an employment concentration of over 35 percent in 2 percent of a metropolitan area seems a good justification for using the monocentric paradigm, even Los Angeles almost passes the test in sufficiently coarse resolution!

Of course Los Angeles is unique, and the above evaluation applies with much greater force to the rest of North American big cities. In most of them land values, structural density, employment density and congestion peak at the CBD; and the area around the CBD, which contains a large proportion of poorly maintained structures, is mainly inhabited by low–income and minority groups. All these characteristics are readily explained by the monocentric city model. We conclude that the tendency to discard the monocentric paradigm on the basis of the declining rôle of the CBD seems unjustified.

The CBD concentration remains strong enough to allow for a drastically simplified empirical description by the monocentric EDF of the complex, polycentric rent and density surfaces that unfold away from the CBD. Fitting a monocentric EDF on polycentric city data has been standard practice since the early days of Clark (1951). Such estimations work in an empirical sense because the monocentric EDF captures to a significant extent the *general trend* of urban rent and density surfaces away from the CBD of a polycentric city. This observation can be substantiated in more precise terms on a theoretical level. As we have seen in section 5.4, the monocentric EDF can be extended to a corresponding polycentric city model which is *exogenous* in the sense that it applies to any given distribution of centres. Consequently, using this polycentric EDF, it is possible to estimate rent and density surfaces of a given metropolitan area at a fine level of resolution, taking fully into account the salient characteristics and the locations of centres in the area. Significantly, the monocentric and the corresponding polycentric EDF can be related to each other in an explicit manner. Namely, under some empirically justifiable conditions about the spacing of centres briefly discussed in section 5.4.3, the monocentric EDF can be derived from its polycentric counterpart by a smoothing procedure.[6] Under these conditions, the Clark formula describes the general trend of population density and land values in a polycentric city. This provides another justification for the standard EDF and the monocentric city model.

In a sort of requiem to the monocentric paradigm, Mieszkowski and Mills (1993, p. 144) claim that "[a]lthough the monocentric model and the exponential density function have been valuable in understanding and documenting past trends in urban decentralization, the rise of "Edge City" makes the model and function increasingly irrelevant." This statement is noteworthy since it comes from one of the founding fathers of the monocentric paradigm. A little further down, on the same page, Mieszkowski and Mills state that "[m]uch evidence and analysis indicate that MSA size, income levels and distribution, transportation evolution and housing demand are important in understanding MSA structure and decentralization." Since they do not present any alternative evidence, we assume that the evidence they refer to is precisely that which they derive from using the EDF. Furthermore, in supporting the hypothesis that "...central city racial mix and suburban land use controls interact to help explain both the extent and pattern of suburbanization in U.S. MSAs" (ibid. p. 144), they use studies on the density gradient (Mills and Price (1984)) and inter–country comparisons (Mills and Ohta (1976), Glickman (1979) and Goldberg and Mercer (1986)). However, gradient estimates in all those studies are based on the assumption that the EDF is sound. Thus discarding the monocentric paradigm as irrelevant is inconsistent with their claim that it explains suburbanisation to a significant extent. Finally, the rise of 'edge cities' on the fringe of an expanding metropolitan area, which Mieszkowski and Mills invoke as a reason for the growing irrelevance of the monocentric paradigm, either represents new subcentres added to the existing polycentric metropolitan framework, or new independent cities which should not be included in the monocentric EDF estimates. In the first case such new developments will, of course, reduce the rôle of the CBD. However, as we have already argued above, we believe that the declining rôle of the CBD does not seem to be a good reason for discarding the monocentric paradigm itself.

Our defence of the monocentric paradigm does not imply that it is good enough for explaining all economic aspects of urban structure and its evolution. On the contrary, we believe that an *endogenous* polycentric model capable of yielding testable hypotheses is a high–priority research challenge. One important reason in support of this statement is that some of the comparative statics available from the monocentric city model do not extend to a polycentric framework. For example, we know from sections 3.2.3 and 3.4.6 that an increase in commuting cost makes the monocentric city more compact. However, in the

[6] This procedure is outlined in Papageorgiou (1971)). A more detailed description is available from YYP upon request.

endogenous polycentric model of Fujita and Ogawa (1982), this is true only up to a certain level of commuting cost. Beyond this level the city is dispersed and we observe the emergence of new, perhaps smaller centres. Thus, in contrast to the monocentric city, the effect of commuting cost is no longer monotonic in the more general case of an endogenous polycentric city. But it is not true that all monocentric comparative statics no longer apply. For example, in the short run where centres are fixed, the effect of transportation on residential land rent at the CBD is the same for both the monocentric and the polycentric city.

In spite of all its shortcomings, we concur with Anas, Arnott and Small (forthcoming, p. 13) that the monocentric paradigm "... provides a rigorous framework for analyzing the spatial aspects of the general–equilibrium adjustments that take place in cities, and for empirically measuring and comparing the degree of centralization across cities and time periods." Furthermore, we are convinced that the monocentric paradigm is far from obsolete: it will continue to play an important rôle for urban economics in the foreseeable future. The monocentric city model provides clear and unequivocal results concerning the relationships among transportation cost, transportation crowding, congestion and land values in an urban setting. It allows for the derivation of robust comparative statics, that is, results which do not depend on specific functions, some of which are not confined to the monocentric characteristics of the model. We present below a short list of characteristic examples in order to substantiate our assertions.

1. Only in the early 1960s it became known that the cost–benefit approach which was routinely used to evaluate the construction of new roads involved double–counting. The simple monocentric model of Mohring (1961) was good enough to identify this serious problem. The relationship between transportation cost and land rent was further clarified both on the disaggregate level by Alonso (1964) and Muth (1969), and on the aggregate level by Arnott and Stiglitz (1979) and Arnott, Pines and Sadka (1986).

2. In the early 1970s, the second generation of monocentric city models elaborated in more detail the concepts of transportation crowding, congestion, allocation of land to roads and the like. New theoretical developments clarified the ambiguity associated with the shadow value of land and its relation to market rent across locations. The issue, which was originally raised by Solow (1973), was eventually settled by Kanemoto (1977) and Arnott (1979). It has significant implications on cost–benefit analyses of allocating land to roads, and it illustrates once again the advantage of the simple monocentric paradigm when dealing with very complicated issues.

3. Wheaton's (1974) comparative statics about the effect of transportation cost on land rent and, especially, his surprising result about the effect of income on central population density and land rent, provide the most convincing explanation for the suburbanisation process and for the deterioration of the housing stock around the CBD of large cities. The quality

deterioration around the CBD was explicitly derived by Arnott, Davidson and Pines (1986) in their study about the spatial aspects of housing maintenance.. Taken together, these two papers complement each other to produce significant insight on one of the most important problems that characterise many MSAs in the US.

The purpose of these selective illustrations was to convince the reader that the monocentric paradigm, due to its simplicity, is an efficient tool for clarifying concepts and for deriving some robust results at a coarse resolution. But as we have already stated, developing a robust polycentric model capable of yielding testable hypotheses is a most important research challenge to–day. We anticipate that, as in the past, the stepping stone for the construction of a polycentric city model will be provided by the monocentric paradigm itself.

8.5 References

W. Alonso, 1960, A Theory of the Urban Land Market. *Papers and Proceedings of the Regional Science Association* 6, 149–158.

W. Alonso, 1964, *Location and Land Use*. Harvard University Press, Cambridge, Mass.

A. Anas, R. J. Arnott and K. A. Small, forthcoming, Urban Spatial Structure. *Journal of Economic Literature.*

R. J. Arnott, 1979, Unpriced Transportation Congestion. *Journal of Economic Theory* 21, 294–314.

R. J. Arnott, R. Davidson and D. Pines, 1986, Spatial Aspects of Housing Quality, Density and Maintenance. *Journal of Urban Economics* 19, 190–217.

R. J. Arnott, D. Pines and E. Sadka, 1986, The Effects of Equiproportional Transport Improvement in a Closed Monocentric City. *Regional Science and Urban Economics* 16, 387–406.

R. J. Arnott and J. E. Stiglitz, 1979, Aggregate Land Rents, Expenditure on Public Goods, and Optimal City Size. *Quarterly Journal of Economics* 93, 471–500.

J. K. Brueckner, 1982, A Note on Sufficient Conditions for Negative Exponential Population Density. *Journal of Regional Science* 22, 353–359.

C. Clark, 1951, Urban Population Densities. *Journal of the Royal Statistical Society* 114 (Series A), 490–496.

A. K. Dixit, 1973, The Optimum Factory Town. *Bell Journal of Economics and Management Science* 4, 637–651.

B. Edmonston, 1975, *Population Distribution in American Cities.* Heath and Company, Lexington Mass.

M. Fujita and H. Ogawa, 1982, Multiple Equilibria and Structural Transition of Non–Monocentric Urban Configurations. *Regional Science and Urban Economics* 12, 161–196.

G. Giuliano and K. A. Small, 1991, Subcenters in the Los Angeles Region. *Regional Science and Urban Economics* 21, 163–182.

N. Glickman, 1979, *The Growth and Management of the Japanese Urban System.* Academic Press, New York.

M. Goldberg and J. Mercer, 1986, *The Myth of the North American City.* The University of British Columbia Press, Vancouver.

D. Harrison and J. Kain, 1974, Cumulative Urban Growth and Urban Density Functions. *Journal of Urban Economics* 1, 61–98.

P. Hartwick and J. Hartwick, 1974, Efficient Resource Allocation in a Multinucleated City with Intermediate Goods. *Quarterly Journal of Economics* 88, 340–352.

J. V. Henderson, 1986, Efficiency of Resources Usage and City Size. *Journal of Urban Economics* 19, 47–70.

J. D. Herbert and B. H. Stevens, 1960, A Model for the Distribution of Residential Activity in Urban Areas. *Journal of Regional Science* 2, 21–36.

H. Imai, 1982, CBD Hypothesis and the Economics of Agglomeration. *Journal of Economic Theory* 28, 275–299.

Y. Kanemoto, 1977, Cost–Benefit Analysis and the Second–Best Land Use in Transportation. *Journal of Urban Economics* 4, 483–507.

P. Krugman, 1995, *Development, Geography and Economic Theory.* Cambridge University Press, Cambridge.

M. Landsberger and S. Lidgi, 1978, Equilibrium Rent Differentials in a Multicenter City. *Regional Science and Urban Economics* 8, 105–116.

P. Mieszkowski and E. S. Mills, 1993, The Causes of Metropolitan Suburbanization. *Journal of Economic Perspectives* 7, 135–147.

P. Mieszkowski and B. Smith, 1991, Analyzing Urban Decentralization: The Case of Houston. *Regional Science and Urban Economics* 21, 183–200.

E. S. Mills, 1967, An Aggregative Model of Resource Allocation in a Metropolitan Area. *American Economic Review* 57, 197–210.

E. S. Mills and K. Ohta, 1976, Urbanization and Urban Problem. In H. Patrick and H. Rosovsky (eds.), *Asia's New Giant: How the Japanese Economy Works.* The Brookings Institution, Washington D.C., 673–751.

E. S. Mills and R. Price, 1984, Metropolitan Suburbanization and Central City Problems. *Journal of Urban Economics* 15, 1–17.

E. S. Mills and J. P. Tan, 1980, A Comparison of Urban Population Density Functions in Developed and Developing Countries. *Urban Studies* 17, 313–321.

H. Mohring, 1961, Land Values and the Measurement of Highway Benefits. *Journal of Political Economy* 69, 236–249.

R. F. Muth, 1961, The Spatial Structure of the Housing Market, *Papers and Proceedings of the Regional Science Association* 7, 207–220.

R. F. Muth, 1969, *Cities and Housing.* The University of Chicago Press, Chicago.

J. Odland, 1978, The Conditions for Multi–Center Cities. *Economic Geography* 54, 234–244.

H. Ogawa and M. Fujita, 1980, Equilibrium Land Use Pattern in a Non–Monocentric City. *Journal of Regional Science* 20, 455–475.

Y. Y. Papageorgiou, 1970, *A Generalization of the Population Density Gradient Concept*. Ph.D. Thesis, The Ohio State University, Columbus, Ohio.

Y. Y. Papageorgiou, 1971, The Population Density and Rent Distribution Models Within a Multicentre Framework. *Environment and Planning* 3, 267–282.

Y. Y. Papageorgiou and D. Pines, 1989, The Exponential Density Function: First Principles, Comparative Statics and Empirical Evidence. *Journal of Urban Economics* 26, 264–268.

M. C. Romanos, 1977, Household Location in a Linear Multi–Center Metropolitan Area. *Regional Science and Urban Economics* 7, 233–250.

R. Sivitanidou, 1996, Do Office–Commercial Firms Value Access to Service Employment Centers? A Hedonic Value Analysis Within Polycentric Los Angeles. *Journal of Urban Economics* 40, 125–149.

R. M. Solow, 1973, Congestion Cost and the Use of Land for Streets. *Bell Journal of Economics and Management Science* 4, 602–618.

T. Tabuchi, forthcoming, Urban Agglomeration and Dispersion: A Synthesis of Alonso and Krugman. *Journal of Urban Economics.*

E. von Boventer, 1976, Transportation Costs, Accessibility, and Agglomeration Economies: Centers, Subcenters, and Metropolitan Structures. *Papers of the Regional Science Association* 37, 167–183.

W. C. Wheaton, 1974, A Comparative Statics Analysis of Urban Spatial Structure. *Journal of Economic Theory* 9, 223–237.

M. J. White, 1977, Firm Suburbanization and Urban Subcenters. *Journal of Urban Economics* 3, 323–343.

Part II

Polycentric Urban Structures

9
Agglomeration

This chapter discusses the basic ideas and reviews the literature on agglomeration. Some important recent dynamic studies of agglomeration are not included here because they are outside the scope of our book.[1] We also provide background for the remaining chapters of part two. In this sense, this chapter serves as an introduction to the second part of our book.

9.1 The Benefits and Costs of Agglomeration

Economic activities tend to concentrate and to form contiguous areas where the intensity of land use (the ratio between other factors and land) is high. On a *regional* level these concentrations represent villages, towns, cities, metropolitan and megalopolitan areas. On a *local* level they represent production, commercial and service centres of different size, from the marketplace of a small village to the CBD of a large city, around which residential areas develop and spread. The characteristics of these concentrations and their distribution over space vary from place to place and from period to period.[2] However, in any case, such concentrations arise mainly because economic agents derive utility from inter-

---

[1] See, for example, Fujita and Mori (1997*b*), Eaton and Eckstein (1997), Krugman (1998) and Fujita, Krugman and Mori (forthcoming).

[2] It is well known, for example, that towns first appeared only around six thousand years ago in Mesopotamia, and that cities in Europe all but disappeared during the middle ages (Davis (1965)). During the twentieth century, accelerating urbanisation caused a dramatic increase in the proportion of urban people on a regional level (agglomeration), while it reduced the importance of the core in many big cities on a local level (deglomeration).

acting with each other either directly or indirectly. The spatial characteristics of human settlement are set by the benefit individuals derive from their interactions, the extent to which these interactions are costly and the negative effects of crowding. In this chapter we elaborate on the interplay among these three factors and how it determines the extent and the character of agglomeration. We confine our discussion to *urban models* dealing with this issue. Accordingly, we do not discuss the literature on the Hotelling (1929) problem, as well as many other pertinent agglomeration studies.[3]

Agents interact with each other in various ways. They exchange tangibles and intangibles in the form of goods, services and ideas. Individuals derive utility from interacting with each other on a personal level or by taking part in collective activities. Likewise firms derive surplus as users of labour and sellers of consumer goods. Firms also derive surplus by trading intermediate goods with each other and by exchanging information.[4] The scope of all these interactions determines the extent of specialisation and, consequently, the level of social performance in its economic and cultural aspects.

When interaction involves the exchange of goods and services, goods change hands through shipment from the warehouse of the seller to that of the buyer, while services are rendered when a service recipient or a service provider move toward each other. The same holds for many competitive or cooperative interactions, whether social, cultural, religious or other. Individuals are gathered in one place to watch a sport event, a concert, or to participate in a religious ceremony.[5]

Since the amount of resources spent for interaction increases with distance between those who interact, agglomeration is one important way to save on such costs. Thus, firms trading intermediate goods are attracted to each other because they want to save on the shipping cost. Firms and households are attracted to each other both as traders in final products and as traders in labour services. The same is true between households and facilities which provide public goods. Observe that, if we interpret transportation costs in a sufficiently broad sense, *saving on these costs can provide the sole motivation for the global tendency of people to concentrate.* For if there were no transportation costs there would be no pressing need for the proliferation of large plants where workers, foremen, stocks of raw material as well as intermediate and final goods are assembled

[3] These include Baesemann (1977), Stahl and Varaiya (1978), Papageorgiou (1979), Stuart (1979), Wolinsky (1981) and Shaked (1982) among others.

[4] The latter plays an important rôle in theoretical discussions about agglomeration. For general discussions on information exchange and agglomeration see Marshall (1890), Lösch (1940) and Jacobs (1969). Information exchange is also used as a conceptual foundation for modelling agglomeration in recent studies as, for example, in Abdel-Rahman and Fujita (1988) and Fujita (1989).

[5] Of course modern technology also allows for interactions which do not require transportation. Watching a football game or a theater performance on the television does not require one's presence in New York, Paris or London—just in the family room. Communication however is not only a substitute for transportation: by fostering the need to interact, it can also become a complement.

close to each other during the production process. Intermediate goods, tools, machines and the like could be shipped to the workers rather than the other way around.[6] Likewise, all necessary interactions could be carried out costlessly at any location—including the original location of interacting parties.

Whereas we emphasize the rôle of transportation cost as the reason why people agglomerate, Fujita and Thisse (1996) emphasize indivisibility.[7] Following Koopmans (1957), they assert that indivisibility is a *necessary* condition for agglomeration.[8] Nevertheless agglomeration can emerge even under perfect divisibility. Beckmann (1976) for example, as well as Boruchov and Hochman (1977), provide early models in which atomistic agents, who benefit from their proximity to each other, agglomerate. Therefore indivisibility is not necessary for modelling agglomeration.

In general, the benefit of agglomeration can be represented by the increase in the consumer surplus it allows as a result of the lower interaction costs. In terms of figure 9.1, the direct benefit is represented by the area ABCD and the further increase in the consumer surplus by the area CDE. Hence the overall benefit is the area ABED. This additional benefit is discussed in Marshall (1890) and elaborated in Chinitz (1961). Marshall, for example, discusses the diffusion of skills propagated by the "...near neighbourhood to one–another..." which operates in a way that "...mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously." (op. cit. p. 271.) Such agglomeration advantages underlie both what is referred to as *localisation economies* and *urbanisation economies*.[9] The former are associated with the agglomeration of similar enterprises, while the latter with the agglomeration of different ones.[10].

---

[6] Of course, in the presence of strong indivisibilities such as those represented by steel mills and oil refineries, it would still make good economic sense to produce at the plant site. Workers, however, could live at any distance from the plant.

[7] Indivisibility is often at the root of scale economies. Up to a certain point, one set of dies can produce any quantity of pressed metal sheets. Consequently, up to a certain point, the larger is the quantity produced the lower becomes the cost of initial investment per unit of pressed metal sheets. Likewise, one has to spend several years in training to become a doctor or a professional football player. It is therefore difficult to become a doctor and a professional football player at the same time.

[8] Koopmans (1957, pp.153-154) states: "...If we imagine all land to be of the same quality, both agricultural and in amount and accessibility of mineral resources, then an activity analysis of production that includes the proportionality postulate [*divisibility*, in our terms] would show a perfectly even distribution of activities to be most economical. Each square of the area would produce the same bundle of commodities and all transportation would thus be avoided!... This suggests that without recognizing indivisibilities—in the human person, in residence, plants, equipment, and in transportation—urban locations, down to those of the smallest villages, cannot be understood."

[9] For a discussion of differences between the two types see Isard (1956). For different empirical evaluations of their relative importance see Henderson (1988) and Glaezer, Kallal, Scheinkman, and Sleifer (1992).

[10] Some specific urbanisation economies are discussed by both Marshall (1890) and Lösch (1940) as, for example, the diversification of employment and the resulting decrease in unemployment risk. Suppose that the cycles of different industries are uncorrelated. If the industrial

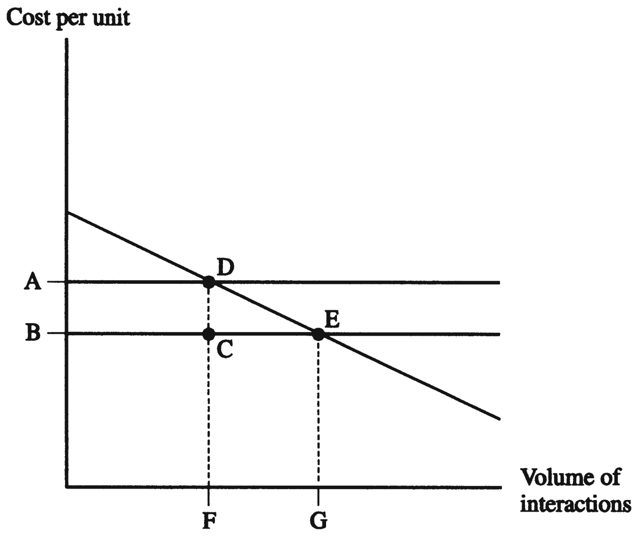

FIGURE 9.1. Direct and Indirect Benefits of Agglomeration.

Agglomeration is not only a consequence of direct attraction between the interacting parties, but of indirect attraction as well. One obvious case is represented by the *vertical* indirect type of interaction, where A sells to B who, in turn, sells to C thus inducing a convergence of A to C. Another case, which is often referred to both in the early literature and in recent models of agglomeration, is the *horizontal* indirect type of interaction, where B interacts with both A and C. Although A and C themselves do not interact, they are indirectly converging to each other because both converge to B independently of each other. These indirect agglomeration effects are important in the discussions of Marshall (1890) and Lösch (1940).[11] A modern formulation of this horizontal indirect type of interaction plays a key rôle in the agglomeration models that followed Krugman (1991). In these studies, which are based on the monopolistic competition model of Dixit and Stiglitz (1977), firms are directly attracted to the same set of customers and, therefore, to each other because customers buy from every firm. The same reason for agglomeration appears in the next chapter of our book.

---

composition of each city's employment is homogeneous, a recession in one industry creates unemployment in one place and labour shortage in another. Expensive relocation costs are required to move the unemployed where they are needed, only to relocate them back during the next cycle. Dumais, Ellison, and Glaezer (1997) found that this reason for agglomeration is statistically more important than the alternative reasons discussed by Marshall.

[11]The accessibility to skilled labour and specialised machinery is one of the main causes Marshall (1890, pp.271–272) suggests for the concentration of industries in certain regions of Europe. He also discusses the advantage of locating textile industries, which employ women, close to the mining and manufacturing industries which employ men. Also, both Marshall (1890, p.273) and Lösch (1940, p.69) suggest that sellers of differentiated products agglomerate to reduce search costs of their customers.

We now turn from the benefits to the costs of agglomeration. These arise because most economic activities occupy space, and their performance is improved with the amount of space they occupy. Sometimes relevant space requirements are expressed in terms of floor area, so that both space abundance and agglomeration can be achieved simultaneously through high structural density. But in many cases the land area matters. Even when the available land area can be augmented by applying higher structural density, there comes a point beyond which the marginal cost of floor space increases with structural density.[12] Hence higher levels of agglomeration imply either less space available per unit of economic activity, or higher cost of supplying a given amount of space, or both. The *cost of crowding* refers to this disadvantage of agglomeration.

In summary, as we mentioned at the beginning of this section, urban spatial structure arises from a fundamental trade–off between the advantage of agglomeration, which augments agents' welfare because it saves on the cost of interaction, and the disadvantage of agglomeration, which reduces agents' welfare because it increases crowding. However, since the interaction flow itself depends on the cost of interaction as illustrated in figure 9.1, we can express the fundamental trade–off more precisely as follows. On the one hand agglomeration implies reducing the cost of interaction, increasing its volume and more crowding. On the other, dispersion implies less crowding, increasing the cost of interaction and reducing its volume.

This trade–off determines agglomeration structure at both the local (intra–city) and regional (inter–city) levels. Take, for example, urban size. Consider a particular agglomeration characterised initially by a given flow of interactions and the benefit they render, a given volume of interaction costs and a given level of crowding. Let population increase. For a given per–unit cost of interaction, such an increase must generate a corresponding higher volume. Unfortunately this desirable effect cannot be realised unless crowding, which is undesirable, increases as well. If though crowding is to remain unchanged, the increased population must be accommodated within a larger urban area, which increases the average distance among agents. Since the cost of interaction increases, the consumer surplus derived from interaction declines. These opposing population effects imply that the feasible level of utility depends on population size and that there may exist some limit to the net benefit of agglomeration. More specifically, as both Lösch (1940, pp. 75–76) and Tiebout (1956) note, there may exist some optimal city size.[13]

The same trade–off determines how various non–residential activities are distributed within the city. Starting with an initial monocentric agglomeration of non–residential activities located at the city centre, consider any relocation of some of these to a non–central location. This may allow more accessibility

[12] Otherwise there would be much higher office buildings than those existing to–day. Even in cities like New York, Tokyo and Hong–Kong all new office buildings are less than one hundred and twenty floors.

[13] For Lösch, an optimal size must be specific to the level of the settlement in the central place hierarchy. For Tiebout, it depends on the type of the population in the community.

for a part of the population with respect to the relocated activities. Moreover, for the part of the population which gained accessibility to the relocated non–residential activities, the marginal benefit of accessibility may decline so that those individuals now tend to substitute less crowding for accessibility, thus increasing the area of the city. Under these conditions relocation both increases consumer surplus and reduces crowding. On the negative side, the same relocation decreases accessibility for another part of the population and for the relocated non–residential activities. We must therefore weigh the benefits against the costs of relocation in order to determine whether or not splitting the city centre in two is desirable.

Suppose that splitting is desirable. An interesting question is when does it signify the emergence of a new centre in the same city or the creation of a separate city. Turning back to our initial state and to the subsequent relocation of some non–residential activities, recall that the increased per–unit cost of interaction reduces the flows between those activities for which interaction has become more expensive. If these flows become insignificant then a new city does emerge. If, on the other hand, their volume remains substantial, splitting creates a new centre in the same city. Of course this statement is not very informative, and more structure is needed in order to understand the conditions leading to either one of the two outcomes. Some of these issues are analysed in the following chapter ten.

9.2 Modelling Agglomeration

In this section we present an overview of formal modelling on the spatial convergence of economic activities. We begin with von Thünen (1826) and the monocentric–city models, then proceed to the ways diverse centripetal and centrifugal forces shape the internal structure of cites, the city size distribution and, finally, the economic geography of cities.

9.2.1 Von Thünen and the Monocentric City

The Isolated State

According to von Thünen's (1826) Isolated State, there exists a marketplace defined as a point in geographical area which is surrounded by agricultural rings. Each ring is devoted to the cultivation of a specific crop associated with the highest bid rent over the ring, where the bid rent is defined as the revenue minus the cost of labour and the cost of transportation—all evaluated per unit of land. In maximising net output, which is equivalent to maximising aggregate land rent, the cultivated rings agglomerate into a compact agricultural area. The radius of the agglomeration that surrounds a single marketplace extends to the point where the bid rent of the marginal crop vanishes.

In this formulation it remains unclear why does the agricultural production agglomerate around a single market instead of dispersing around as many markets as possible. The latter possibility saves on transporting the crops to the market, hence on delivery prices to the consumers. Only recently Fujita and Krugman (1995) have provided an answer to this question by specifying in a precise way certain aspects of the Isolated State which have been ignored for a long time.[14] Using the monopolistic competition approach of Dixit and Stiglitz (1977) to explain why the production of manufactured goods agglomerates and to account for the trade of manufactured for agricultural goods, they define conditions under which a single city does emerge amidst an agricultural hinterland and forms with the latter a closed region. Inside this region, the fundamental centripetal force is provided by the cost of transporting the manufactured good, while the fundamental centrifugal force is provided by the need for land and labour which arises from the demand for the agricultural goods. This derived demand disperses part of the population over the agricultural area.

The Monocentric City

The original concept of the monocentric city, as developed by Alonso (1960) and Muth (1961*b*), is similar to von Thünen's (1826) framework in that the centre is exogenously defined as a point in geographical space and is either the origin or the destination of all trips and shipments. Both approaches are partial: whereas von Thünen focuses on the spatial structure of the hinterland and ignores the marketplace, the monocentric city model ignores the hinterland and focuses on the spatial structure of the marketplace. In that sense the monocentric city model can be taken as a partial complement of von Thünen's framework with respect to the Isolated State. Thus while in von Thünen the centre is the location of all non–agricultural activities, in the monocentric city model it accommodates all non–residential activities.[15] Individuals live around the centre and their residential crowding decreases with distance from it. Once again, as in the case of von Thünen, one wonders why does the population agglomerate in such cities where equilibrium utility is a decreasing function of population size (see section 3.2.1).[16] In other words, why an infinite number of "atomistic" cities does not replace the inefficient concentration of people around some geographical point.[17]

---

[14] Fujita and Krugman emphasize that, contrary to what has been written many times, the marketplace in von Thünen's formulation *trades with the hinterland manufactured goods for agricultural goods.* Thus a fully specified model reflecting the ideas of von Thünen must solve for the output of the manufactured goods in the city, as well as for the spatial distributions of population, wage rate, price and consumption of the manufactured goods.

[15] The earliest model that analyses simultaneously the two sectors has been proposed by Muth (1961*a*) who assumed a circular area of housing surrounded by a single agricultural ring.

[16] This objection applies to the basic monocentric city model alone, where income per capita is exogenous. When urban production is explicitly introduced, as in section 5.2, the utility–population profile is no longer necessarily monotonic. In that case the size of an optimal agglomeration will correspond to the highest feasible equilibrium utility.

[17] In Alonso (1964), the convergence of non–residential activities around the centre is explained along von Thünen's lines. Namely, the bid–rent functions of these activities are steeper

A step further is made by Dixit (1973) who assumes that the production of the composite good is characterised by scale economies. While this approach accounts for agglomerations on a city–wide level, it does not explain why the production of the composite good takes place at the centre of the monocentric city.[18] A similar criticism can be applied to a number of early models in the mainstream of urban economics which attempted to relax strict monocentricity while, at the same time, assuming the existence of a main centre—rather than explaining why it exists. The first such extension can be found in Muth's (1969) book, where some of the workplaces are scattered over the residential rings of the monocentric city. Under these circumstances individuals work either at the centre or at their place of residence. Competition among workers and among employers makes the equilibrium wage rate decline with distance from the city centre. This approach was further developed by Solow (1973*b*), in whose model the CBD is occupied only by firms which export their output from a node located there.[19] Firms which serve the city population occupy parts of the residential rings. The formation of the CBD is justified by the existence of the export node combined with sufficiently high shipping costs of the export goods to the node (relative to the commuting cost). Similarly, the existence of firms inside the residential rings is justified by the demand for their services combined with a sufficiently high cost of shopping travel. Solow's equilibrium conditions do not account for the possibility that some exporting firms may be attracted to peripheral locations by the lower wage rate. This motivation plays a key rôle in White (1976) who calculates the bid rent of an exporting firm on the basis of spatially variable shipping cost and wage rate. In this model exporting firms and individuals repel each other because of their competition for land while, at the same time, they attract each other in order to save on transport cost. Combining these two opposing effects likely produces alternating employment and residential rings.[20]

---

than the residential bids. But there is no explanation as to why this happens, and why employment and residences combine to agglomerate.

[18] The same reservation equally applies to an alternative formulation of the monocentric city suggested by Arnott (1979) and Starrett (1988), who replace production at the centre with a central public facility visited by all individuals in the city. If the public–good facility is not spatially indivisible, one wonders once again why is the supply of the public good not dispersed over the urban area.

[19] The same idea of an export node is used in Henderson (1986).

[20] The discussion above refers to the first part of White's paper. In the rest of the paper more than one export nodes do exist and, without interactions among firms or individuals, it is not clear why the city is not decomposed into as many agglomerations as there are export nodes. The same criticism applies to Vieand (1988) who analyses the formation of subcentres in more detail. See also the modelling of agglomeration factors in Sasaki and Mun (1996) and in Fujita, Thisse and Zenou (1997).

9.2.2 *Internal City Structure*

One basic characteristic of the models we discussed in section 9.2.1 is the *exogenous* character of the city centre. Here we discuss models in which the centre or centres emerge *endogenously.* To the best of our knowledge, the first such model was built by Herbert and Stevens (1960) with intent to provide a detailed account of how transportation affects urban land use and vice–versa. The outcome was a discrete–space urban planning model which can generate endogenously multiple centres. Although this operational model was further developed and reinterpreted by Harris (1966) and by Harris, Nathanson and Rosenberg (1966) six years after its publication, ten more years elapsed before its potential applications to theory development became recognised in the literature.[21]

The first example of an endogenous agglomeration as a pure trade–off between accessibility and crowding has been modelled by Beckmann (1976) for the simplest case of a single type of interacting agents. Beckmann obtained a bell–shaped equilibrium density of population in a linear city *without a predetermined centre.* In this model, spatial interaction between agents over a bounded landscape creates an advantage for central locations which are associated with lower interaction cost. Competition for land eliminates this advantage through agglomeration. This model was extended to a circular city by Boruchov and Hochman (1977). Later on, Papageorgiou and Thisse (1985) and Fujita (1988) introduced a second type of agent in Beckmann's linear and bounded landscape, and obtained the bell–shaped equilibrium density configuration when the pattern of spatial interaction between households and firms is sufficiently dispersed.[22]

The main contribution of Ogawa and Fujita (1980), Fujita and Ogawa (1982) and Imai (1982) was to allow for interactions among firms and to specify how these, in conjunction with commuting cost (required for interactions between individuals and firms), determine whether the city will be *monocentric* or *polycentric.* The modelling of interaction costs and benefits in these papers draws from the seminal work of Solow and Vickrey (1971) which develops an explicit framework for interactions among firms. Whereas Solow and Vickrey deal with the allocation of land between production and roads, Fujita, Ogawa and Imai model the allocation of land between production and housing. They adjust the framework of Solow and Vickrey to consider interactions not only among firms (which choose to interact with every other firm irrespectively of cost) but also employer–employee interactions.

Two versions of interaction costs are considered in these models: a constant rate $\bar{a}$ per unit of distance x between the firms (Ogawa and Fujita (1980) and Imai (1982)), and a diminishing rate $\bar{b}\,(1 - \exp(-\bar{a}x))$ where $\bar{a}$ now represents

[21] See Wheaton (1974) and Fujita (1989).

[22] The spatial interaction effects between households and firms in these two models are determined on the entire distribution of both agent types, which can be justified in the case of product differentiation. Papageorgiou and Thisse use general formulations and introduce the cost of interaction in a reduced form. Fujita uses specific formulations and derives a similar reduced form by modelling explicitly market interactions.

the decay rate in the benefit of interaction (Fujita and Ogawa (1982)). Both versions imply that, given $\bar{a}$, a monocentric city cannot be sustained for sufficiently high commuting cost. For, in that case, it is preferable to save on the high cost of interaction between individuals and firms by creating a polycentric city—even though the moderate cost of interaction among firms does increase. More surprising is the implication found by Fujita and Ogawa (1982) that, with the second version, the monocentric city cannot be sustained *both* when $\bar{a}$ is sufficiently low or sufficiently high. The reason can be understood by inspecting the structure of $\bar{b}\,(1 - \exp(-\bar{a}x))$. If $\bar{a}$ vanishes, so does the cost of interaction between firms so that distance between interacting firms is no longer relevant. Since only the commuting cost matters, the firms will completely disperse over the residential locations. Likewise, if $\bar{a}$ becomes arbitrarily large, the cost of interaction between any two firms tends to $\bar{b}$ thus becoming independent of the distance between them. Then once again, only the commuting cost matters. We shall discuss this issue further in chapter ten, where we present a similar result based on the locational endogeneity of all centres.[23]

For alternative values of the commuting rate and $\bar{a}$, Fujita and Ogawa (1982) derived the monocentric city and various polycentric structures within the same modelling framework.[24] The importance of this path–breaking contribution to the understanding of how and why firms agglomerate within cities cannot be exaggerated. However, there still remains the issue of indirect agglomeration forces which may perhaps be no less important than the direct ones. As we explained in section 9.1, firms can be attracted to each other even though they do not interact directly. In particular, suppliers of different services to the same individuals may locate close to their client—hence close to each other. Fujita (1989) examines the case where employment is found at the centre of the city and the locations of two public facilities are endogenously determined. The question is under what circumstance will the public facilities locate at the centre and when will they diverge. In particular, the model investigates the relative importance of the demand for housing, the demand for public services, and the rate at which the level of service deteriorates with distance between the residential and public facility locations. It turns out that only when the share of housing is sufficiently important relative to that of the public good and, at the same time, the 'rate of decay' in public–facility service deteriorates fast with distance, will the public facilities diverge from each other. When the share of the public good is high relative to housing and the 'rate of decay' is low, the facilities converge to the centre. This result can be clarified by taking into account that high demand for housing has a similar effect to low commuting cost. Thus the

[23] Mori (1997) obtains the same result in a different context.

[24] In particular, they derive (i) a monocentric city; (ii) a city with mixed land use in the central area surrounded by two pure business districts which, in turn, are surrounded by exclusively residential areas; (iii) a city where land uses are completely integrated everywhere; (iv) a symmetric duocentric city; (v) a CBD and two symmetric subcentres bordering residential districts on both sides; and (vi) three subcentres, each bordering residential districts on both sides.

possibility of divergence increases with the 'rate of decay' and decreases with both commuting cost and the demand for the public good. Fujita explains the latter effect as follows. The combined quantity of the public good available at the centre increases by concentrating the facilities there, which becomes more important the higher is the share of the public good.

In the above studies, the attraction between firms has been modelled to reflect the rôle of information exchange to agglomeration. This is well–recognised as an important reason why firms benefit from their interaction (see also Marshall (1890), Lösch (1940) and Jacobs (1969)). Other important reasons why firms agglomerate include their input–output relationships and saving on the cost of search for households—thereby becoming more attractive to them (Marshall (1890, p. 173)). Both input–output relationships among firms and the tendency of firms to agglomerate in order to reduce the cost of search for households have been taken into account by Anas and Kim (1996). In their model individuals distribute their purchases over all firms, and the relative utility they derive from the good purchased in a particular store increases with the volume of sales in the area. Anas and Kim investigate the implications of scale economies on firms. As expected, in the absence of scale economies and since every individual interacts with every firm, land uses are mixed and their intensity declines with distance from the centre. If the scale economies become large enough then a compact CBD emerges which, once again, appeals to intuition.

9.2.3 Distribution of City Sizes

In the previous section we reviewed models of agglomeration within a given city. In the present and in the next section we describe models of agglomeration concerned with how cities are formed and how is the population distributed among them. We shall organise the presentation of these models chronologically into *first–*, *second–*, and *third–generation* studies. The main questions to be asked in this inter–city context are:

1. What are the economic forces which affect the creation of cities and the distribution of population among them? In particular, what is the effect of transportation cost on these?

2. Do market forces tend to induce excessive concentration in cities or excessive dispersion among cities?

3. Why do cities specialise in their economic activities?

The *first generation* of studies about systems of cities developed within the framework of central place theory.[25] The origins of this theory go as far back as Cantillon (1755) who demonstrated a clear grasp of settlement hierarchies, from hamlets to towns to cities, and how all these central places combine to serve their hinterland. The cornerstone of central place theory rests on the concept of

[25] A central place is a cluster of producers.

a functional interdependence between city and countryside, which defines the nature of economic activities to be found in central places. Since these activities (central place functions) exist to serve the rural population of the hinterland, they do not include export manufacturing oriented toward more distant markets. The hinterland is perfectly homogeneous with respect to its geography and the rural population evenly distributed over it. Starting with this ideal landscape, Christaller (1933) proceeded to construct his monumental framework for the study of city systems. He observed that each central place function can be associated with a specific *threshold* and a specific *range* which, because of homogeneity, will apply to every producer of the function anywhere he or she produces it.[26] All functions, in turn, can be organised into homogeneous groups with respect to their threshold and range, so that within–group variation is small relative to between–group variation. Thus groups can be ordered according to the size of their representative threshold and range, from smallest (lowest order functions) to largest (highest order functions), which leads to the notion of a central place hierarchy. This settlement system develops over a regular lattice of evenly spaced locations, each at the centre of an hexagonal market area that contains the threshold population of the lowest order group. The lattice and its spacing have been justified by homogeneity, the condition that everyone in the system must have access to all goods, transport cost minimisation and free entry which, through competition, drives profit to zero—hence market area to its threshold.

Taking into account that larger central places in the real world typically contain almost everything that smaller places can offer but not vice–versa, Christaller (1993) proposed a central place hierarchy in which all settlements of a particular order contain all lower groups of central functions in addition to the group that characterises their order. Thus every settlement of a particular order must be identical to every other settlement of the same order. Lowest–order settlements contain only the lowest–order group of functions, while a highest–order settlement contains all central place functions. The geometry of the regular lattice and Christaller's requirement that higher–order places contain all lower–order goods impose strict constrains on how can a central place system be organised over the landscape. In particular, the size of every market area of any order must be a given number of times larger than the market area of the immediately lower order; and this given number of times is not arbitrary, but assumes only certain integer values, the *characteristic values*, which form a sequence beginning with $3, 4, 7$ and so on. It follows that a single integer determines the entire spatial structure of a central place system up to an arbitrary origin and orientation. Furthermore, given some assumptions about employment demand for the various central place functions, it also determines how the entire urban population is distributed among the cities of the region (Beckmann (1958)). It is possible to compare these theoretical city size distributions

[26] Threshold refers to the minimum level of demand required to support the production of the good. Range refers to the maximum distance over which a producer can sell that good.

against corresponding empirical regularities such as the rank–size rule;[27] and even though Christaller's central place theory predicts a stepwise distribution of city sizes, which is at odds with the continuous distribution approximated by the rank–size rule, Beckmann maintains that the two may well be compatible because centres of the same order will not have the exact same population in reality. As a result, the steps predicted by the theory will be smoothed out to approximate a continuous distribution of city sizes.

The rigidity imposed on a central place system by the requirement that all settlements of a particular order contain all lower order groups makes the consistency of the entire scheme problematic. For it is only by coincidence that the threshold populations at any two consecutive levels will be expected to yield the same characteristic value. And if they do not, it is unclear why the excess profits generated by some larger–than–threshold market areas do not affect the neat conceptual scheme of Christaller. Rigidity was relaxed to a considerable extent by Lösch (1940), who combined several characteristic values into the same central place system. The basic implication is that central places need no longer contain all lower order functions in addition to those that characterise their order, but different combinations which allow settlements of a similar size to produce different bundles of goods. In this manner Lösch allowed for specialisation within a central place system, which was impossible under Christaller's scheme. And the distribution of city sizes as predicted by his theory comes closer to a continuous distribution without the need to invoke random error terms. Nevertheless, even behind this more sophisticated and flexible version of central place theory, spatial arrangement and city size distributions are still determined by the geometry of the regular lattice and by the principle that central places exist only to serve their hinterland. But if we allow for export industries, the distribution of city sizes must change because the population size of some cities will increase to accommodate those who produce export goods; and the same will happen if we recognise the various agglomeration advantages enjoyed by firms, which were described in the previous section. In any case it is unclear how a central place system can be supported as an equilibrium because there is no underlying mechanism that ensures the agglomeration of firms required by central place theory. The need to serve all the uniformly distributed individuals with the maximum possible number of central places does not seem sufficient. Production locations for different goods could still be arranged on their regular lattices over the region, competition could still drive their market areas to their threshold size, but it is not obvious why will these lattices coincide at various locations to generate central places that provide various combinations of goods. Lattices could indeed unfold over the land without constraint imposed by the position of other lattices. Therefore, as in the agricultural land use theory and the monocentric city model of section 9.2.1, central place theory offers no explicit reason as to why different central place functions agglomerate

[27] In its simplest case, the rank–size rule says that the population of a given city in a region times its rank equals the population of the largest city in that region (Zipf (1949)).

rather than disperse. Finally, another obvious criticism of central place theory is that it rests on the functional interdependence between city and countryside. In to–day's increasingly urban world this functional interdependence has lost much of its force, and we can no longer argue convincingly that it affects locational patterns the way it did in the past. But the great conceptual scheme of Christaller, about the spatial organisation of economic activity in a hierarchy of centres, can still be useful in a polycentric urban context as we saw in chapter eight.

We now turn our attention to more recent studies. These are mainly concerned about the interplay between the fundamental centripetal and centrifugal forces which shape the way productive activities and populations are distributed among cities, as well as about their operation through the market mechanism. We can classify these studies in two groups, *second–generation* and *third–generation* models, which differ from one–another with respect to:

A. The fundamental centripetal and centrifugal forces that affect city structure and their impact on the market;
B. The explicit representation of the spatial and geographical aspects of their allocation, where geography here means distances between any pair of cities; and
C. Their positive or normative viewpoint.

Regarding aspect A, second–generation models use a reduced form of those centripetal and centrifugal forces in order to study how they shape the relationship between utility and city size, and in order to specify the implications of this relationship on the distribution of economic activities among cities. By contrast, third–generation models start from the fundamental mechanisms that generate the centripetal and centrifugal forces, and then examine the effects of these on the city system. Regarding aspect B, second–generation models abstract from geography either by assuming that cities are autarchic or that trade among cities is costless. By contrast, in third–generation models, inter–city trade and the cost of shipping goods (as a function of distance between the cities involved) are explicitly taken into account. However, although geography matters, it remains exogenous in the earlier studies. Only is some more recent studies the spacing between agglomerations is endogenously determined. Finally, regarding aspect C, second–generation models are (with some exceptions) normative, although the question of sustainability in the market is also important. By contrast third–generation models are positive, and the specific market structure plays a crucial rôle in shaping the city system. In the rest of this section we discuss only second–generation models. Third generation is discussed in the following section that deals with the economic geography of cities.

Common to the *second generation* models is the $\cap$–shape configuration between utility and city size. This configuration reflects the dominance of scale economies over diseconomies when the city population is small, and the dominance of scale diseconomies over economies when the city population is large. The source of these economies differ from one model to another. On the one hand, scale economies in some models stem from the production of a composite good and are external to the firm, while scale diseconomies stem for residential housing (see, for example, Dixit (1973), Henderson (1974) and Anas (1992)). On the other hand, in local public good and club theory models, the source of scale economies is the cost–sharing advantage of the collective good, while scale diseconomies reflect either residential crowding and transportation costs (Arnott (1979)), or the congestion in consuming the collective good (Berglas (1976) and Scotchmer and Wooders (1987)), or the diminishing marginal productivity on a fixed amount of land (Stiglitz (1977) and Wilson (1987)).

Whatever the source of the ∩–shape configuration, with perfect replicability, efficient allocations in the tradition of Tiebout (1956) emerge under competitive markets. Their long–run equilibrium is either a system of identical cities, each of optimal size, or a system of different groups of identical cities, where cities belonging to the same group accommodate a distinct socioeconomic type and specialise in the production of a distinct private good. However, perfect replicability is relevant only if the total population size is enough to justify a large number of cities and if the population is an integer multiple of optimal population size or sizes. Thus, in many cases, perfect replicability cannot apply. This explains a tendency for restricting the analysis to a fixed number of cities or, as an illustration, to only two. One such example is provided by Stiglitz (1977), who examines a wide variety of configurations between utility and city size and illustrates cases where the market fails to distribute efficiently the population among cities. In all cases where stable equilibria are inefficient, the bias is unidirectional: markets induce excessive concentration.[28] More precisely, stable equilibria may correspond to either partial or full concentration, but full dispersion is Pareto–superior to both. The opposite possibility, that markets induce insufficient concentration, does not appear in Stiglitz. Nevertheless, as we shall see in chapter twelve, excessive concentration does not necessarily hold under different specifications of scale economies and crowding effects.

The standard distinction between intra– and inter–city levels is somewhat artificial. For example, checking the configurations derived by Ogawa and Fujita (1980), we find that the same forces can disperse economic activities to create either a subcentre in the same city or an entirely new city. As we shall see in the next section, even more complicated spatial systems emerge within the paradigm of the new economic geography, where the distinction between intra– and inter–city levels is thoroughly blurred.

A theoretical framework that integrates the intra– and inter–city levels of agglomeration has been suggested by Hochman, Pines and Thisse (1995). We elaborate on this framework in chapter ten. Similarly to the monopolistic com-

[28] A similar issue is discussed by Anas (1992) where, once again, the land constraint provides the source of crowding. Crowding is related to the residential use of land while, in contrast to Stiglitz, the production of the private good exhibits scale economies. Anas displays some of the configurations between utility and city size proposed by Stiglitz and concludes that the market tends to induce excessive concentration—at least during some phases of population growth.

petition framework described above, the agglomeration advantage in this study also stems from a horizontal indirect attraction between collective facilities which are simultaneously patronised by the same individuals. Firms now can be thought of as special cases of a public facility, where the collective good represents non–labour inputs. At the intra–city level, the internal structure of the city (including the location of individuals, the location of public facilities and the market area of each facility) is determined simultaneously as a polycentric urban structure. The effect of transport cost to a given public facility on the city structure is examined, and the emergence of a polycentric city for either sufficiently large or sufficiently small transport cost is obtained as in Fujita and Ogawa (1982) and Mori (1997). At the inter–city level, the urban system is determined along the lines of club theory as an integer number of economically independent cities. However, this system need not be homogeneous because cities can specialise in the production of private goods. The reasons and the characteristics of specialisation, which were first examined by Wilson (1987), are discussed in chapter eleven.

9.2.4 Economic Geography of Cities

In section 9.2.3 we discussed central place theory in the context of city size distribution. The other main function of this theory is to provide an understanding of spatial settlement patterns as they are represented by the distance between any pair of cities. We have already noticed that the theory suffers from several shortcomings. For example, the various thresholds are assumed to have certain strict relationships to one–another and there is no agglomeration advantage proposed to explain why each higher–order central place must include central places of lower order. More importantly, the fundamental dispersion force in central place theory is the supply of evenly distributed farmers with manufacturing goods. To–day, when the importance of agriculture in value added is well below ten percent in most developed countries, this is an exaggeration. It seems therefore preferable to shift the emphasis from agriculture, and to recognise that the fundamental dispersion force in the economic geography of cities is represented by the quest for open space in congested urban environments.

Fundamentals of Agglomeration

Consider a population of identical individuals distributed over a line segment. Individuals interact with each other, and they derive utility from their interaction as in Beckmann (1976). Locations with a higher *potential to interact* are associated with lower interaction cost per unit of interaction—hence they have an advantage over locations with a lower potential to interact.[29] The utility level of individuals is given in a reduced form as a function of a local crowding

[29] The potential to interact associated with a particular location is a measure of accessibility for that location. Since the potential to interact for an individual located at A with an individual located at B decreases with distance between A and B and increases with the population density at B, the overall potential to interact for an individual at A depends

effect (represented by the density of population in situ) and of a global interaction effect (represented by the corresponding potential to interact). Start with a uniform distribution of individuals over the line segment. Crowding is the same for everyone, but individuals in central locations enjoy a higher potential to interact—hence a higher level of utility. In equilibrium the fundamental trade–off between accessibility and crowding yields an equal–utility agglomeration. Notice how the combination of a bounded space and spatial interaction draws the individuals toward the centre. The line segment here represents a *non–homogeneous* landscape akin to those created by the models of sections 9.2.2 and 9.2.3.

Now bend the linear landscape into a circle. Since boundaries vanish the space becomes *homogeneous*, so that locations lose their identity in terms of relative position and become indistinguishable from all other locations. In particular, every location is associated with the same distribution of distances to all other locations, so that the uniform distribution of individuals over the circle is an equilibrium. It follows that spatial homogeneity disperses individuals even when they need to interact. Conversely, no need for spatial interaction among individuals also disperses individuals even if the space is non–homogeneous.

The example suggests that both spatial interaction among individuals and a bounded space are needed for agglomeration. This differs from central place theory, where centrality arises in conjunction with a settlement pattern over the boundless, perfectly homogeneous landscape: whereas centrality determines agglomeration in the models of sections 9.2.2 and 9.2.3, agglomeration defines centrality in the classic landscapes of Christaller (1933) and Lösch (1940). Now if the initial population distribution is uniform, *how can agglomerations emerge over an homogeneous landscape*? Since the uniform distribution is an equilibrium, agglomerations can emerge only if it becomes unstable. And since, in this context, the only reason for agglomeration is the propensity of individuals to interact with each other, the question is about the conditions under which the propensity to interact induces an instability of the equilibrium. This is the central question in Papageorgiou and Smith (1983). They find that the stability characteristics of the uniform equilibrium at a particular location depend on (1) the marginal effect of increased congestion on the utility of individuals caused by a population increase there and (2) the corresponding marginal effect of an increased potential to interact. The former effect is negative, while the latter is positive since spatial interaction is desirable. The uniform distribution is a locally stable equilibrium if and only if the marginal effect of increased congestion dominates over the marginal effect of an increased potential to interact.[30] When the positive effect of spatial interaction outweighs the negative effect of congestion, the uniform population distribution becomes locally unstable. Under these

---

on the entire spatial distribution of the population and the corresponding distribution of distances from A to every other location.

[30] Papageorgiou and Smith also show that the uniform equilibrium is locally stable if and only if an arbitrarily small perturbation lowers total utility. Taken together, these two results indicate that agglomeration begins when it "should".

circumstances, if the locally unstable equilibrium is perturbed, agglomeration will ensue.

Although Papageorgiou and Smith (1983) lay down the precise conditions for the transition from a spatially uniform equilibrium to an agglomerated one, they say nothing about the nature of the agglomerated equilibrium. Once the spatially uniform equilibrium is locally unstable, *what are the characteristics of the emerging settlement pattern*? In particular, what is the spacing of locations which, at the critical instability, define up to an arbitrary origin the only candidates for agglomeration? De Palma and Papageorgiou (1991*a*) study this problem on a linear and boundless landscape. Using a different method, they obtain the same necessary and sufficient condition for local instability. However, for spatial interactions obeying a negative exponential rate of decrease over distance, they find that the spacing of locations where the uniform population distribution first becomes unstable is infinitely long. This implies that the emerging settlement pattern is degenerate. The same conclusion holds over alternative specifications of spatial interaction. Therefore it seems that *spatial interaction* alone among individuals of a *single type* over a *perfectly homogeneous* landscape is not enough to produce a regular settlement pattern. The implication is that, if settlement systems can indeed arise endogenously over an homogeneous landscape, there must be some other fundamental reason for agglomeration which is absent from Papageorgiou and Smith (1983). A natural possibility to consider is the specialisation of economic activity, which is basic to the models we discussed in sections 9.2.2 and 9.2.3, and the importance of which seems well–established in the historical literature about early societies and city formation.[31]

The simplest such case, involving spatial interaction between two types (say, firms and households) over an homogeneous landscape has been examined by de Palma and Papageorgiou (1991*b*). The potential of an agent to interact now involves both own–type and cross–type interactions. For analytical tractability, de Palma and Papageorgiou assume that the local congestion effect is no longer determined by local density alone, but rather by the entire potential to interact: high potential to interact also implies high local congestion and vice–versa. Since utility is a function of the potential to interact alone, the balance between crowding and accessibility, which explicitly appears in Papageorgiou and Smith (1983) and in de Palma and Papageorgiou (1991*a*), is replaced by a balance between the negative and positive aspects of spatial interaction. Thus the transition between local stability and local instability here occurs when the net value of the positive and negative marginal effects (caused by an increased potential to interact) changes sign. It is seen that the departure from the spatially uniform equilibrium can take the form of initial growth on a regularly spaced pattern of locations, in which case a settlement system emerges endogenously. For negative exponential rates of spatial interaction, *emergence* depends both on the preferences and on the characteristics of spatial interaction among

[31] See, for example, Mumford (1961) and Parsons (1977).

agents. By contrast, the *spacing* of settlements depends on the characteristics of spatial interaction alone.

New Economic Geography

The 'new economic geography' represents the most challenging item on the research agenda since the seminal paper of Krugman (1991). These are the *third generation* models we referred to in section 9.2.3. The basic approach underlying the new economic geography models makes use of a modern monopolistic competition framework developed by Dixit and Stiglitz (1977), and introduce the cost of trade between regions (cities)—thus recognising that geography matters. This approach provides a micro–foundation to the scale economies associated with population size. Although it was originally applied to scale economies in production (Abdel–Rahman and Fujita (1988) and Fujita (1989)), its application to scale economies in consumption has dominated the field since Krugman (1991) (see, for example, Fujita and Krugman (1995), Fujita and Mori (1997*a*,*b*), Hadar (1997), Mori (1997), Helpman (1998) and Tabuchi (forthcoming)). We shall discuss only models that explain scale economies of the latter type.

The new economic geography models are based on the following five basic assumptions:

1. In terms of our fundamental agglomeration concept, each individual interacts not only with one firm as an employee or, perhaps, with several firms as a consumer, but with every manufacturing firm in the economy wherever it is located.
2. Due to indivisibility the production function of each firm exhibits scale economies.
3. The importation of goods from other cities involves a real cost in the form of 'melting iceberg', that is, the quantity delivered in the importing city is smaller than the quantity loaded in the exporting city.
4. The manufacturing goods are differentiated products (brands).
5. The market structure is monopolistic competition. Each firm equates its marginal cost to its marginal revenue and, due to free entry, its profits are driven to zero.

Underlying all the models included in this group is some common structure based on Dixit and Stiglitz (1977). Formally, the utility function is given by

$$U = u\left[Z_A, Z_M\right] \tag{9.1}$$

where Z_A is the amount of an homogeneous *non–manufactured good* consumed by an individual and Z_M is the corresponding amount of a complex of *manufactured goods* defined as

$$Z_M = \left(\int_0^{\bar{m}} y^{\bar{\alpha}}\left[\varrho\right] d\varrho\right)^{\bar{\alpha}} \quad \text{for } 0 < \bar{\alpha} < 1 \tag{9.2}$$

where $\bar{m}$ is the available *number of brands* (manufacturing firms) and $y[\varrho]$ is the *quantity consumed* of each brand ϱ.

It takes the firm $\bar{\beta} + X$ units of labour to produce X units of its brand and, with the wage rate equal to $\bar{\omega}$, the brand costs $\left(\bar{\beta} + X\right)\bar{\omega}$. Hence the marginal cost of producing a brand is $\bar{\omega}$ and the average cost is $\left(\bar{\beta}/X + 1\right)\bar{\omega}$. The shipment of a brand from one city to another costs a fixed proportion of the quantity shipped per unit of distance travelled. If the proportion of the shipped commodity which 'melts' is $\bar{\gamma}$ per unit of distance, the proportion of the shipment that survives the 'melting' at distance x from the exporting firm is $\exp(-\bar{\gamma}x)$, the proportion that 'melts' is equal to $1-\exp(-\bar{\gamma}x)$ and the corresponding delivery price is $\bar{P}\cdot\exp(\bar{\gamma}x)$, where $\bar{P}$ is the *mill price* of the brand.

The marginal revenue of a brand sold in market j is $\bar{P}(1 - 1/|\eta_{ij}|)$, where η_{ij} is the *own–price elasticity* of a brand sold in market j by a producer located in city i. The above specification implies that $\eta_{ij} = \bar{\eta} = 1/(1 - \bar{\alpha})$ and that the marginal revenue corresponding to a given mill price is $\bar{\alpha}\bar{P}$. Since the marginal revenue of a firm does not depend on where it is located, there is no need for mill–price discrimination.[32] Furthermore since, under monopolistic competition, firms equate their marginal cost to their marginal revenue, we have $\bar{P} = \bar{\omega}/\bar{\alpha}$; and since free entry drives profit to zero, the average cost equals the price of the brand. These imply $X = \bar{\alpha}\bar{\beta}/(1 - \bar{\alpha})$. Thus the output of each firm depends only on the parameters. Finally, the labour inputs per brand are equal to $\bar{\beta}/(1 - \bar{\alpha})$ and the number of brands in a location with $\bar{M}$ workers is given by $\bar{m} = \bar{M}(1 - \bar{\alpha})/\bar{\beta}$.

The common structure underlying the models of the new economic geography and its implications corroborate some basic ideas of Jane Jacobs (1969) about cities. For example, the last result described in the previous paragraph implies that the relative range of brands produced in a particular place is proportional to its relative share in population. Moreover, an increase in the population share of a place does not cause a proportional increase in the production of existing 'works' (Jacobs' terminology for 'brands') there, but rather a corresponding increase in the range of works offered at this location. This is closely related to Jacobs' (1969, p.122) idea that adding new works is more important for growth than expanding the volume of existing works. Indeed, the formulations of new economic geography are consistent with this idea because they imply that, given the aggregate quantity produced, the consumption of manufactured goods at a particular location increases with the number of brands produced

[32] See Dixit and Stiglitz (1977) for an outline of the proof and for a reference to more detailed analyses of the relationship between the elasticity of substitution and the own–price elasticity. Equal elasticities of locational demands from the viewpoint of producers follow by the transport cost structure, which implies a fixed relative delivery–mill price ratio.

in that location.[33] For this and other reasons Jacobs can be considered as the prophet of the new economic geography.

The specification we described above is shared by a wide range of models which followed Krugman (1991). It implies a unique, *fundamental* agglomeration advantage: by concentrating the population in a few cities, instead of dispersing it among many, a larger proportion of differentiated goods are produced at a real cost of $1/\bar{\alpha}$ per unit rather than $\exp(\bar{\gamma}x)/\bar{\alpha}\,(> 1/\bar{\alpha})$—thus saving real resources for a given supply of differentiated products. This fundamental advantage is manifested in the *market* cumulatively. An increase of the urban population implies an increased demand for locally produced and an increase in the local supply of labour. These changes attract more firms and increase the local share of brands, which reduces the average cost of differentiated products for the local population (reduces the price index) and increases real income—thus enhancing utility. The increased utility, in turn, encourages further immigration.[34]

In contrast to the unique agglomeration advantage they explore, the reasons for dispersion underlying these models, as well as how utility is specified and how initial endowments are distributed, bring to mind a gallery of truly differentiated products! Krugman's (1991) dispersion force is provided by an immobile agricultural population who consumes the manufactured goods locally. The income derived from the non–manufactured (homogeneous) agricultural good is spent in the region where it has been generated. By contrast Helpman (1998) adopts an approach closer to urban economics. He defines an inter–city model without agriculture, where the force of dispersion is the residential crowding caused by a fixed supply of residential land in each city. Similarly Hadar (1997) and Tabuchi (forthcoming) use residential crowding as a dispersion force. But in their model crowding can be partly relieved by transportation.[35] Whereas in Krugman some of the population is immobile and both goods are mobile, in Helpman, Hadar and Tabuchi the population is mobile but housing supply is immobile.

It turns out that the main implications drawn from the above models crucially depend on their specification. Krugman (1991) and Helpman (1998) ar-

---

[33] Let the aggregate volume of the manufacturing good produced be equal to $\overline{\mathbb{V}}$, a constant. Then the optimal amount of the manufactures good consumed by an individual is given by

$$Z_M^* = \max_{y[\varrho]} \left(\int_0^{\bar{m}} y^{\bar{\alpha}}[\varrho]\, d\varrho \right)^{\bar{\alpha}} \text{ subject to } \int_0^{\bar{m}} y[\varrho]\, d\varrho = \overline{\mathbb{V}}.$$

This optimal amount increases with the number of works, that is,

$$\frac{dZ_M^*}{d\bar{m}} = \frac{1-\bar{\alpha}}{\bar{\alpha}} (y[\varrho]) \left(\int_0^{\bar{m}} y^{\bar{\alpha}}[\varrho]\, d\varrho \right)^{1/(\bar{\alpha}-1)} > 0$$

since $\bar{\alpha} < 1$.

[34] These effects are referred to as 'circular causation', where the attraction to individuals resulting from the increase in the number of firms represents the 'forward linkage' while the attraction to firms resulting from a larger demand represents the 'backward linkage'.

[35] Transportation allows for an increase in the effective land supply.

rive at quite different conclusions, while the results derived by Hadar (1997) and Tabuchi (forthcoming) are at odds with both Krugman and Helpman. In what follows we elaborate on this characteristic vulnerability to specification.

The model of Krugman (1991) implies that when the share of the differentiated product in consumption is large and the elasticity of substitution among brands is small, the agglomeration advantage dominates and a single city represents the only stable and efficient allocation. This holds for any transportation cost of the differentiated product. When both the above share and elasticity are small while the transportation cost is large enough, full dispersion replaces full agglomeration as the only stable and efficient equal–utility equilibrium. In the latter case, for intermediate values of transportation cost, both full dispersion and full agglomeration can persist as stable equilibria. Overall, if the degree of agglomeration is affected, it *declines* with increasing transportation cost.[36] An opposite effect is obtained by Helpman (1998) when the share of the differentiated product in expenditure is moderate and the elasticity of substitution among brands is sufficiently large. Under these circumstances low transportation cost leads to dispersion and high transportation cost leads to agglomeration.

In contrast to Krugman, Helpman's specification also allows for stable equilibria with unequal population distribution. It turns out that such equilibria are inefficient as they are Pareto–dominated by a full agglomeration. Helpman (1998, p.54) concludes that "... in economies of this type, if anything, free markets provide too little agglomeration." However using a CES (rather than Cobb–Douglas) utility specification, Hadar (1997) shows that the relationship between utility and population size can assume a $\cap$–shape. In this case, as in Stiglitz (1977) and Anas (1992), stable equilibria *may* lead to excessive agglomeration.[37] Finally Tabuchi (forthcoming), who extends Krugman's (1991) model by including a residential sector within each monocentric city, derives similar results. Like Hadar, he finds that the relationship between utility and population size can assume a $\cap$–shape or a 'wave' shape (formed by a combination of $\cup$ and $\cap$ shapes)—hence that the market incentive may lead to excessive agglomeration.

More recently Ottaviano and Thisse (1998) examine alternative pricing policies in a monopolistic competition framework that departs from Dixit and Stiglitz (1977) and approaches the paradigm of location theory. By using quite different specifications of utility, production and transportation cost in a model based on an immobile agricultural population, they show that Krugman's (1991) results are robust. In particular, contrary to Hadar (1997), Helpman (1998) and Tabuchi (forthcoming), they find that lower transport cost of the manufactured good is conducive to agglomeration—which corroborates Krugman. But, as we

[36] Some of these results are reported in Helpman (1998) who used simulations based on Krugman's (1991) specification. Helpman also provides a clear explanation of the reasons behind this seemingly counter-intuitive result.

[37] Excessive agglomeration in this case can be a stable full agglomeration which is Pareto–dominated by either a partial agglomeration or by a full dispersion, or a stable partial agglomeration which is Pareto–dominated by a full dispersion.

know, if the agriculture–oriented framework is abandoned in favour of another type of fundamental dispersion force then Krugman's results must change.

What conclusion can we draw from this brief comparison of agglomeration studies that explore alternative scale diseconomies? Firstly, we observe that the effect of transport cost on agglomeration is highly sensitive to the specification of the model. The policy implication of this finding is that, contrary to common belief, making the periphery more accessible need not encourage dispersion. Secondly, the direction of the bias induced by market forces is not clear. And although the latter observation has already been made by Mills and Hamilton (1994, p.403), the above models provide quite a different reasoning.

These models of the new economic geography incorporate costs which depend on the distances between trading cities. But inter–city distances remain exogenous, and so are the costs of trade between the potential cities. However, in some recent works that belong to the same paradigm, this important restriction is removed.[38]

Following Fujita and Krugman (1995), an excellent group of studies by Fujita and Mori (1997*a*) and Mori (1997) explore the economic factors that create the megalopolis. Relative to Krugman (1991), Fujita and Krugman (1995) make progress by introducing costly transportation of the agricultural good and by allowing farmers to migrate. These modifications change the character of the dispersion force in the model since they replace the immobility of farmers with the immobility of land as a factor in producing the agricultural good. Farmers now are no longer 'serfs' as in Krugman (1991). Both land and labour are required in fixed proportions to accommodate the demand for the agricultural product, which implies that some of the workers are dispersed outside of the city to produce the agricultural good. Those workers need to be provided with the manufactured good for, otherwise, they will not agree to toil the land. Thus the fundamental source of dispersion is the derived demand for labour in the periphery which may be reflected in two complementary market incentives. The first is associated with the demand price for the manufactured good in the hinterland. From the viewpoint of a single brand supplier, the demand price at a given location in the hinterland is higher than in the city.[39] The second can be a lower wage rate in the hinterland, which is guaranteed in Fujita and Krugman (1995) by assuming that the transportation cost of the agricultural good sufficiently exceeds that of the manufactured good.[40]

---

[38] As we have seen, an earlier study that generates the spacing of potential cities endogenously is by de Palma and Papageorgiou (1991*b*)..

[39] The vertical difference between the two reflects the transportation cost.

[40] One may wonder how can the wage rate be lower in the hinterland and, at the same time, because of transport cost, the delivered price of the manufactured good be higher there than in the city. For only when farmers can attain the same level of utility enjoyed by city residents will they be induced to move to the hinterland. This puzzle is resolved by taking into account the lower mill price of the agricultural good in the hinterland. Fujita and Krugman (1995) show that if the transportation cost of the agricultural good sufficiently exceeds that of the manufactured good, the utility levels in the city and in the hinterland can be equalised. See also Mori (1997).

Fujita and Krugman (1995) define *sufficient* conditions under which the agglomeration advantage dominates over the above two dispersion effects. Fujita and Mori (1997*a*) and Mori (1997) examine cases in which these conditions are not satisfied, so that either agglomeration or dispersion can prevail. As it turns out, a wide range of interesting configurations emerge—including megalopolitan configurations.

In what follows we describe the findings reported by Mori (1997), some of which draw from Fujita and Mori (1997*a*). Starting with a complete agglomeration that forms a single city, Mori considers a firm which moves to the hinterland and evaluates its potential profit function with respect to its new location. He distinguishes between two cases: one where the transport cost of the manufacturing good is sufficiently *small* relative to the transport cost of the agricultural good, and another where it is sufficiently *large*. In the first case, it is always profitable for the firm to move to any location in the hinterland, so that the complete agglomeration cannot be an equilibrium. In the second case there are two possibilities: (1) The brands of the manufactured good are poor substitutes and/or the population is small enough. Under these circumstances the complete agglomeration is an equilibrium. This happens because the relocating firm cannot benefit enough from the local substitution of its brand for the brands imported from the city. Even if the wage declines with distance from the city, the gains from its decline are not sufficient to compensate for the corresponding decrease in the demand of the urban population for the relocating firm's brand. (2) The brands of the manufactured good are close substitutes and the population is sufficiently large. Then there exists some range of the hinterland in which a relocating firm can make positive profits, so that the complete agglomeration is not an equilibrium.[41]

The next question addressed by Mori is about the kind of urban structure that emerges in these cases. Notice that, under the second possibility above, it is also true that if the transport cost of the manufactured good is *either* sufficiently large *or* sufficiently small relative to the transport cost of the agricultural good then the complete agglomeration is not an equilibrium.[42] Suppose we start from a complete agglomeration. On the one hand, increasing sufficiently the transport cost of the manufactured good yields a two–city equilibrium such that the potential profit over the interval between them is negative. In this case the two cities remain distinct. On the other hand, decreasing sufficiently the transport cost of the manufactured good yields a two–city equilibrium such that the interval between them is occupied by a continuous manufacturing belt where the density of workers is lower than the density in the two cities. In this case the

[41] This range however is not adjacent to the city, because the substitution of the relocating firm's brand for the brands produced in the city becomes negligible sufficiently close to the city. The population should be large enough to generate sufficient demand for the manufactured good in the hinterland.

[42] The same result is also obtained in Fujita and Ogawa (1982) and in our chapter ten below under quite different modelling frameworks.

two cities form the rudiments of a *megalopolitan* structure, first described by Gottman (1961) in an empirical context and elaborated by Doxiades (1968).[43]

These papers by Fujita and Mori provide us with considerable insight regarding the processes of urbanisation that characterise many countries around the world. Nevertheless one could still question their excessive dependence on agriculture, which completely overlooks the industrial and residential uses of land. Casual observation indicates that plants located along the main arteries connecting big metropolitan areas consume large tracts of land, and so does housing needed for their employees. These imply that urban crowding, which operates as a dispersion force *inside* the cities, is important relative to the alternative dispersion force generated by the agricultural pool as a market for manufactured products. It would therefore be interesting to verify whether and to what extent the implications of the agriculture–oriented model are robust to a modification that replaces agricultural consumption of land with urban housing consumption.

9.3 References

H. Abdel–Rahman and M. Fujita, 1988, Product Differentiation, Monopolistic Competition and City Size. *Regional Science and Urban Economics* 18, 69–86.

W. Alonso, 1960, A Theory of the Urban Land Market. *Papers and Proceedings of the Regional Science Association* 6, 149–158.

W. Alonso, 1964, *Location and Land Use*. Harvard University Press, Cambridge, Mass.

A. Anas, 1992, On the Birth and Growth of Cities: Laissez–Faire and Planning Compared. *Regional Science and Urban Economics* 22, 243–258.

A. Anas, R. J. Arnott and K. A. Small, forthcoming, Urban Spatial Structure. *Journal of Economic Literature.*

A. Anas and J. I. Kim, 1996, General Equilibrium Models of Polycentric Urban Land Use with Endogenous Congestion and Job Agglomeration. *Journal of Urban Economics* 40, 232–256.

R. J. Arnott, 1979, Optimal City Size in a Spatial Economy. *Journal of Urban Economics* 6, 65–89.

R. C. Baesemann, 1977, The Formation of Small Market Places in a Competitive Economic Process—The Dynamics of Agglomeration. *Econometrica* 45, 361–374.

M. J. Beckmann, 1958, City Hierarchies and the Distribution of City Size. *Economic Development and Cultural Change* 6, 243–248.

[43] These are enormous urban networks, each encompassing several major cities and the corridors along the main transportation lines connecting them. Such urban networks are developing in many parts of the world including Europe (for example Brussels–Frankfurt–Munich–Zürich, Glasgow–Edinburgh), North America (for example Washington–New York–Boston, Los Angeles–San Francisco, Windsor–Toronto–Montréal) and Asia (for example Tokyo–Osaka). They represent the fastest–growing type of human settlement, and they are expected to become dominant in the not–so–distant future (Doxiades (1968)).

M. J. Beckmann, 1976, Spatial Equilibrium in the Dispersed City. In Y. Y. Papageorgiou (ed.) *Mathematical Land–Use Theory.* D. C. Heath, Lexington, Mass., 117–125.

E. Berglas, 1976, On the Theory of Clubs. *American Economic Review* 66, 116–121.

E. Boruchov and O. Hochman, 1977, Optimum and Market Equilibrium in a Model of a City Without a Predetermined Center. *Environment and Planning A* 9, 849–856.

R. Cantillon, 1755, *Essay on the Nature of Trade in General.* Reprinted by Augustus Kelley, Bookseller, New York, 1964.

B. Chinitz, 1961, Contrasts in Agglomeration: New York and Pittsburgh. *American Economic Review, Papers and Proceedings* 51, 279–289.

W. Christaller, 1933, *Die Centralen Orte in Süddeutschland.* Translated in English as *Central Places in Southern Germany* by C. W. Baskin, Prentice–Hall, Englewood Cliffs, N.J., 1966.

K. Davis, 1965, The Urbanization of Human Population. *Scientific American* 213, 41–53.

A. de Palma and Y. Y. Papageorgiou, 1991*a*, Agglomeration as Local Instability of Spatially Uniform Steady–States: A Note on Degeneracy. *Working Paper* 91–08, Department of Economics, McMaster University.

A. de Palma and Y. Y. Papageorgiou, 1991*b*, Settlement Systems as Spontaneous Order. McMaster University, (unpublished).

A. K. Dixit, 1973, The Optimum Factory Town. *Bell Journal of Economics and Management Science* 4, 637–651.

A. K. Dixit and J. E. Stiglitz, 1977, Monopolistic Competition and Optimum Product Diversity. *American Economic Review* 67, 297–308.

C. A. Doxiades, 1968, *Ekistics: An Introduction to the Science of Human Settlements.* Hutchinson, London.

G. Dumais, G. Ellison, and E. L. Glaezer, 1997, Geographic Concentration as a Dynamic Process. Harvard University and NBER, (unpublished).

J. Eaton and Z. Eckstein, 1997, Cities and Growth: Theory and Evidence from France and Japan. *Regional Science and Urban Economics* 27, 443–474.

M. Fujita, 1988, A Monopolistic Competition Model of Spatial Agglomeration. *Regional Science and Urban Economics* 18, 87–124.

M. Fujita, 1989, *Urban Economic Theory.* Cambridge University Press, Cambridge.

M. Fujita and P. Krugman, 1995, When is the Economy Monocentric? von Thünen and Chamberlin Unified. *Regional Science and Urban Economics* 25, 525–528.

M. Fujita, P. Krugman and T. Mori, forthcoming, On the Evolution of Hierarchical Urban Systems. *European Economic Review.*

M. Fujita and T. Mori, 1997*a*, Transport Development and the Evolution of Economic Geography. Institute of Economic Research, Kyoto University, (unpublished).

M. Fujita and T. Mori, 1997*b*, Structural Stability and Evolution of Urban Systems. *Regional Science and Urban Economics* 27, 399–442.

M. Fujita and H. Ogawa, 1982, Multiple Equilibria and Structural Transition of Non-Monocentric Urban Configurations. *Regional Science and Urban Economics* 12, 161–196.

M. Fujita and J.-F. Thisse, 1996, Economics of Agglomeration. *Journal of the Japanese and International Economics* 10, 339–378.

M. Fujita, J.-F. Thisse and Y. Zenou, 1997, On the Endogenous Formation of Secondary Employment Center. *Journal of Urban Economics* 41, 337–357.

E. D. Glaezer, H. D. Kallal, J. A. Scheinkman, and A. Shlefer, 1992, Growth and Cities. *Journal of Political Economy* 100, 1126-1152.

J. Gottman, 1961, *Megalopolis: The Urbanized Northeastern Seaboard of the United States.* MIT Press, Cambridge, Mass.

J. Grihfield and G. Tolley, 1987, City Size and Place as Policy Issue. In E. S. Mills (ed.) *Handbook of Regional and Urban Economics* II. Elsevier Science Publishers, Amsterdam, 1285–1311.

Y. Hadar, 1997, The Formation of Cities. Tel–Aviv University, (unpublished).

B. Harris, 1966, Basic Assumptions for a Simulation of the Urban Residential Housing and Land Market. The University of Pennsylvania, (unpublished).

B. Harris, J. Nathanson and L. Rosenberg, 1966, Research on Equilibrium Model of Metropolitan Housing and Location Choice: Interim Report. The University of Pennsylvania, (unpublished).

E. Helpman, 1998, The Size of Regions. In D. Pines, E. Sadka, and I. Zilcha (eds.) *Topics in Public Economics.* Cambridge University Press, New York, 33–54.

J. V. Henderson, 1974, The Sizes and Types of Cities. *American Economic Review* 64, 640–656.

J. V. Henderson, 1977, *Economic Theory and the Cities.* Academic Press, New York. Second edition in 1986.

J. V. Henderson, 1986, Efficiency of Resources Usage and City Size. *Journal of Urban Economics* 19, 47–70.

J. V. Henderson, 1988, *Urban Development.* Oxford University Press, New York.

J. V. Henderson, 1997, Medium Size Cities. *Regional Science and Urban Economics* 27, 583–612.

J. D. Herbert and B. H. Stevens, 1960, A Model for the Distribution of Residential Activity in Urban Areas. *Journal of Regional Science* 2, 21–36.

O. Hochman, D. Pines and J.-F. Thisse, 1995, On the Optimal Structure of Local Governments. *The American Economic Review* 85, 1224–1240.

H. Hotelling, 1929, Stability in Competition. *Economic Journal* 39, 41–57.

H. Imai, 1982, CBD Hypothesis and the Economics of Agglomeration. *Journal of Economic Theory* 28, 275–299.

W. Isard, 1956, *Location and Space Economy.* MIT Press, Cambridge, Mass.

J. Jacobs, 1969, *The Economy of Cities.* Random House, New York.

T. C. Koopmans, 1957, *Three Essays on the State of Economic Science.* McGraw Hill, New York.

P. Krugman, 1991, Increasing Returns and Economic Geography. *Journal of Political Economy* 99, 483–499.

P. Krugman, 1995, *Development, Geography and Economic Theory.* Cambridge University Press, Cambridge.

P. Krugman, 1998, A "Slime Mold" Model of City Formation. In D. Pines, E. Sadka, and I. Zilcha (eds.), *Topics in Public Economics.* Cambridge University Press, New York, 15–32.

P. Krugman and M. Fujita, 1993, When Is the Economy Monocentric?: von Thünen and Chamberlin Unified. *Regional Science and Urban Economics* 25, 505–528.

A. Lösch, 1940, *Die Raumliche Ordnung der Wirtschaft.* Translated in English as *The Economics of Location* by W. H. Woglom and W. F. Stopler. Yale University Press, New Haven, 1954.

A. Marshall, 1890, *Principles of Economics.* Reprinted by MacMillan, London, 1946.

E. S. Mills and B. W. Hamilton, 1994, Urban Economics. Harper Collins College Publishers, New York.

T. Mori, 1997, A Modelling of Megalopolis Formation: The Maturing of City Systems. *Journal of Urban Economics* 42, 133–157.

L. Mumford, 1961, *The City in History.* Harcourt, Brace and World, New York.

R. F. Muth, 1961*a*, Economic Change and Urban-Rural Land conversions. *Econometrica 29*, 1-23.

R. F. Muth, 1961*b*, The Spatial Structure of the Housing Market, *Papers and Proceedings of the Regional Science Association* 7, 207–220.

R. F. Muth, 1969, *Cities and Housing.* The University of Chicago Press, Chicago.

H. Ogawa and M. Fujita, 1980, Equilibrium Land Use Pattern in a Non–Monocentric City. *Journal of Regional Science* 20, 455–475.

G. Ottaviano and J.–F. Thisse, 1998, Agglomeration and Trade Revisited. CORE (unpublished).

Y. Y. Papageorgiou, 1979, Agglomeration. *Regional and Urban Economics* 9, 41–59.

Y. Y. Papageorgiou and T. R. Smith, 1983, Agglomeration as Local Instability of Spatially Uniform Steady–States. *Econometrica* 51, 1109–1119.

Y. Y. Papageorgiou and J.–F. Thisse, 1985, Agglomeration as Spatial Interdependence Between Firms and Households. *Journal of Economic Theory* 37, 19–31.

T. Parsons, 1977, *The Evolution of Societies.* Prentice–Hall, Englewood Cliffs, NJ.

A. H. Pascal and J. J. McCall, 1980, Agglomeration Economies, Search Costs and Industrial Location. *Journal of Urban Economics* 8, 383–388.

K. Sasaki and S.–I. Mun, 1996, A Dynamic Analysis of Multiple Center Formation in a City. *Journal of Urban Economics* 40, 257–278.

S. A. Scotchmer and M. H. Wooders, 1987, Competitive Equilibrium and the Core in Club Economies with Anonymous Crowding. *Journal of Public Economics* 34, 159–174.

A. Shaked, 1982, Human Environment as a Local Public Good. *Journal of Mathematical Economics* 9, 275–283.

R. M. Solow, 1973*a*, Congestion Cost and the Use of Land for Streets. *Bell Journal of Economics and Management Science* 4, 602–618.

R. M. Solow, 1973*b*, On Equilibrium Models of Urban Location. In M. J. Parkin (ed.), *Essays in Modern Economics*. Longmans, London, 2-16.

R. M. Solow and W. S. Vickrey, 1971, Land Use in a Long Narrow City. *Journal of Economic Theory* 3, 430–447.

K. Stahl and P. Varaiya, 1978, Economics of Information: Examples in Location and Land–Use Theory. *Regional Science and Urban Economics* 8, 43–56.

D. M. Starrett, 1988, *Foundations of Public Economics*. Cambridge University Press, Cambridge.

J. E. Stiglitz, 1977, The Theory of Local Public Goods. In Martin Feldstein and Robert P. Inman (eds.) *The Economics of Public Services*. Macmillan, London, 274–333.

C. Stuart, 1979, Search and the Spatial Organization of Trading. In S. A. Lippman and J. J. McCall (eds.), *Studies in the Economics of Search*. North–Holland, New York, 17–33.

T. Tabuchi, forthcoming, Urban Agglomeration and Dispersion: A Synthesis of Alonso and Krugman. *Journal of Urban Economics*.

J. H. von Thünen, 1826, *Der Isolierte Staadt in Beziehung auf Landwirtschaft und Nationalöconomie*. Translated in English as *von Thünen's Isolated State* by C. W. Wartenberg. Pergamon Press, Oxford, 1966.

C. M. Tiebout, 1956, A Pure Theory of Local Public Goods. *Journal of Political Economy* 64, 416–424.

K. F. Vieand, 1987, An Extension of the Monocentric Urban Spatial Equilibrium Model to a Multicenter Setting: The Case of the Two–Center City. *Journal of Urban Economics* 21, 259–271.

W. C. Wheaton, 1974, Linear Programming and Locational Equilibrium: The Herbert–Stevens Model Revisited. *Journal of Urban Economics* 1, 278–287.

M. J. White, 1976, Firm Suburbanization and Urban Subcentres. *Journal of Urban Economics* 3, 323–343.

M. J. White, 1988, Location Choice and Commuting in Cities with Decentalized Employment. *Journal of Urban Economics* 24, 129–152.

J. D. Wilson, 1987, Trade in a Tiebout Economy. *American Economic Review* 77, 431–441.

A. Wolinsky, 1981, Retail Trade Concentration Due to Consumer's Imperfect Information. *Economics Discussion Paper* 195, Bell Laboratories.

G. K. Zipf, 1949, *Human Behavior and the Principle of Least Effort*. Addison–Wesley, Reading, Mass.

10

The Polycentric City

In this chapter we treat urban centres as facilities collectively used by the whole or part of the urban population. We employ the framework of club theory, which accounts for the collective use of facilities as a way of enhancing the utility level of club members.[1] We discuss the fundamentals of non–spatial club theory, and we extend it to the spatial domain. A single club in this context provides the analog of a monocentric city, while a system of clubs corresponds to a polycentric city. Although the spatial structure of the polycentric system we present here resembles that of chapter eight, in the sense that both are based on nested market areas and individuals patronise several centres in both, it differs with respect to spatial organisation because, unlike the polycentric city of chapter eight, no central facility here supplies more than a single collective good. We characterise the optimal size of such a system, and we explain why decentralisation must be undertaken at the level of the urban territory itself, rather than at the level of individual facilities as suggested by fiscal federalism. We then study in detail a simplified version of this model in order to understand what characterises the transition from a monocentric to a polycentric urban structure and what determines whether the location of a public facility will be central or peripheral.

---

[1] In what follows we use the term 'club theory' also for local public good theory since both are concerned with the efficient use of collective goods (see Berglas and Pines (1981)).

10.1 Club Theory

Club theory is concerned with the collective use of facilities by small groups, often referred to as "consumption groups". It differs both from the standard public–good theory (where the group is large) and from the standard private–good theory (where the group reduces to a single individual). Club theory explains the conditions under which a collective use of facilities by consumption groups is advantageous, the characteristics of club allocations and the conditions under which such allocations can be achieved within a competitive market structure.

The single main advantage for the collective use of a facility is sharing its cost, while the disadvantage is a fundamental aversion towards crowding and/or its cost. The collective use of a facility is preferable to its private use whenever the above advantage exceeds the disadvantages. When the advantage exceeds the disadvantages for small groups but not for large groups, the institution of a club emerges as an efficient way of consuming the good and, probably, as the prevailing institution in the market place.

Decentralising the supply of club goods through competitive markets requires several conditions, among which is the possibility of monitoring the users of the collective facilities inexpensively. The description of a competitive market structure in which collective goods are supplied and which yields an efficient allocation is discussed and demonstrated in Berglas (1976), Berglas and Pines (1981) and, especially, Scotchmer and Wooders (1987).[2] Club decentralisation in this context can be conceived as a standard market institution, where agents trade the right to impose crowding. As long as it is possible to monitor this effect, there is nothing to differentiate clubs from standard markets in which private goods are traded. Decentralisation becomes problematic only when the crowding effect cannot be monitored. In particular, this may be the case when the crowding effect is not anonymous, that is, when it depends on the types of individuals who form the consumption group. In this case monitoring requires information both on the individuals who generate the crowding effect and on those affected by it.[3] In addition Scotchmer (1994) and Conley and Wooders (1998) argue that decentralisation in this case becomes extremely difficult when the composition of optimal clubs and the composition of the population at large are inconsistent.

---

[2] Their results hinge on the assumption that the population can sort itself into an *integer* number of efficient clubs. But this assumption is no more restrictive than an analogous one required to establish the existence of equilibrium in the standard private case, where a competitive equilibrium with free entry can prevail only if the quantity demanded at the minimum average cost, divided by the firm's cost-minimising output, is an integer.

[3] For example, the crowding effect of a visitor to a swimming pool may well depend on the gender of the visitor and on the gender composition of the other visitors. Moreover, it may well depend on the specific attractiveness of the visitor to each of the other visitors in the swimming pool.

10.1.1 *Elements of Club Theory*

The following exposition draws from Buchanan (1965), Berglas (1976), Berglas and Pines (1980, 1981), Scotchmer and Wooders (1987), and Starrett (1988).[4]

The Model

Consider $\bar{N}$ individuals who are identical in both their preferences and initial endowment. The well–behaved utility function of a representative individual is given by $u\left[Z, G, \bar{\Upsilon}, \bar{G}\right]$, where Z is the *amount of the composite good consumed by the individual*, G is the *frequency of visits to a collectively used facility*, $\bar{\Upsilon}$ specifies the *characteristics of the facility* and $\bar{G}$ denotes the *aggregate visits to the collectively used facility*, which can also be interpreted as the determinant of *crowding*. Notice that $\bar{G} \equiv \bar{M}G$, where $\bar{M} \leq \bar{N}$ is the *total number of customers serviced by the facility* (the consumption group size). The *total cost for operating the facility* is $\bar{C} = c\left[\bar{\Upsilon}, \bar{G}\right]$, and it is given in terms of the composite good. Hence, if individuals are treated equally, the consumption of the composite good is $\bar{\Omega} - c\left[\bar{\Upsilon}, \bar{G}\right]/\bar{M}$ and utility can be written as

$$u\left[Z, G, \bar{\Upsilon}, \bar{G}\right] = u\left[\bar{\Omega} - c\left[\bar{\Upsilon}, \bar{G}\right]/\bar{M}, G, \bar{\Upsilon}, \bar{G}\right]. \tag{10.1}$$

Efficient Allocation for a Given Consumption Group Size

For a given consumption group size, $\left(G^*, \bar{\Upsilon}^*\right)$ is an *efficient club allocation* if and only if it maximises the RHS of (10.1). It is preferable, however, to define the optimisation problem in a somewhat less restrictive way. Instead of choosing G directly, its price $\bar{P}$ is chosen leaving the choice of G to the individuals. Accordingly, for a given consumption group size $\bar{M}$, $\left(\bar{P}^*, \bar{\Upsilon}^*\right)$ is an *efficient club allocation* if and only if it maximises U subject to

$$\begin{aligned} \bar{M} \cdot z\left[\bar{P}^*, \bar{\Upsilon}^*, \bar{G}^*, U\right] + c\left[\bar{\Upsilon}^*, \bar{G}^*\right] &= \bar{\Omega} \\ \bar{M} \cdot g\left[\bar{P}^*, \bar{\Upsilon}^*, \bar{G}^*, U\right] &= \bar{G}^* \end{aligned} \tag{10.2}$$

[4]Since geographical space does not play a rôle in the first three subsections of section 10.1, we do not need to distinguish between those variables that vary with location and those which do not. Therefore absence of a bar in subsections 10.1.1–10.1.3 does not imply that the corresponding variable varies with location. For consistency throughout the chapter, we still use bars in the first three subsections of section 10.1 so that the same symbols appear in both non–spatial and spatial models.

where $z[\cdot]$ and $g[\cdot]$ denote the compensated demands for Z and G respectively. The first-order conditions of this optimisation problem are

$$\begin{aligned} \bar{M}\cdot \text{MRS}\langle Z^*,\bar{\Upsilon}^*\rangle &\equiv -\bar{M}\cdot\frac{\partial e^*}{\partial\bar{\Upsilon}} \\ &= \frac{\partial c^*}{\partial\bar{\Upsilon}} \qquad (a) \\ \text{MRS}\langle Z^*,G^*\rangle &\equiv \bar{P}^* \\ &= \frac{\partial c^*}{\partial\bar{G}}+\bar{M}\cdot\frac{\partial e^*}{\partial\bar{G}} \qquad (b) \end{aligned} \tag{10.3}$$

where $Z^* \equiv z\left[\bar{P}^*,\bar{\Upsilon}^*,\bar{G}^*,U\right]$ and $e[\cdot] \equiv z[\cdot]+\bar{P}g[\cdot]$ is the minimum expenditure function. Expression (10.3(a)) represents Samuelson's (1954) rule for providing a collective good, where the LHS is the marginal benefit of the facility to the group in terms of the composite good and the RHS is the marginal cost. Expression (10.3(b)) implies marginal cost pricing. Here the LHS is the subjective evaluation of visits to the facility by a consumption group member and the RHS is the marginal social cost. The latter consists of two components, namely, the marginal resource cost of using the facility and the crowding diseconomies a user imposes on all other users. Thus (10.3(b)) states that the facility should be used by someone up to the point where the marginal benefit he or she derives from such use equals the sum of its direct and indirect cost to the group.

Private, Public and Club Goods: The Optimal Consumption Group Size

Equations (10.3) and $G^* = \bar{G}^*/\bar{M}$ can be used to solve for $\bar{\Upsilon}^*$ and $\bar{G}^*$ in terms of $\bar{M}$. Substituting these into (10.1) we obtain the reduced form of the utility level attained by the consumption group members as a function of group size, which we denote by $\bar{U}^*\left[\bar{M}\right]$. Four alternative configurations of $\bar{U}^*\left[\bar{M}\right]$ are shown in figure 10.1. Under configuration PRG, utility is decreasing with the consumption group size. Since in this case the optimal group size is the smallest possible, it corresponds to a private good. Configuration PUG implies that the optimal group is the entire population, hence it corresponds to a public good. According to configuration CLG, the optimal group size is $\bar{M}^\circ$. Assuming that it is small relatively to the total population in the economy, so that the total number of clubs $\bar{\kappa} \equiv \bar{N}/\bar{M}^\circ$ is a large integer number, the good represented by CLG is a club good.[5] Finally, the optimal group size corresponding to ING is indeterminate.

[5] Two important possibilities have so far been ignored. One is the case where $\bar{\kappa}$ is a small integer. The second is when $\bar{\kappa}$ is both small and not a integer. Both these possibilities give rise to what in the literature has been called the *integer problem*. We shall discuss the first case later on in this chapter. In chapter twelve we shall discuss the implications of the second case.

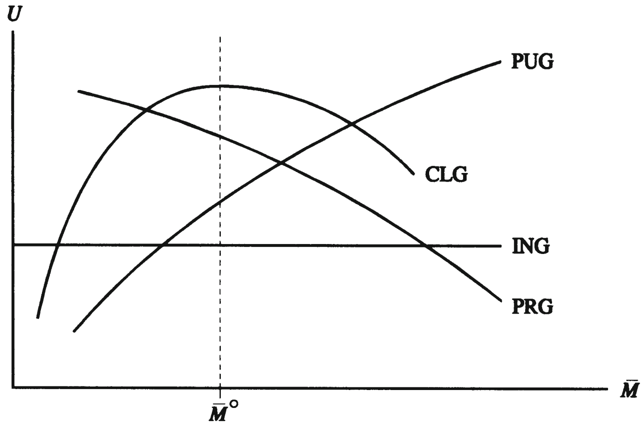

FIGURE 10.1. Alternative Relations Between Utility and Consumption Group Size.

Observe that the classification of a good in one of these categories does not necessarily depend on the good itself but rather on the economic background. For example, a swimming pool is often a private good in an affluent community where the demand for privacy is high, it is club good in a middle–class neighborhood and, perhaps, a (local) public good serving several low–income neighborhoods. The same can even apply to the classical "publicness" of law and order: affluent groups may establish gated neighborhoods defended by private police.

Our formulation of the club model implies that the shape of $\bar{U}^*\left[\bar{M}\right]$ depends on both the utility and cost functions. Specifically, applying the envelope theorem to $u\left[\bar{\Omega} - c\left[\bar{\Upsilon}, \bar{G}\right]/\bar{M}, G, \bar{\Upsilon}, \bar{G}\right]$ and using (10.3) together with $G = \bar{G}/\bar{M}$, we get

$$\begin{aligned}
\frac{d\bar{U}^*}{d\bar{M}} \div \frac{\partial u^*}{\partial Z} &= \frac{\bar{C}^*}{\bar{M}^2} - G^*\left(\frac{1}{\bar{M}}\frac{\partial c^*}{\partial \bar{G}} - \text{MRS}\left[Z^*, \bar{G}^*\right]\right) \\
&= \left(\bar{C}^* - \bar{G}^*\frac{\partial c^*}{\partial \bar{G}} - \bar{\Upsilon}^*\frac{\partial c^*}{\partial \bar{\Upsilon}}\right)/\bar{M}^2 \\
&+ \left(\bar{G}^*\frac{\partial u^*}{\partial \bar{G}} + \bar{\Upsilon}^*\frac{\partial u^*}{\partial \bar{\Upsilon}}\right)/\left(\bar{M}\frac{\partial u^*}{\partial Z}\right) \\
&= \frac{\bar{C}^*\left(1 - \varepsilon_{\bar{c}}\right)}{\bar{M}^2} + u\left[\cdot\right]\varepsilon_{u:\left(\bar{\Upsilon},\bar{G}\right)} \div \left(\bar{M}\frac{\partial u^*}{\partial Z}\right),
\end{aligned} \tag{10.4}$$

where ε_c is the degree of homogeneity of $\bar{C}$ and $\varepsilon_{u:\left(\bar{\Upsilon},\bar{G}\right)}$ is the degree of homogeneity of $u\left[Z, G, \bar{\Upsilon}, \bar{G}\right]$ with respect to $\bar{\Upsilon}$ and $\bar{G}$.

TABLE 10.1. Examples for Various Types of Goods.

$u[\cdot]$	$c[\cdot]$	$d\bar{U}^*/d\bar{M}$	TYPE
$u\left[Z, G, \frac{\bar{\Upsilon}}{\bar{G}}\right]$	$\bar{\Upsilon}^2 + \bar{G}\left(\frac{\bar{G}}{\bar{\Upsilon}}\right)$	negative	PRG
$u\left[Z, G, \frac{\bar{\Upsilon}}{\bar{G}^{.5}}\right]$	$\bar{\Upsilon} + \bar{G}\left(\frac{\bar{G}}{\bar{\Upsilon}}\right)^{\gamma}$	positive	PUG
$Z^{.5}G^{.5}$	$\bar{\Upsilon} + A + \frac{\bar{G}^4}{\bar{\Upsilon}}$	$\geq (\leq) 0 \Leftrightarrow \bar{M} \leq (\geq) 4A/\bar{\Omega}$	CLG
$u\left[Z, G, \frac{\bar{\Upsilon}}{\bar{G}}\right]$	$\bar{\Upsilon}^2 + \bar{G}\left(\frac{\bar{G}}{\bar{\Upsilon}}\right)^{\gamma}$	zero	ING

Examples that illustrate the analysis based on (10.4) appear in table 10.1. In the first case the degree of homogeneity of the utility function with respect to $\bar{\Upsilon}$ and $\bar{G}$, $\bar{\Upsilon}\partial u/\partial\bar{\Upsilon} + \bar{G}\partial u/\partial\bar{G}$, vanishes while the degree of homogeneity of the cost function, $\bar{\Upsilon}\partial c/\partial\bar{\Upsilon} + \bar{G}\partial c/\partial\bar{G}$, exceeds unity. Therefore utility decreases with the group size, which corresponds to a private good (PRG in figure 10.1). In the second case the homogeneity of the utility function is positive and that of the cost function is unitary. Thus utility increases with the group size, which corresponds to a public good (PUG in figure 10.1). In the third case the homogeneity of the utility function is zero while the homogeneity of the cost function is positive (zero,negative) for $\bar{G} < (=,>)\,(A/2)^{1/2}$. It is, therefore, impossible to infer directly from (10.4) whether scale economies or diseconomies are dominant. However, since $G^* = \left(\bar{\Omega}/2\bar{M}\right)^{1/2}/2$, it follows that the homogeneity of the utility function is positive (zero,negative) for $\bar{M} < (=,>) 4A/\bar{\Omega}$. This implies a $\cap$-shape for $U^*\left[\bar{M}\right]$, which is a necessary condition for a club good (CLG in figure 10.1). Finally, in the fourth case, the homogeneity of the utility function vanishes and that of the cost function is unitary, so that group size does not affect utility (ING in figure 10.1).

The Optimal Club

An allocation $\left(G^{\circ}, \bar{\Upsilon}^{\circ}, \bar{G}^{\circ}, \bar{M}^{\circ}, \bar{\kappa}^{\circ}\right)$ is an *optimal club* if and only if it maximises $\bar{U} = u\left[\bar{\Omega} - c\left[\bar{\Upsilon}, \bar{G}\right]/\bar{M}, G, \bar{\Upsilon}, \bar{G}\right]$ subject to

$$\begin{aligned} \bar{\kappa}\bar{M} &= \bar{N} \quad (a) \\ \bar{M}G &= \bar{G} \quad (b) \end{aligned} \tag{10.5}$$

where u and c are known functions while $\bar{N}$ and $\bar{\Omega}$ are given parameters. We want to characterise the *optimal consumption group size* $\bar{M}^{\circ}$. Rewriting (10.4) gives

$$\frac{d\bar{U}^{*}}{d\bar{M}} \div \frac{\partial u^{*}}{\partial Z} = \frac{\bar{C}^{*}}{\bar{M}^{2}} - \left(\frac{G^{*}}{\bar{M}}\frac{\partial c^{*}}{\partial \bar{G}} - G^{*} \cdot \mathrm{MRS}\left\langle Z^{*}, \bar{G}^{*}\right\rangle\right). \tag{10.6}$$

The RHS of (10.6) is the difference between the positive cost–sharing effect and the negative crowding effect associated with an increase in the size of the consumption group. The first term in the parenthesis is the direct effect on the facility's resource cost, while the second is the crowding effect on the utility of the users. The difference between the cost sharing and the crowding effects vanishes at the optimal consumption group size. There, (10.6) collapses to

$$\bar{C}^{\circ} = \bar{G}^{\circ}\left(\frac{\partial c^{\circ}}{\partial \bar{G}} - \bar{M}^{\circ} \cdot \mathrm{MRS}\left\langle Z^{\circ}, \bar{G}^{\circ}\right\rangle\right). \tag{10.7}$$

According to the RHS of (10.7), the cost of the facility is just balanced by the total imputed user charge, that is, the aggregate value of the marginal crowding effect. Consequently, if the club's members are charged their marginal crowding effect, the revenue would be just sufficient to finance the cost of the facility at the optimal club size. This result will be related to the Henry George rule when, in addition to club crowding, there exists spatial crowding as well.

Decentralisation

The possibility of decentralising the provision of club goods was the main issue examined by Tiebout (1965) in his classic paper.[6] More recently, decentralisation has become one of the central problems in club theory. Since Berglas (1976), it continues occupying the literature (see Wooders (1978), Berglas and Pines (1980, 1981), Scotchmer and Wooders (1987), Barham and Wooders (1998), Conley and Wooders (1998), and Gilles and Scotchmer (1998)). In this section we begin discussing the issue.

We use a price–taking equilibrium based on goods of variable quality, which was used by Rosen (1974) and Berglas and Pines (1980). We are concerned with pure price–taking equilibria, where the only signal relevant to an agent's decision–making is price. Here price is not necessarily a parameter, but rather a function of the attributes that characterise a good. To fix ideas consider the good "a visit to the facility". Its quality depends on the attributes of the facility $\bar{\Upsilon}$ and on crowding $\bar{G}$, so that the corresponding price is given by $P\left[\bar{\Upsilon}, \bar{G}\right]$. By choosing the attributes of the facility, individuals determine the price to be paid per visit. By choosing also the number of visits to the facility G, their feasible utility level $u\left[\bar{\Omega} - P\left[\bar{\Upsilon}, \bar{G}\right] G, G, \bar{\Upsilon}, \bar{G}\right]$ is uniquely determined. Likewise, each entrepreneur selects the attributes of the facility in order to maximise profit $P\left[\bar{\Upsilon}, \bar{G}\right] \bar{G} -\ c\left[\bar{\Upsilon}^{e}, \bar{G}^{e}\right]$, which is driven to zero because of competition under

[6] This issue was only alluded in Buchanan (1965) who was the first to formulate club theory in the strict sense (i.e. when the group is not a local jurisdiction).

free entry. Formally, an allocation $\left(G^e, \bar{\Upsilon}^e, \bar{G}^e, \bar{M}^e, \bar{\kappa}^e\right)$ and a price $P\left[\bar{\Upsilon}, \bar{G}\right]$ is a *club equilibrium* if and only if

$$\begin{aligned}
u\left[\bar{\Omega} - P\left[\bar{\Upsilon}^e, \bar{G}^e\right] G^e, G^e, \bar{\Upsilon}^e, \bar{G}^e\right] &\geq u\left[\bar{\Omega} - P\left[\bar{\Upsilon}, \bar{G}\right] G, G, \bar{\Upsilon}, \bar{G}\right] \\
& \quad \text{for all } \left(G, \bar{\Upsilon}, \bar{G}\right) \qquad (a) \\
P\left[\bar{\Upsilon}^e, \bar{G}^e\right] \bar{G}^e - c\left[\bar{\Upsilon}^e, \bar{G}^e\right] &\geq P\left[\bar{\Upsilon}, \bar{G}\right] \bar{G} - c\left[\bar{\Upsilon}, \bar{G}\right] \\
& \quad \text{for all } \left(\bar{\Upsilon}, \bar{G}\right) \qquad (b) \\
\bar{M}^e \bar{\kappa}^e &= \bar{N} \qquad (c) \\
\bar{M}^e G^e &= \bar{G}^e \qquad (d) \\
P\left[\bar{\Upsilon}^e, \bar{G}^e\right] \bar{G}^e &= c\left[\bar{\Upsilon}^e, \bar{G}^e\right] \qquad (e)
\end{aligned} \tag{10.8}$$

where P, u and c are known functions, $\bar{\Omega}$ and $\bar{N}$ are given parameters, and where $\bar{\kappa}^e$ is large integer number. Condition (a) ensures utility maximisation, (b) profit maximisation, (c) and (d) clearing the club and the facility use markets respectively, and (e) zero profits because of free entry. In section 10.4.1 of the appendix we show that a club equilibrium is optimal (first welfare theorem), and that a club optimum can be supported by a competitive price system (second welfare theorem). We conclude that the provision of club goods can be decentralised.

10.1.2 Extensions

Heterogeneous Population

Suppose now that the population is heterogeneous, in the sense that individuals can be partitioned into a number of types differing with respect to either their initial endowment or tastes or both as in section 5.3.3. A remarkable implication of club theory in that case is given by the *segregation theorem* which states that, under certain conditions, an efficient allocation requires distributing individuals among clubs so that every club contains a single type.[7] In what follows we provide some intuition for this result by discussing a simple example along the lines of Berglas (1976).

Suppose that there are two types of individuals and that the fixed population of each type is given by $\bar{N}_i$, $i = 1, 2$. Consider any system of integrated clubs which contain both types. Since we still ignore the integer problem, if the club system under consideration is efficient, it must contain clubs identical with respect to their type composition. Consider any two such identical clubs, named a and b, in which members of type one consume the bundle (Z_1, G_1) and members

[7] For details see Berglas and Pines (1981), and Scotchmer and Wooders (1987).

of type two consume (Z_2, G_2). Transfer one individual of type one from a to b and G_1/G_2 individuals of type two from b to a. Suppose that every transferred individual continues to consume the same bundle as before the transfer. Under these circumstances the change in the total use of the facilities in those two clubs is given by

$$\begin{aligned} \Delta\bar{G}_a &= -G_1 + G_2\frac{G_1}{G_2} = 0 \quad (a) \\ \Delta\bar{G}_b &= G_1 - G_2\frac{G_1}{G_2} = 0. \quad (b) \end{aligned} \tag{10.9}$$

Since the total use of the facilities remains the same in both clubs, this reallocation provides the same utility as before to both types of club members.

Now if the original allocation were efficient, conditions (10.3) would have been satisfied. Since however the transfer has changed the population compositions in each of the two clubs a and b, it must be the case that these conditions are violated after the transfer. Thus the allocation after the transfer can be increased and, therefore, it cannot be efficient. Since the utilities achievable before the transfer are equal to those achievable in a feasible and inefficient allocation, we conclude that the original allocation cannot be efficient.[8]

Ignoring the integer problem, the segregation theorem implies that the optimal allocation is characterised by a system of segregated clubs, each serving identical patrons. A movement along the UPF (utility possibility frontier) defined on types cannot be done by mixing different individuals but, rather, by redistributing the initial endowment while still allowing different individuals to form their exclusive clubs.

The segregation theorem can also be explained in positive terms. The isomorphism between an optimum and a competitive equilibrium for an homogeneous population is established in section 10.4.1 of the appendix using the concept of variable–quality goods. This result is valid for an heterogeneous population as well. Segregation follows from the difference in the choice of quality by individuals with different income and tastes. Only by coincidence individuals with different income and tastes will choose identical attributes of the commodity.

Multi–Product Clubs

Since there are many collective goods, one can think about an economic system where each individual belongs to several clubs, each engaging in the supply of a single collective good. For the case of two collective goods, achievable utility

[8] The only exception arises when the marginal valuations of both facility and congestion are the same for both types in the original mixed clubs, that is,

$$\frac{\partial u_1/\partial \Upsilon_1}{\partial u_1/\partial Z_1} = \frac{\partial u_2/\partial \Upsilon_2}{\partial u_2/\partial Z_2} \text{ and } \frac{\partial u_1/\partial \bar{G}_1}{\partial u_1/\partial Z_1} = \frac{\partial u_2/\partial \bar{G}_2}{\partial u_2/\partial Z_2} .$$

can be represented by

$$\bar{U}^{1,2} = \max_{G_1,G_2,\bar{\Upsilon}_1,\bar{\Upsilon}_2,\bar{M}_1,\bar{M}_2} u\left[\bar{\Omega} - \frac{\bar{C}_1}{\bar{M}_1} - \frac{\bar{C}_2}{\bar{M}_2}, G_1, G_2, \bar{\Upsilon}_1, \bar{\Upsilon}_2, \bar{M}_1 G_1, \bar{M}_2 G_2\right]. \tag{10.10}$$

However economies of scope may cause the emergence of diversified clubs, each supplying more than a single good (see Berglas and Pines (1981), and Brueckner and Lee (1991))). Utility in that case, where the consumption group size must be the same for both goods, is written as

$$\bar{U}^{1+2} = \max_{G_1,G_2,\bar{\Upsilon}_1,\bar{\Upsilon}_2,\bar{M}} u\left[\bar{\Omega} - \frac{\bar{C}^{1+2}}{\bar{M}}, G_1, G_2, \bar{\Upsilon}_1, \bar{\Upsilon}_2, \bar{M}G_1, \bar{M}G_2\right] \tag{10.11}$$

where $\bar{C}^{1+2} \equiv c\left[\bar{\Upsilon}_1, \bar{\Upsilon}_2, \bar{M}G_1, \bar{M}G_2\right]$. Following Berglas and Pines (1981), the advantage of the diversified over the specialised club is given by

$$\begin{aligned} &\bar{U}^{1+2} - \bar{U}^{1,2} \\ = &\left(\bar{U}^{1+2} - \max_{G_1,G_2,\bar{\Upsilon}_1,\bar{\Upsilon}_2,\bar{M}} u\left[\bar{\Omega} - \frac{\bar{C}_1}{\bar{M}} - \frac{\bar{C}_2}{\bar{M}}, G_1, G_2, \bar{\Upsilon}_1, \bar{\Upsilon}_2, \bar{M}G_1, \bar{M}G_2\right]\right) \\ - &\left(\bar{U}^{1,2} - \max_{G_1,G_2,\bar{\Upsilon}_1,\bar{\Upsilon}_2,\bar{M}} u\left[\bar{\Omega} - \frac{\bar{C}_1}{\bar{M}} - \frac{\bar{C}_2}{\bar{M}}, G_1, G_2, \bar{\Upsilon}_1, \bar{\Upsilon}_2, \bar{M}G_1, \bar{M}G_2\right]\right). \end{aligned} \tag{10.12}$$

The first parenthesis on the RHS of (10.12) is the gain from combining the two separate clubs into one or, equivalently, the economies of scope.[9] The second parenthesis gives the corresponding loss from forcing the same consumption group size on two different clubs. The diversified club will prevail over the two specialised clubs whenever the economies of scope exceed the loss from inefficient club sizes. In that case, if the population is heterogeneous, the segregation theorem may no longer apply (see Berglas and Pines (1981)).

10.1.3 Clubs as Institutions of Collective Economic Activity

In its original formulation club theory explains the collective use of facilities for enhancing the utility level of club members. There is, however, another consistent interpretation of club theory: a group of workers in a factory can also be thought of as a club in which equipment is used collectively by the workers. This interpretation becomes apparent when we reconsider the objective function in (10.1), $u\left[\bar{\Omega} - c\left[\bar{\Upsilon}, \bar{G}\right]/\bar{M}, G, \bar{\Upsilon}, \bar{G}\right]$, where $\bar{\Upsilon}$ is now *equipment and other non-labour inputs*, G is *per-capita labour input*, $\bar{G}$ is *total labour input*, $-c\left[\bar{\Upsilon}, \bar{G}\right]$ is the *value of output minus non-labour inputs* and $\bar{M} = \bar{G}/G$ is the

[9] An example of scope economies in this context arises when there is an initial fixed cost for each specialised club that can be reduced by combining the two into a single diversified club.

number of workers.[10] Hence club theory refers to any cooperation of a group of individuals in carrying out an economic activity, whether consumption or production, if the outcome exhibits scale economies with respect to group size when it is sufficiently small and scale diseconomies when it is sufficiently large.

The concept of a production club also applies when the group is engaged in several economic activities, some of which exhibit scale economies and others scale diseconomies, as long as scale economies dominate for a small group size and scale diseconomies dominate for a large group size. A good example is the model of Stiglitz (1977) which pertains to identical communities, each located on a separate island. Individuals are engaged in two activities: production which is characterised by scale diseconomies, and the collective consumption of a pure local public good which exhibits scale economies. The utility of a group is given by $\bar{U} = u\left[\left(f\left[\bar{M}\right] - \Upsilon\right)/\bar{M}, \Upsilon\right]$, where $f\left[\bar{M}\right]$ is the *average product of land* satisfying $f'\left[\bar{M}\right] > 0$ and $d\left(f\left[\bar{M}\right]/\bar{M}\right)/d\bar{M} < 0$, and where Υ is the *resources spent in the production of a pure public good.* Applying the envelope theorem gives

$$\frac{d\bar{U}^\circ}{d\bar{M}} = \frac{\partial u}{\partial Z^\circ}\left(\frac{\Upsilon^\circ}{\bar{M}^2} - \frac{f'\left[\bar{M}\right] - f\left[\bar{M}\right]/\bar{M}}{\bar{M}}\right). \tag{10.13}$$

The first term inside the parenthesis of (10.13) represents the cost–sharing advantage (scale economies), while the second represents scale diseconomies associated with the decreasing average product of labour. If both goods are essential and if $\lim_{\bar{M}\to 0} f\left[\bar{M}\right] = \lim_{\bar{M}\to\infty} f\left[\bar{M}\right]/\bar{M} = 0$ then the group is not viable when its size is either arbitrarily small or arbitrarily large. On the one hand, if group size is arbitrarily small then either Υ is arbitrarily close to zero or $f\left[\bar{M}\right] - \Upsilon$ is negative, which implies Z negative. On the other hand, if group size is arbitrarily large then Z cannot be positive. Since both goods are essential, we conclude that the group size must be bounded and, since utility is a continuous function of group size, there is some finite group size which corresponds to a maximum utility level. Although these considerations do not imply strictly a $\cap$–shape for $\bar{U}^\circ\left[\bar{M}\right]$, they are sufficient to secure the existence of a finite optimal group size.

[10] Under these definitions, we have $c\left[\bar{\Upsilon}, \bar{G}\right] < 0$, $\partial c/\partial\bar{\Upsilon} < 0$, $\partial^2 c/\partial\bar{\Upsilon}^2 > 0$, $\partial c/\partial\bar{G} < 0$, $\partial^2 c/\partial\bar{G}^2 > 0$ and $\partial u/\partial\bar{\Upsilon}^\circ = \partial u/\partial\bar{G}^\circ = 0$. With this specification, we obtain the following a first order conditions. Fistly,

$$\text{MRS}\left\langle Z_\circ, G^\circ\right\rangle = -\frac{\partial c^\circ}{\partial\bar{G}}$$

that is, the marginal valuation of leisure is equal to the marginal labour productivity. Secondly, the optimal $\bar{\Upsilon}$ implies its use up to where its marginal productivity equals its price. Finally, the optimal group size condition now requires that the marginal labour productivity is equal to its average productivity. Thus each worker gets the marginal product of labour.

10.1.4 Spatial Clubs

In this section we introduce geographical space and we explore the correspondence between a spatial club and a monocentric renter city.[11] Suppose that the location of a single collective facility represents the centre of the spatial club, which can be thought of as a monocentric city by itself. A population of $\bar{M}$ individuals (the consumption group) is located around the centre. The interaction of individuals with the centre does not only involve a single source of crowding from the use of the collective facility itself, as in the case of non–spatial clubs, but also crowding to save costly trips between any residential location and the centre. Given identical individuals, their utility now is $\bar{U} = u\left[Z, H, G, \bar{\Upsilon}, \bar{G}\right]$, where $\left(\bar{\Upsilon}, \bar{G}\right)$ replaces the shift parameter $\bar{\gamma}$ of section 2.1.2. Under these circumstances the individual compensated demands become

$$
\begin{aligned}
Z^i &= z\left[R, \bar{P}, U, \bar{\Upsilon}, \bar{G}\right] \quad (a) \\
H^i &= h\left[R, \bar{P}, U, \bar{\Upsilon}, \bar{G}\right] \quad (b) \\
G^i &= g\left[R, \bar{P}, U, \bar{\Upsilon}, \bar{G}\right] \quad (c)
\end{aligned}
\tag{10.14}
$$

where g is the *compensated demand function for the collective good.*

For $x \in \mathcal{X}$, an allocation $\left(n^\circ, \bar{\Upsilon}^\circ, \bar{G}^\circ, \bar{N}^\circ, \bar{x}^\circ\right)$ and a price system $\left(R^\circ, \bar{P}^\circ\right)$ are a *spatial club optimum* if and only if they maximise $\bar{U}$ subject to

$$
\begin{aligned}
nh\left[\cdot\right] &= \theta \quad (a) \\
\int_{\mathcal{X}} ndx &= \bar{N} \quad (b) \\
\int_{\mathcal{X}} ng\left[\cdot\right]dx &= \bar{G} \quad (c) \\
\int_{\mathcal{X}} n\left(z\left[\cdot\right] + T\right)dx + c\left[\bar{\Upsilon}, \bar{G}\right] + \Theta\left\langle\bar{x}\right\rangle\bar{R} &= \bar{N}\bar{\Omega} \quad (d)
\end{aligned}
\tag{10.15}
$$

where u, z, h, g, T, c, θ and Θ are known functions while $\bar{\Omega}$ and $\bar{R}$ are given parameters.

---

[11] Variants of spatial clubs have been discussed in Stiglitz (1977), Arnott (1979), Arnott and Stiglitz (1979), Hochman (1982*a*,*b*), Starrett (1988), Fujita (1989), and Hochman, Pines and Thisse (1995). Since our formulations here are explicitly spatial, we return to our standard convention of distinguishing between spatial and non–spatial variables by using a bar on the latter.

Adjusting the arguments of section 10.1.1 to the present formulation, we arrive at

$$
\begin{aligned}
-\int_{\mathcal{X}} n \cdot \frac{\partial e^{\circ}}{\partial \bar{\Upsilon}} dx &\equiv \int_{\mathcal{X}} n \cdot \mathrm{MRS}\left\langle Z^{\circ}, \bar{\Upsilon}^{\circ}\right\rangle dx \\
&= \frac{\partial c^{\circ}}{\partial \bar{\Upsilon}} \qquad (a) \\
\mathrm{MRS}\left\langle Z^{\circ}, G^{\circ}\right\rangle &\equiv \bar{P} \\
&= \frac{\partial c^{\circ}}{\partial \bar{G}} - \int_{\mathcal{X}} n \cdot \mathrm{MRS}\left\langle Z^{\circ}, \bar{G}^{\circ}\right\rangle dx \qquad (b)
\end{aligned}
\tag{10.16}
$$

which extend the Samuelson (1954) and the marginal cost pricing rules in (10.3) to the spatial domain. The effect of club size on utility now becomes

$$
\frac{d\bar{U}^{*}}{d\bar{M}} \div \frac{\partial u^{*}}{\partial Z} = \frac{\bar{C}^{*}}{\bar{M}^{2}} - \left(\frac{G^{*}}{\bar{M}} \frac{\partial c^{*}}{\partial \bar{G}} - G^{*} \cdot \mathrm{MRS}\left\langle Z^{*}, \bar{G}^{*}\right\rangle\right) - \frac{\overline{\mathrm{DLR}}}{\bar{M}^{2}} . \tag{10.17}
$$

Comparing this expression with (10.6), we observe that the effect of group size on utility is extended here to include $\overline{\mathrm{DLR}}$. In other words, whereas crowding is confined to the facility itself in a non–spatial club, crowding in a spatial club involves housing as well. For, in order to increase their accessibility, individuals converge towards the club facility at the expense of diminished housing consumption. The marginal cost of this type of crowding is reflected in $\overline{\mathrm{DLR}}$.

10.1.5 The Henry George Rule

For an optimal group size (10.17) becomes

$$
\begin{aligned}
\overline{\mathrm{DLR}}^{\circ} &= \bar{C}^{\circ} - \bar{G}^{\circ}\bar{P} \\
&= \bar{G}^{\circ}\left(\frac{\partial c^{\circ}}{\partial \bar{G}} - \bar{M}^{\circ} \cdot \mathrm{MRS}\left\langle Z^{\circ}, \bar{G}^{\circ}\right\rangle\right) .
\end{aligned}
\tag{10.18}
$$

When the collective good is pure of any crowding effect, so that only residential crowding matters, (10.18) collapses to $\overline{\mathrm{DLR}}^{\circ} = \bar{C}^{\circ}$, which is reminiscent of the original single–tax proposal of the classical economist Henry George (1896). Influenced by Thomas Malthus (1798) and David Ricardo (1821), George believed that land owners can take all the excess value derived by economic development through land rent. He also contended, along with other nineteenth–century rent theorists, that it is wrong for land owners to gain profit by renting their land, since land itself is not a product of labour. Consequently, George proposed that all land rents be taxed away and be used to finance government expenditures.[12]

---

[12] At the time, it was estimated that the single tax would be sufficient to support all levels of government in the United States.

Until the 1970s, the single–tax concept was treated as the outcome of a peculiar *normative* approach. Since then it is treated as a *positive* statement about the relationship between the deficits associated with the supply of a collective good and the differential land rent.[13] Observe that the Henry George rule does not emerge because of housing itself or the spatial structure of the model; rather, it emerges because of the crowding effect caused by the presence of a fixed factor of production. Thus the non–spatial specification of Stiglitz (1977), where land is a fixed factor in the production of the composite good, yields a similar Henry George result. The same is true whenever the production of any good other than housing is associated with scale diseconomies (see Berglas and Pines (1981)).

Finally, the Henry George rule (10.18) suggests that, under decentralisation, the entrepreneur who provides the collective good must also act as a developer who provides housing. In this more general context, the supporting price system must be modified to include the effect of the developer's decisions on land rent. As in the case of club equilibrium (10.8), it is possible to show that under certain assumptions a spatial club equilibrium does exist and it is optimal under the appropriate price system. We conclude that the provision of spatial club goods can be decentralised.

10.2 The City as an Optimal Complex of Spatial Clubs

Using the framework developed in section 10.1.2, one could imagine a system of urban clubs, each supplying a single collective good to a consumption group that consists of identical individuals who pay a user charge to finance its cost. Individuals belong to several clubs that differ from one–another with respect to the good they supply and, therefore, with respect to their consumption group size.

This description of city seems perfectly consistent with the fundamental policy proposal of fiscal federalism, about the decentralised provision of collective goods recommended as a basis for reforming the present system of local governments.[14] The appropriate user charge would not only suffice to finance fully each club, but it would also provide the incentive required to ensure optimal deci-

---

[13] Alternative versions of this rule in an urban context have been proposed by, among others, Flatters, Henderson and Mieszkowski (1974), Stiglitz (1977), Arnott and Stiglitz (1979), Berglas and Pines (1981), and Hochman (1981).

[14] Olson (1969, p.483) maintains that "...there is a need for a separate governmental institution for every collective good within a unique boundary, so that there can be a match between those who receive the benefits of a collective good and those who pay for it." Furthermore, Oates (1972, p.42) concludes that "...the optimal structure of the public sector was found to be one in which there was a level of government (or a collective decision making mechanism) for each jurisdiction over which the consumption of a public good could be defined." (See Hochman, Pines and Thisse (1995, p. 1226).)

sions for the provision of the collective goods (see section 10.1.1). However, one should recall that this policy and its justification crucially depend on the non–spatial nature of fiscal federalism. For when geographical space is introduced, the user charge is no longer sufficient and it must be supplemented by land rent proceeds in order to sustain club operations (see section 10.1.4). Indeed, since fiscal federalism in a geographical context implies overlapping market areas, the rent generated by the same piece of land should be divided to finance several clubs. But as Hochman, Pines and Thisse (1995, p.1225) point out, even when this difficulty is resolved, there "... still remains the issue of finding incentives to make the optimal decisions regarding the location and other attributes of the LPG's. The rent–sharing mechanism must relate the revenue to actions taken by each supplier of the LPG's so that this relationship reflects precisely the net social benefit of each of its actions. The major difficulty lies in implementing such a mechanism when market areas of different LPG's do overlap." This is why decentralisation in the manner prescribed by fiscal federalism fails when geography is explicitly taken into account, and there arises a necessity for city–wide governments to co–ordinate the activities of all urban clubs that provide the various collective goods in the city. Thus decentralisation when geography matters is based on *territories*, rather than goods, which exactly reverses the original prescription of fiscal federalism.

Taking into account geographical space does not only negate the fundamental policy proposal of fiscal federalism, but it also gives rise to the possibility of a polycentric city because facilities dispersed over the urban area, instead of concentrated on a single location, might save some cost. We now begin to discuss these issues more analytically. In the rest of section 10.2 we combine aspects of club theory and land–use theory along the lines suggested by Hochman, Pines and Thisse (1995) to examine decentralisation. In section 10.3 we present conditions for the emergence of multiple centres, and we highlight the determinants of their convergence and divergence under polycentricity.

10.2.1 The Model

We consider a linear city of unit width extending over $\mathcal{X} = [-\bar{x}, \bar{x}]$, providing $\bar{n}$ types of LPGs and inhabited by $\bar{M}$ identical individuals. The city boundaries are fixed and the opportunity cost of land $\bar{R}$ is set equal to zero. Utility is given by $\bar{U} = u[Z, H, \Upsilon]$, where $\Upsilon \equiv (\Upsilon_1, ..., \Upsilon_{\bar{n}})$ denotes *the quantities of* LPGs *consumed* by an individual. All goods are essential. Every individual is endowed with $\bar{\Omega}$ units of the composite good which can be used for private consumption, for producing the LPGs and for transportation.

Each of the $\bar{n}$ LPGs is supplied by $\bar{\kappa}_i$ *facilities* and is identified by ij, where $i \in \{1, ..., \bar{n}\}$ denotes the *type* of the LPG it provides and $j \in \{1, ..., \bar{\kappa}_i\}$ denotes the *index* of the specific facility. A facility ij provides Υ_{ij} units of the ith LPG to $\bar{M}_{ij}$ individuals residing within its market area. The *provision cost* of the facility ij is $C_{ij} = c_i \left[\Upsilon_{ij}, \bar{M}_{ij}\right] > 0$. If transportation cost increases with distance, utility maximisation implies that market areas of the same type cannot overlap. Therefore each boundary between two adjacent market areas of a given type represents the singular location of an individual who is indifferent between two corresponding facility locations. Furthermore, a residential location on either side of the boundary must belong to the market area of the facility located on the same side of the boundary. And if there is no vacant land then each market area is an interval and each facility is contained in its market area. Taking these into account, we denote the location of the facility ij as $\bar{x}_{i,2j}$ and the two boundaries of its market area as $\bar{x}_{i,2j-1}$ and $\bar{x}_{i,2j+1}$ respectively. Thus an individual located at $x \in [\bar{x}_{i,2j-1}, \bar{x}_{i,2j+1}]$ travels to the facility ij at cost $T_i[|x - \bar{x}_{i,2j}|]$. A possible layout of the ith–type facility system with $\bar{\kappa}_i = 3$ is shown in figure 10.2. Given that all goods are essential, the union of type–i market areas must coincide with the urban area $[-\bar{x}, \bar{x}]$. Consequently, we have

$$\bar{x}_{i,1} = -\bar{x} \text{ and } \bar{x}_{i,2\bar{\kappa}+1} = \bar{x}. \tag{10.19}$$

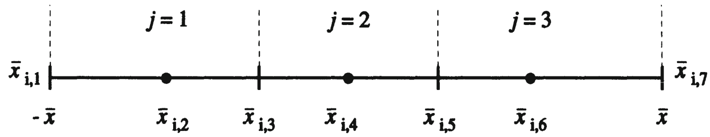

FIGURE 10.2. A System of Type–i Facilities with $\bar{\kappa}_i = 3$.

Since a facility ij provides Υ_{ij} units of the ith LPG to $\bar{M}_{ij}$ individuals who belong within its market area $[\bar{x}_{i,2j-1}, \bar{x}_{i,2j+1}]$, we can refer to the spatial and the economic characteristics of the facility ij as the (spatial) *club* ij. In particular, the location of a club describes the location of the corresponding facility. Our specification refers directly to public facilities such as schools, parks, museums and the like, where the individual makes a home–facility trip in order to enjoy the services rendered by the facility. It also describes situations where the social cost associated with a home–facility trip is a delivery cost as in the case of garbage collection and snow removal, or where the level of service deteriorates with distance as in the case of emergency care and fire or police protection. Although all these examples belong to consumption–group cases, a club may also represent a productive activity, where $\bar{M}_{ij}$ individuals use Υ_{ij} units of non–labour input collectively to produce $-C_{ij}$ units of the composite good. Hence this model is appropriate for analysing the employment–location issue as well.

An urban *club complex* is a given system of spatial clubs ($\bar{\kappa}_i$ for $i \in \{1, ..., \bar{n}\}$) that supply the $\bar{n}$ LPGs to the $\bar{M}$ individuals distributed over the urban area $[-\bar{x}, \bar{x}]$. Figure 10.3 provides an example of spatial structure for an urban club complex with $\bar{n} = 3, \bar{\kappa}_1 = 3, \bar{\kappa}_2 = 2$ and $\bar{\kappa}_3 = 1$. Further to (10.19), a *feasible*

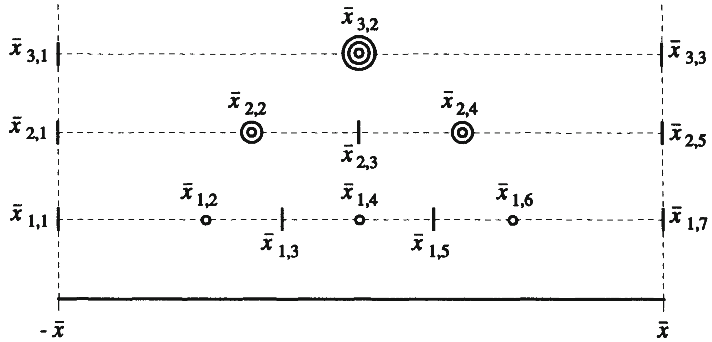

FIGURE 10.3. A Complex with $\bar{n} = 3, \bar{\kappa}_1 = 3, \bar{\kappa}_2 = 2$ and $\bar{\kappa}_3 = 1$.

club complex must satisfy

$$\begin{aligned} nh[\cdot] &= 1 \text{ for } x \in [-\bar{x}, \bar{x}] && (a) \\ \int_{\mathcal{X}} n dx &= \bar{M} && (b) \\ \int_{\bar{x}_{i,2j-1}}^{\bar{x}_{i,2j+1}} n dx &= \bar{M}_{ij} \quad \begin{matrix} \text{for } i \in \{1, ..., \bar{n}\} \\ \text{and } j \in \{1, ..., \bar{\kappa}_i\} \end{matrix} && (c) \\ \int_{\mathcal{X}} nz[\cdot] dx &+ \\ \sum_{i=1}^{\bar{n}} \sum_{j=1}^{\bar{\kappa}_i} \int_{\bar{x}_{i,2j-1}}^{\bar{x}_{i,2j+1}} nT_i[\|x - \bar{x}_{i,2j}\|] dx &+ \\ \sum_{i=1}^{\bar{n}} \sum_{j=1}^{\bar{\kappa}_i} C_{ij} &= \bar{M}\bar{\Omega}, && (d) \end{aligned} \tag{10.20}$$

where z and h are the compensated demands for the composite good and land given by[15]

$$\begin{aligned} Z &= z\left[R, \bar{U}, \Upsilon_{1,j_1(x)}, ..., \Upsilon_{\bar{n},j_{\bar{n}}(x)}\right] && (a) \\ H &= h\left[R, \bar{U}, \Upsilon_{1,j_1(x)}, ..., \Upsilon_{\bar{n},j_{\bar{n}}(x)}\right] && (b) \end{aligned} \tag{10.21}$$

and where T_i and c_i are known functions while $\bar{\Omega}$ and $\bar{M}$ are given parameters. Also, $j_i(x)$ is the index for the specific club of type i patronised by an individual who resides at x.[16]

---

[15] Observe that, if we include employment centres in our definition of clubs, $\bar{\Omega} = 0$.

[16] Notice that the supply of land in the land use constraint (10.20(a)) is given by $\theta = 1$ because the city is linear with unitary width. Also notice that the material balance constraint (10.20(d)) does not include a term for the opportunity cost of land because $\bar{R} = 0$.

10.2.2 The Optimal Club Complex

An *optimal club complex* is a feasible complex in which the total area and population size, the spatial distribution of population and land values, the club locations and their market areas are chosen to maximise the common level of utility. In fact, any replication of an optimal complex is an optimal complex. However, we shall confine our definition to an optimal complex which is not an integer replication of an optimal complex. Given the optimal number of spatial clubs ($\bar{\kappa}_i^{\circ}$ for $i \in \{1, ..., \bar{n}\}$), it can be shown that the first-order conditions of this problem imply the following fundamental characteristics of an optimal club complex.

We know that the supply of every collective good in an optimal complex must obey Samuelson's rule, namely,

$$-\int_{\bar{x}_{i,2j-1}}^{\bar{x}_{i,2j+1}} n^{\circ} \frac{\partial e^{\circ}}{\partial \Upsilon_{ij}} dx = \frac{\partial c_i^{\circ}}{\partial \Upsilon_{ij}} \text{ for } i \in \{1, ..., \bar{n}\} \text{ and } j \in \{1, ..., \bar{\kappa}_i\}. \tag{10.22}$$

Also since, in contrast to the cases presented in earlier sections, individuals here are restricted to one visit per unit of time for all club types, the marginal cost pricing rule reduces to the prescription that individuals must pay their marginal congestion cost. Under marginal cost pricing, the initial endowment net of transport cost is just sufficient to finance the optimal budget:

$$e\left[R^{\circ}, \bar{U}^{\circ}, G^{\circ}_{1,j_1(x)}, ..., G^{\circ}_{\bar{n},j_{\bar{n}}(x)}\right] + \sum_{i=1}^{\bar{n}} \frac{\partial c_i^{\circ}}{\partial \bar{M}_{i,j_i(x)}} = \bar{\Omega} - \sum_{i=1}^{\bar{n}} T_i \left[\left|x - \bar{x}_{i,j_i(x)}\right|\right] \tag{10.23}$$

where $e[\cdot]$ denotes the *minimum expenditure* on the composite good and housing. This, together with (10.20(d)), gives

$$\sum_{i=1}^{\bar{n}} \sum_{j=1}^{\bar{\kappa}_i} \left(\bar{M}_{ij}^{\circ} \frac{\partial c_i^{\circ}}{\partial \bar{M}_{ij}} - C_{ij}^{\circ} \right) + \overline{\mathrm{ALR}}^{\circ} = 0 \tag{10.24}$$

where $\overline{\mathrm{ALR}}$ is the *aggregate land rent,* which can be derived from the aggregate differential land rent by setting agricultural rent equal to zero. This is the Henry George rule for an optimal club complex. Namely, the aggregate operating deficits from the provision of the LPGs are equal to the aggregate land rent. The deficit is the difference between the aggregate user charge and the cost of providing the LPG. Moreover since the opportunity cost of land has been set equal to zero, urban rent vanishes at the boundaries of the optimal complex, that is,

$$R^{\circ}\langle -\bar{x} \rangle = R^{\circ}\langle \bar{x} \rangle = 0. \tag{10.25}$$

A complete characterisation of an optimal complex seems difficult at this level of generality. It is still possible however to say something about the spatial structure of such a complex. Firstly, we know that there is no vacant land within an optimal complex. For otherwise, one could imagine that the length of the urban area is shortened to eliminate the vacant land; and if everything located on occupied land before this operation is kept fixed during the operation, aggregate transport cost must be reduced. Secondly, we know that every sub–area of the urban territory is involved in extensive spatial interaction with other sub–areas in the territory. In particular, as we prove in section 10.4.2 of the appendix, we have

> **Result 10.1** (Hochman, Pines and Thisse (1995)): Given the territory $\mathcal{X}^\circ$ of an optimal complex, there is no strictly smaller sub–area of $\mathcal{X}^\circ$ such that all the trips made in this sub–area both originate and end there.

Finally, we can specify the optimal location principle for clubs within the complex. Observe that the location of a club $\bar{x}_{i,2j}$ appears only in the material balance constraint (10.20(d)), so that its effect on utility can be exerted only through this constraint. More precisely, club location in (10.20(d)) only determines the corresponding aggregate transport cost associated with that club. Hence a necessary condition for optimality is to chose a club location that minimises this aggregate transport cost, which implies

$$\int_{\bar{x}_{i,2j-1}}^{\bar{x}_{i,2j}} n^\circ t_i\left(\left|\bar{x}_{i,2j}-x\right|\right) dx = \int_{\bar{x}_{i,2j}}^{\bar{x}_{i,2j+1}} n^\circ t_i\left(\left|x-\bar{x}_{i,2j}\right|\right) dx$$

$$\text{for } i \in \{1,...,\bar{n}\} \text{ and } j \in \{1,...,\bar{\kappa}_i\}. \tag{10.26}$$

The intuitive interpretation of (10.26) is straightforward. By relocating the club $\bar{x}_{i,2j} - x$ meters rightward, all those individuals living to its left have to pay $t_i\left(\bar{x}_{i,2j} - x\right)$ more for travelling to the facility, while all those to its right save $t_i\left(x - \bar{x}_{i,2j}\right)$ each. Minimising transport cost requires that the marginal cost of relocation is just equal to the corresponding marginal benefit.

10.2.3 Decentralisation

It is readily verified from (10.24) that the total cost of providing the LPGs must be larger that the total revenue from the optimal user charges. Therefore at least some clubs cannot be sustainable if their revenue is limited to the user charge. In fact, the excess of total cost over total revenue holds not only for the optimal complex as a whole, but for every club in it. To see this observe that the user charge exceeds cost only if the average cost $C_{ij}/\bar{M}_{ij}$ is increasing in $\bar{M}_{ij}$, which is impossible at the optimum. For if a club operates at increasing average cost then it is inefficient because a larger number of clubs of the same type can serve the same population at a lower cost. Furthermore, with more clubs of the same type, aggregate transport cost will be reduced. Even at the minimum average provision cost the number of clubs is still too small because transportation cost, which is strictly increasing with club size, must be added to the club's provision cost. We conclude that *each club needs to be partially financed by land rent.*

The Henry George rule (10.24) implies that the above deficit problem can be resolved on the aggregate level by using the appropriate lump–sum transfers. But although such transfers can be decentralised by the appropriate allocation of land ownership to the various clubs, the appropriate incentive structure cannot. For suppose that any given club ij is entitled to, say, a share θ_{ij} (< 1) of land in its market area, which allows it to obtain θ_{ij} of the corresponding $\overline{\text{ALR}}$. Then, since the social benefit (or cost) related to any change of the club's policy is fully reflected by the land values inside its market area, the club cannot realize the full marginal impact of its policy on the welfare of its patrons as a marginal profit (or loss)—but only as a share θ_{ij} of it.[17] This distorts its decision making away from optimality.[18] We conclude that decentralisation of an optimal complex must be undertaken at the level of the urban territory itself, rather than at the level of individual clubs inside the territory as suggested by fiscal federalism.

10.3 Monocentricity Versus Polycentricity

As we explained in section 10.2.2, it is difficult to characterise in detail the spatial structure of an optimal club complex at the level of generality adopted in that section. However questions such as what characterises the transition from a monocentric to a polycentric urban structure, or what determines whether the location of a public facility will be central or peripheral, are fundamental to an understanding of the city.

In this section we provide an answer to these questions by using a drastically simplified model of an optimal club complex. As in Fujita (1989) and Thisse and Wildasin (1992), we restrict both the number of non–residential activities and the number of urban configurations. We allow only two types of *pure* LPGs in the complex, we assume that a *given* quantity $\bar{G}_i$ for $i = 1, 2$ is necessary for survival and that the marginal utility of G_i vanishes for $G_i > \bar{G}_i$. In this manner we fix the consumption of LPGs in the complex. We also preclude multi–product clubs (see, for example, Brueckner and Lee (1991)). It then follows that the aggregate cost of supplying a given type of LPG depends solely on the number of clubs that provide it: scale economies can only be realised by reducing the number of clubs. Moreover, we restrict the number of clubs to $\bar{\kappa}_i \leq 2$. Thus, if we denote the setup cost of a type–i facility by $\bar{C}_i$, its corresponding aggregate cost of provision is either $\bar{C}_i$ or $2\bar{C}_i$. Finally, we allow only two configurations. The first, *monocentric configuration*, contains two facilities (one for each type) both located at the mid–point of the interval. The second, *polycentric configuration*, contains three facilities such that two of the same type are symmetrically located

[17] See the discussion in Hochman, Pines and Thisse (1995).

[18] The problem is similar to the impossibility of finding a pay–off formula (sharing rule) for a team which both exhausts the common output and, at the same time, provides the appropriate incentive for efficiency. (See Holmstrom (1982)).

at a distance on either side of the mid–point and a single one of the other type is centrally located. Under both configurations transport cost is linear, $T_i = \bar{A}_i x$.

Turning now to individual behaviour, we assume that the elasticity of substitution between land and the composite good is zero as in section 2.3.5. Given the supply of essential LPGs, the utility level is determined by $\bar{U} = \min\{Z, H/\bar{B}\}$ where $\bar{B}$ is a parameter. Accordingly, equations (10.21) reduce to

$$\begin{aligned} Z &= \bar{U} \quad (a) \\ H &= \bar{B}\bar{U}. \quad (b) \end{aligned} \tag{10.27}$$

In this simple case, where the complementarity hypothesis of Haig (1926) still holds, consumption is the same for every individual—hence residential density is the same across locations (but it changes with parameters).

10.3.1 Monocentric Configuration

We have $\bar{\kappa}_i = 1$ for $i = 1, 2$ and we know that both facilities are centrally located. Using (10.27) and taking into account that there is no vacant land, (10.20(b)) can be written as

$$\frac{2\bar{x}^\circ}{\bar{B}\bar{U}^\circ} = \bar{M}^\circ \tag{10.28}$$

at the optimum. Since the collective goods are pure, there is no user charge and the optimal budget equation (10.23) reduces to

$$\bar{U}^\circ + R^\circ \bar{B}\bar{U}^\circ = \bar{\Omega} - \left(\bar{A}_1 + \bar{A}_2\right)|x| \text{ for } |x| \leq \bar{x}^\circ. \tag{10.29}$$

At the RHS boundary of the urban area (10.29) simplifies into

$$\bar{U}^\circ = \bar{\Omega} - \left(\bar{A}_1 + \bar{A}_2\right)\bar{x}^\circ \tag{10.30}$$

because the optimal rent vanishes there. Combining (10.29) and (10.30) gives

$$R^\circ = \frac{1}{\bar{B}\bar{U}^\circ}\left(\bar{A}_1 + \bar{A}_2\right)|\bar{x}^\circ - x| \text{ for } |x| \leq \bar{x}^\circ. \tag{10.31}$$

Upon integration of (10.31) we get the aggregate land rent

$$\begin{aligned} \overline{\text{ALR}}^\circ &= \int_{\mathcal{X}^\circ} R^\circ dx \\ &= \frac{1}{\bar{B}\bar{U}^\circ}\left(\bar{A}_1 + \bar{A}_2\right)\bar{x}^{\circ 2}. \end{aligned} \tag{10.32}$$

We can therefore write the Henry George rule as

$$\frac{1}{\bar{B}\bar{U}^\circ}\left(\bar{A}_1 + \bar{A}_2\right)\bar{x}^{\circ 2} = \bar{C}_1 + \bar{C}_2. \tag{10.33}$$

Finally, combining (10.30) with (10.33) implies

$$\left(\bar{\Omega}-\bar{U}^{\circ}\right)^{2}=\left(\bar{A}_{1}+\bar{A}_{2}\right)\bar{B}\left(\bar{C}_{1}+\bar{C}_{2}\right)\bar{U}^{\circ} \tag{10.34}$$

which is a quadratic equation in $\bar{U}^{\circ}$. We then can use (10.30) once again to solve for $\bar{x}^{\circ}$. With $\bar{U}^{\circ}$ and $\bar{x}^{\circ}$ known, $\bar{M}^{\circ}$ and R° can be determined from (10.28) and (10.31) respectively.

10.3.2 Polycentric Configuration

We have $\bar{\kappa}_1 = 1$ and $\bar{\kappa}_2 = 2$, and we know that the type–one facility is centrally located while the type–two facilities are located symmetrically on either side and at a distance from the centre. We restrict the exposition to the RHS of the complex, so that $x \geq 0$. Applying (10.26) to the second type yields

$$2\bar{x}_{2,4}^{\circ}=\bar{x}^{\circ} \tag{10.35}$$

which implies that the two peripheral facilities are located at $-\bar{x}^{\circ}/2$ and $\bar{x}^{\circ}/2$, that is, at the mid–points of their respective market areas. It follows that the four urban subregions determined by the two boundaries and the three centres are equally populated. The optimal budget equation (10.23) now reduces to

$$\bar{U}^{\circ}+R^{\circ}\bar{B}\bar{U}^{\circ}=\begin{cases}\bar{\Omega}-\bar{A}_{1}x+\bar{A}_{2}\left(\bar{x}_{2,4}^{\circ}-x\right) & \text{for } x\leq\bar{x}_{2,4}^{\circ}\\[2ex] \bar{\Omega}-\bar{A}_{1}x+\bar{A}_{2}\left(x-\bar{x}_{2,4}^{\circ}\right) & \text{for } x>\bar{x}_{2,4}^{\circ}.\end{cases} \tag{10.36}$$

At the RHS boundary of the urban area (10.36) simplifies into

$$\bar{U}^{\circ}=\bar{\Omega}-\bar{A}_{1}\bar{x}^{\circ}+\bar{A}_{2}\left(\bar{x}^{\circ}-\bar{x}_{2,4}^{\circ}\right). \tag{10.37}$$

Combining (10.36) and (10.37) gives

$$R^{\circ}=\begin{cases}\dfrac{1}{\bar{B}\bar{U}^{\circ}}\left(\left(\bar{A}_{1}+\bar{A}_{2}\right)\bar{x}^{\circ}-2\bar{A}_{2}\bar{x}_{2,4}^{\circ}+\left(\bar{A}_{2}-\bar{A}_{1}\right)x\right) & \text{for } x\leq\bar{x}_{2,4}^{\circ}\\[2ex] \dfrac{1}{\bar{B}\bar{U}^{\circ}}\left(\bar{A}_{1}+\bar{A}_{2}\right)\left(\bar{x}^{\circ}-x\right) & \text{for } x>\bar{x}_{2,4}^{\circ}.\end{cases} \tag{10.38}$$

Taking into account symmetry, we integrate (10.38) to obtain

$$\begin{aligned}\overline{\mathrm{ALR}}^{\circ} &= 2\left(\int_{0}^{\bar{x}_{2,4}^{\circ}}R^{\circ}dx+\int_{\bar{x}_{2,4}}^{\bar{x}^{\circ}}R^{\circ}dx\right)\\ &= \frac{1}{\bar{B}\bar{U}^{\circ}}\left(\left(\bar{A}_{1}+\bar{A}_{2}\right)\bar{x}^{\circ}-2\bar{A}_{2}\left(\bar{x}_{2,4}^{\circ}\right)^{2}\right).\end{aligned} \tag{10.39}$$

Thus the Henry George rule in this case can be written as

$$\frac{1}{\bar{B}\bar{U}^{\circ}}\left(\left(\bar{A}_{1}+\bar{A}_{2}\right)\bar{x}^{\circ}-2\bar{A}_{2}\left(\bar{x}_{2,4}^{\circ}\right)^{2}\right)=\bar{C}_{1}+2\bar{C}_{2}. \tag{10.40}$$

Finally, equations (10.28), (10.37) and (10.40) can be reduced to

$$\left(\bar{\Omega}-\bar{U}^{\circ}\right)^{2}=\left(\bar{A}_{1}+\frac{\bar{A}_{2}}{2}\right)\bar{B}\left(\bar{C}_{1}+2\bar{C}_{2}\right)\bar{U}^{\circ} \tag{10.41}$$

which, as in the monocentric case, it is a quadratic equation in $\bar{U}^{\circ}$. We then can use (10.35), (10.37) and (10.41), to obtain $\bar{x}^{\circ}$ and $\bar{x}^{\circ}_{2,4}$. With $\bar{U}^{\circ}, \bar{x}^{\circ}$ and $\bar{x}^{\circ}_{2,4}$ known, $\bar{M}^{\circ}$ and R° can be determined from (10.28) and (10.38) respectively.

10.3.3 Optimal Configuration

In this subsection we investigate conditions for monocentricity versus polycentricity and, in the case where the latter occurs, which of the two facility types is centrally located. Using (10.34) and (10.41), we have

$$\begin{aligned}\left(\bar{\Omega}-\bar{U}^{\circ}_{(1)}\right)^{2}-\left(\bar{\Omega}-\bar{U}^{\circ}_{(3)}\right)^{2} &= \left(\bar{A}_{1}+\bar{A}_{2}\right)\bar{B}\left(\bar{C}_{1}+\bar{C}_{2}\right)\bar{U}^{\circ}_{(1)} \\ &- \left(\bar{A}_{1}+\frac{\bar{A}_{2}}{2}\right)\bar{B}\left(\bar{C}_{1}+2\bar{C}_{2}\right)\bar{U}^{\circ}_{(3)} \\ &= \left(\bar{A}_{1}+\bar{A}_{2}\right)\bar{B}\left(\bar{C}_{1}+\bar{C}_{2}\right)\left(\bar{U}^{\circ}_{(1)}-\bar{U}^{\circ}_{(3)}\right) \\ &+ \frac{\bar{A}_{1}\bar{B}\bar{C}_{1}}{2}\left(\frac{\bar{A}_{2}}{\bar{A}_{1}}-\frac{2\bar{C}_{2}}{\bar{C}_{1}}\right)\bar{U}^{\circ}_{(3)}\,.\end{aligned} \tag{10.42}$$

If $\bar{U}^{\circ}_{(1)} < \bar{U}^{\circ}_{(3)}$ then the LHS of (10.42) must be positive, which implies that the second term on the RHS of (10.42) must also be positive because the corresponding first term is negative. Similarly, if $\bar{U}^{\circ}_{(1)} > \bar{U}^{\circ}_{(3)}$ then the second term on the RHS of (10.42) must be negative. Therefore when a facility of the first type is located at the mid–point of the polycentric configuration, the optimal configuration is determined by

$$\bar{U}^{\circ}_{(1)} \geq (\leq)\,\bar{U}^{\circ}_{(3)} \Leftrightarrow \frac{\bar{A}_{2}}{\bar{A}_{1}} \leq (\geq)\,\frac{2\bar{C}_{2}}{\bar{C}_{1}}\,. \tag{10.43}$$

Taking into account the symmetry between the two LPGs, we also conclude that

$$\bar{U}^{\circ}_{(1)} \geq (\leq)\,\bar{U}^{\circ}_{(3)} \Leftrightarrow \frac{\bar{A}_{2}}{\bar{A}_{1}} \geq (\leq)\,\frac{\bar{C}_{2}}{2\bar{C}_{1}} \tag{10.44}$$

when a facility of the second type is located at the mid–point of the polycentric configuration. Combining these two conditions, we arrive at

Result 10.2: The optimal configuration is monocentric with two facilities (one for each type) both located at the mid–point of the optimal complex if and only if

$$\frac{\bar{C}_{1}}{2\bar{C}_{2}} < \frac{\bar{A}_{1}}{\bar{A}_{2}} < \frac{2\bar{C}_{1}}{\bar{C}_{2}}\,.$$

If the ratio of transportation rates is outside these bounds then the optimal configuration is polycentric with a single facility located at the mid–point of the optimal complex and two facilities of the other type symmetrically located on either side of the centre at the mid–point of their respective market areas. In particular, the inequalities

$$\frac{\bar{A}_1}{\bar{A}_2} < \frac{\bar{C}_1}{2\bar{C}_2} \text{ or } \frac{\bar{A}_1}{\bar{A}_2} > \frac{2\bar{C}_1}{\bar{C}_2}$$

determine respectively whether facility type one or facility type two is the central facility in the polycentric configuration.

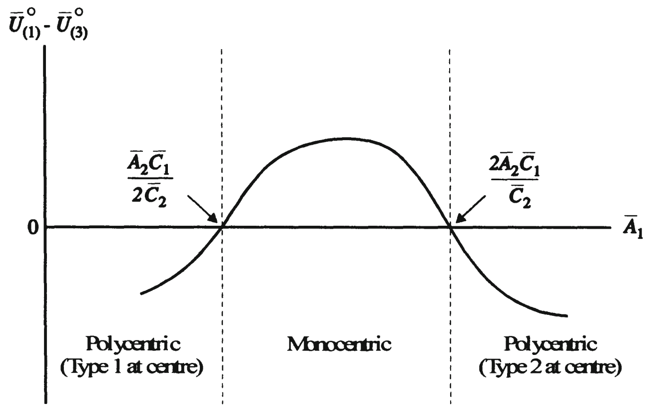

FIGURE 10.4. Determination of Monocentricity Versus Polycentricity.

Our result is illustrated in figure 10.4. Notice that the establishment of each alternative optimal configuration depends only upon the relative values of transportation rates and provision costs. In particular, multiplying $\bar{A}_1$ and $\bar{A}_2$ or $\bar{C}_1$ and $\bar{C}_2$ by the same constant will not affect the choice of an optimal configuration.

10.3.4 Comparative Statics

In this section we discuss the economic mechanisms that produce result 10.2 using the comparative statics of a polycentric optimal configuration. If we combine (10.28), (10.37) and (10.40) we derive

$$\bar{x}^{\circ 2} + \bar{B}\left(\bar{C}_1 + 2\bar{C}_2\right)\bar{x}^{\circ} - \frac{\Omega\bar{B}\left(\bar{C}_1 + 2\bar{C}_2\right)}{\bar{A}_1 + \bar{A}_2/2} = 0. \tag{10.45}$$

Upon inspection of (10.45), we can determine the parameter effects on the optimal size of urban land use as shown on the first row of table 10.2. The second row of that table follows from (10.41). For example, an increase of any

TABLE 10.2. Comparative Statics of a Polycentric City.

	$\bar{\Omega}$	$\bar{A}_1$	$\bar{A}_2$	$\bar{B}$	$\bar{C}_1$	$\bar{C}_2$
$\bar{x}^{\circ}$	+	−	−	+	+	+
$\bar{U}^{\circ}$	+	−	−	−	−	−
$\bar{M}^{\circ}$	−	−	−	−	+	+

parameter on the RHS of (10.41) cannot raise utility because the RHS would increase while the LHS would decrease. For the third row, the effects of $\bar{A}_1$ and $\bar{A}_2$ on $\bar{M}^{\circ}$ can be determined from the first two rows together with (10.28), (10.35) and (10.37). The effects of $\bar{\Omega}$ and $\bar{B}$ follow from the first row together with (10.28) and (10.35). Finally, the effects of $\bar{C}_1$ and $\bar{C}_2$ can be derived using the first two rows together with (10.28) and (10.35).

In order to characterise more precisely the parametric effects on the optimal urban structure, we draw upon result 10.2 and table 10.2 . Because of symmetry, we need only examine the parametric effects for a single type. We begin with figure 10.5, which illustrates the way the transport rate $\bar{A}_1$ affects both the optimal configuration and the length of the urban area—hence, by (10.35), the spacing of centres. The curve segments that correspond to a polycentric city determine the distance between two centres that provide a different type of collective good. For the range over which the optimal configuration is monocentric, both types collapse to the mid–point of the interval $[-\bar{x}^{\circ}, \bar{x}^{\circ}]$ as indicated in the figure.

Figure 10.5 implies that, with the remaining parameters fixed, the monocentric city structure can prevail only if the transport rate to one of the facilities is neither too large nor too small, given the size of the remaining parameters. The following is an intuitive explanation of this result.

A. When $\bar{A}_1$ is zero it pays to have a system of polycentric cities, each with a single facility of the first type at the centre. Suppose not. Then either we have a system of polycentric cities, each with a single facility of the second type at the centre, or a system of monocentric cities. In the first case it pays to replace the two facilities of the first type by one located at the centre and save on the cost of a type–one facility without any increase in transportation costs. In the second case we can combine two adjacent cities, and replace two facilities of the first type located at the centres of the two cities with a single type–one facility at the border between the two adjacent cities. Once again, the cost of one type–one facility is saved without any increase in transport costs.

B. When $\bar{A}_1$ is sufficiently large, a polycentric city with a type–two facility at the centre prevails. Suppose not. Then either the efficient structure implies polycentric cities, each with a facility of the first type at the centre, or monocentric cities. We will show that neither can prevail. First, consider the net benefit of two type–two facilities at the periphery rather than one facility at the city

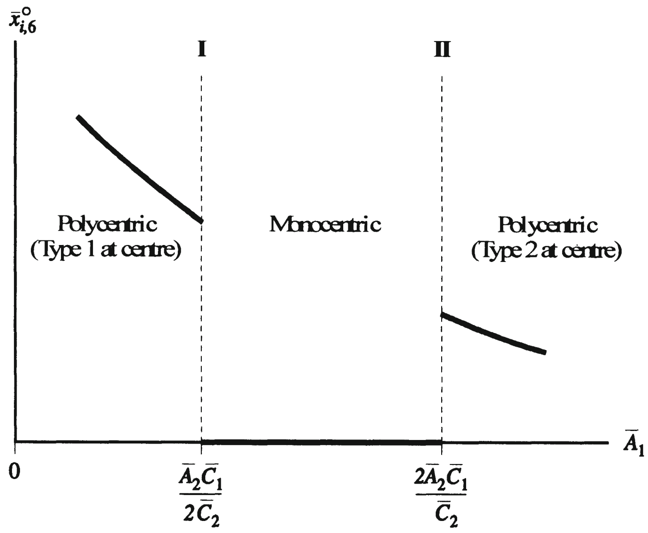

FIGURE 10.5. The Effect of Transport Rate on the Spatial Extent and the Optimal Configuration of the City.

centre, namely, $\bar{A}_2\bar{x}^{\circ}/4 - \bar{C}_2/\bar{M}^{\circ}$. The first term is the saving in transportation cost which results from dispersing the facilities, whereas the second is the cost of an additional facility. As $\bar{A}_1$ increases, both $\bar{x}^{\circ}$ and $\bar{M}^{\circ}$ decrease (see table 12.2), and $\bar{x}^{\circ}$ must tend to zero as $\bar{A}_1$ increases beyond any bounds (see (10.37)). Hence, for sufficiently large $\bar{A}_1$, $\bar{A}_2\bar{x}^{\circ}/4 - \bar{C}_2/\bar{M}^{\circ}$ becomes negative. It follows that since the saving of $\bar{C}_2/\bar{M}^{\circ}$ exceeds the extra transportation cost $\bar{A}_2\bar{x}^{\circ}/4$, it pays to replace the two facilities of the second type at the periphery by one facility located at the centre. Thus a monocentric structure dominates the polycentric structure with a facility of the first type at the centre. However, as $\bar{A}_1$ continues to increase, even $\bar{A}_2\bar{x}^{\circ}/2 - \bar{C}_2/2\bar{M}^{\circ}$ becomes eventually negative, so that $\bar{C}_2/\bar{M}^{\circ} > \bar{A}_2\bar{x}^{\circ}$. But then, consider two adjacent monocentric cities. They constitute one city with both facilities dispersed and population $2\bar{M}^{\circ}$. The net benefit of replacing two facilities of the second type from the centres of the two cities by one facility of the second type located at the border between the cities is $\bar{A}_2\bar{x}^{\circ}/2 - \bar{C}_2/2\bar{M}^{\circ} > 0$. We thus obtain a polycentric city with a facility of the second type at the centre. In summary, for sufficiently large $\bar{A}_1$, a monocentric city dominates a polycentric city with type–one facility at the centre, and is dominated by a polycentric city with a type–two facility at the centre.

Analogous arguments can be applied to setup costs. Similarly to the case of transportation rates, both small and large setup costs induce polycentricity. However, using table 10.2, we know that the effect of setup costs on the size of the city is opposite to the effect of transport rates.

In the simple case we presented here all centres locate on the middle of their market area. Since the population is evenly distributed, it is also true that exactly half of the consumption group locates on either side of any centre. Therefore, when demand is insensitive to prices, all centres locate at both their *geographical* and *population* midpoints. In the more general case where demand is sensitive to prices, these two become distinct and centres locate on their population midpoints. Since density declines away from the main centre, subcentral population midpoints locate closer to the main centre than before. Higher transport rates, which tend to concentrate population around the main centre, may induce a strong centripetal movement of subcentres. In particular, using a Cobb–Douglas utility function, Loay Alemi of Tel–Aviv University found that increasing $\bar{A}_1$ to its first critical value (point I in figure 10.5) decreases $\bar{x}^{\circ}_{2,6}/\bar{x}^{\circ}$, that is, the centres converge to each other not only in absolute but also in relative terms; and when $\bar{A}_1$ increases beyond its second critical value, the centres diverge. In other words, as the system moves from a polycentric toward a monocentric configuration, the subcentres converge until a critical point is reached and the city becomes monocentric.

Conventional wisdom says that a fall in transport cost (stemming from interactions between individuals and firms) *encourages* agglomeration (see Fujita and Thisse (1996)). This conclusion is supported by several studies (see Ogawa and Fujita (1980), Imai (1982), Fujita (1989) and Fujita, Thisse and Zenou (1997)). Our results are at odds with this conclusion. For example, an increase in home–based travel costs induces agglomeration when they are low, first by the convergence of the non–residential activities and then by their concentration in the geographical centre. On the contrary, when these costs increase further, they contribute to a dismantling of the unique centre and to the dispersion of the non–residential activities. Finally, when non–residential activities are dispersed, any further increase in the travel costs once again induces convergence of the centres—although they remain separate.[19] The main reason for the difference is the distinct nature of the two non–residential activities and the symmetry between them. As the cost of interaction between individuals and one activity increases, the centrally located activity eventually changes. This characteristic is missing in the above studies, where the same type of activity is always located at the centre. However, if our setup costs are taken to represent the centripetal force, our results regarding setup costs are consistent with Fujita and Ogawa (1982), where the cost effect of inter–firm interactions on agglomeration is non–monotonic, and full agglomeration requires an intermediate level of such cost. Likewise, our results are reminiscent of Mori (1997) where full agglomeration occurs only under an intermediate level of transportation cost for the manufacturing goods (stemming from interactions between individuals and firms).

[19] To see the difference between our results and the ones reported in the existing literature consider, for example, Fujita, Thisse, and Zenou (1997). We can interpret the firm's cost of distance from the center (localisation economies) as our scale economies. Then their Proposition 1 implies a monotone relationship between the location of the entrant and commuting cost on the one hand, and localisation economies on the other (distance from the main center increases with commuting cost and decreases with localisation economies).

10.4 Appendices

10.4.1 Optimality of the Club Equilibrium and Decentralisation

In this section we prove the first and second welfare theorems for the club equilibrium (10.8).

> I. Let $(G^e, \bar{\Upsilon}^e, \bar{G}^e, \bar{M}^e, \bar{\kappa}^e)$ and $P\left[\bar{\Upsilon}^e, \bar{G}^e\right]$ be a club equilibrium. Then there is no other feasible allocation that yields higher utility than $(G^e, \bar{\Upsilon}^e, \bar{G}^e, \bar{M}^e, \bar{\kappa}^e)$.

PROOF: Suppose not. Then there is a feasible allocation $(G^1, \bar{\Upsilon}^1, \bar{G}^1, \bar{M}^1, \bar{\kappa}^1)$ such that[20]

$$u\left[\bar{\Omega} - c\left[\bar{\Upsilon}^1, \bar{G}^1\right]/\bar{M}^1, G^1, \bar{\Upsilon}^1, \bar{G}^1\right] > u\left[\bar{\Omega} - c\left[\bar{\Upsilon}^e, \bar{G}^e\right]/\bar{M}^e, G^e, \bar{\Upsilon}^e, \bar{G}^e\right], \tag{10.46}$$

while (10.8(a)), (10.8(d)) and (10.8(e)) imply

$$u\left[\bar{\Omega} - c\left[\bar{\Upsilon}^e, \bar{G}^e\right]/\bar{M}^e, G^e, \bar{\Upsilon}^e, \bar{G}^e\right] \geq u\left[\bar{\Omega} - P\left[\bar{\Upsilon}^1, \bar{G}^1\right]G^1, G^1, \bar{\Upsilon}^1, \bar{G}^1\right]. \tag{10.47}$$

Taken together, (10.46) and (10.47) yield

$$P\left[\bar{\Upsilon}^1, \bar{G}^1\right]G^1 > c\left[\bar{\Upsilon}^1, \bar{G}^1\right]/\bar{M}^1. \tag{10.48}$$

However, using (10.8(b)) and (10.8(e)), we conclude that

$$P\left[\bar{\Upsilon}^1, \bar{G}^1\right]G^1 \leq c\left[\bar{\Upsilon}^1, \bar{G}^1\right]/\bar{M}^1, \tag{10.49}$$

a contradiction. □

> II. Let $(G^\circ, \bar{\Upsilon}^\circ, \bar{G}^\circ, \bar{M}^\circ, \bar{\kappa}^\circ)$ be a club optimum. Then there exists a supporting price $\bar{P}^\circ = P\left[\bar{\Upsilon}, \bar{G}\right]$ under which the club optimum is a club equilibrium.

PROOF: We shall argue that $P\left[\bar{\Upsilon}, \bar{G}\right] = c\left[\bar{\Upsilon}, \bar{G}\right]/\bar{G}$ is a supporting price function. We know that the equilibrium conditions (10.8(b)–(e)) are satisfied by the definition of a club optimum and the above supporting price function. Therefore we only have to establish that condition (10.8(a)) is satisfied if we replace $(G^e, \bar{\Upsilon}^e, \bar{G}^e)$ with $(G^\circ, \bar{\Upsilon}^\circ, \bar{G}^\circ)$. Suppose not. Then there is a $(G^1, \bar{\Upsilon}^1, \bar{G}^1, \bar{M}^1)$ such that

$$u\left[\bar{\Omega} - P\left[\bar{\Upsilon}^1, \bar{G}^1\right]G^1, G^1, \bar{\Upsilon}^1, \bar{G}^1\right] > u\left[\bar{\Omega} - c\left[\bar{\Upsilon}^\circ, \bar{G}^\circ\right]/\bar{M}^\circ, G^\circ, \bar{\Upsilon}^\circ, \bar{G}^\circ\right]. \tag{10.50}$$

Using $P\left[\bar{\Upsilon}, \bar{G}\right] = c\left[\bar{\Upsilon}, \bar{G}\right]/\bar{G}$ and (10.8(d)) on (10.50), we obtain

$$u\left[\bar{\Omega} - c\left[\bar{\Upsilon}^1, \bar{G}^1\right]/\bar{M}^1, G^1, \bar{\Upsilon}^1, \bar{G}^1\right] > u\left[\bar{\Omega} - c\left[\bar{\Upsilon}^\circ, \bar{G}^\circ\right]/\bar{M}^\circ, G^\circ, \bar{\Upsilon}^\circ, \bar{G}^\circ\right] \tag{10.51}$$

---

[20] A feasible allocation satisfies the material balance condition $\bar{M}Z + \bar{C} = \bar{M}\bar{\Omega}$, which implies $Z = \bar{\Omega} - \bar{C}/\bar{M}$.

which implies that $\left(G^{\circ}, \bar{\Upsilon}^{\circ}, \bar{G}^{\circ}, \bar{M}^{\circ}, \bar{\kappa}^{\circ}\right)$ is not optimal because the allocation $\left(G^{1}, \bar{\Upsilon}^{1}, \bar{G}^{1}, \bar{G}^{1}/G^{1}\right)$, which allows for a higher utility is feasible—a contradiction. □

10.4.2 Proof of Result 10.1

The proof is taken from Hochman, Pines and Thisse (1995). Suppose that there is a sub-area $\mathcal{X}$ strictly smaller that $\mathcal{X}^{\circ}$ and such that all trips that originate in $\mathcal{X}$ also end inside it. Then either (i) $\mathcal{X}$ is an autarchy in the sense that the resources supplied in $\mathcal{X}$ equal the resources used, or (ii) $\mathcal{X}$ is not an autarchy.

In the first case, the optimal utility level must be attained in a club complex that satisfies (10.19) and (10.20). This however is impossible because the total population in $\mathcal{X}$ is smaller that the optimal population.

In the second case, resources either flow from $\mathcal{X}^{\circ} - \mathcal{X}$ into $\mathcal{X}$ or from $\mathcal{X}$ into $\mathcal{X}^{\circ} - \mathcal{X}$. If the first possibility applies, the original allocation is not optimal because a utility level higher than $\bar{U}^{\circ}$ can be achieved in $\mathcal{X}^{\circ} - \mathcal{X}$ by not exporting resources into $\mathcal{X}$.[21] A similar argument eliminates the second possibility.

10.5 References

R. J. Arnott, 1979, Optimal City Size in a Spatial Economy. *Journal of Urban Economics* 6, 65-89.

R. J. Arnott and J. E. Stiglitz, 1979, Aggregate Land Rent, Expenditure on Public Goods, and the Optimal City Size. *Quarterly Journal of Economics* 93, 471-500.

V. Barham and M. H. Wooders, 1998, First and Second Welfare Theorems for Economies with Collective Goods. In D. Pines, E. Sadka, and I. Zilcha (eds.), *Topics in Public Economics*. Cambridge University Press, New York, 57–88.

E. Berglas, 1976, On the Theory of Clubs. *American Economic Review* 66, 116–121.

E. Berglas and D. Pines, 1980, Clubs as a Case of Competitive Industry with Goods of Variable Quality. *Economics Letters* 5, 363-366.

E. Berglas and D. Pines, 1981, Clubs, Local Public Goods, and Transportation Models: A Synthesis. *Journal of Public Economics* 15, 141–162.

J. K. Brueckner and K. Lee, 1991, Economies of Scope and Multiproduct Clubs. *Public Finance Quarterly* 19, 193-209.

J. M. Buchanan, 1965, An Economic Theory of Clubs. *Economica* 32, 1-14.

J. P. Conley and M. H. Wooders, 1998, Anonymous Pricing in Tiebout Economies with Clubs. In D. Pines, E. Sadka, and I. Zilcha (eds.), *Topics in Public Economics*. Cambridge University Press, New York, 89–120.

F. Flatters, J. V. Henderson and P. Mieszkowski, 1974, Public Goods, Efficiency and Regional Fiscal Equalization. *Journal of Public Economics* 3, 99–112.

[21] Eliminating $\mathcal{X}$ releases additional resources into $\mathcal{X}^{\circ} - \mathcal{X}$ without affecting the initial allocation in any other respect.

M. Fujita, 1989, *Urban Economic Theory.* Cambridge University Press, Cambridge.

M. Fujita and H. Ogawa, 1982, Multiple Equilibria and Structural Transition of Non–Monocentric Urban Configurations. *Regional Science and Urban Economics* 12, 161–196.

M. Fujita, J.-F. Thisse and Y. Zenou, 1997, On the Endogenous Formation of Secondary Employment Center. *Journal of Urban Economics* 41, 337–357.

H. George, 1896, *Progress and Poverty.* Reprinted by the Robert Schalkenbach Foundation, New York, 1954.

R. P. Gilles and S. A. Scotchmer, 1998, Decentralization in Club Economies: How Multiple Private Goods Matter. In D. Pines, E. Sadka, and I. Zilcha (eds.), *Topics in Public Economics.* Cambridge University Press, New York, 121–138.

R. M. Haig, 1926, Toward an Understanding of the Metropolis. *Quarterly Journal of Economics* 40, 421–423.

O. Hochman, 1981, Land Rents, Optimal Taxation and Local Fiscal Independence in an Economy with Local Public Goods. *Journal of Public Economics* 15, 59–85.

O. Hochman, 1982*a*, Congestable Local Public Goods in an Urban Setting. *Journal of Urban Economics* 11, 290-310.

O. Hochman, 1982*b*, Clubs in an Urban Setting. *Journal of Urban Economics* 12, 85-101.

O. Hochman, D. Pines and J.-F. Thisse, 1995, On the Optimal Structure of Local Governments. *The American Economic Review* 85, 1224–1240.

B. Holmstrom, 1982, Moral Hazard in a Team. *Bell Journal of Economics* 13, 324–340.

H. Imai, 1982, CBD Hypothesis and the Economics of Agglomeration. *Journal of Economic Theory* 28, 275–299.

T. Malthus, 1798, *An Essay on the Principle of Population as it Affects the Future Improvement of Society.* Reprinted by Pelican, London, 1970.

T. Mori, 1997, A Modelling of Megalopolis Formation: The Maturing of City Systems. *Journal of Urban Economics* 42, 133–157.

W. E. Oates, 1972, *Fiscal Federalism.* Harcourt Brace Jovenovich, New York.

H. Ogawa and M. Fujita, 1980, Equilibrium Land Use Pattern in a Non–Monocentric City. *Journal of Regional Science* 20, 455–475.

M. Olson Jr., 1969, The Principle of 'Fiscal Equivalence': The Division of Responsibility Among Different Levels of Government. American Economic Review, Papers and Proceedings 59, 479–487.

D. Ricardo, 1821, *Principles of Political Economy and Taxation.* Reprinted by George Bell & Sons Ltd, London, 1891.

S. Rosen, 1974, Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy* 82, 34–55.

P. A. Samuelson, 1954, A Pure Theory of Public Expenditure. *Review of Economics and Statistics* 36, 387–389.

S. A. Scotchmer, 1994, Public Goods and the Invisible Hand. In J. Quigley and E. Smolensky (eds.) *Modern Public Finance.* Harvard University Press, Cambridge, Mass., 93-125.

S. A. Scotchmer and M. H. Wooders, 1987, Competitive Equilibrium and the Core in Club Economies with Anonymous Crowding. *Journal of Public Economics* 34, 159–174.

D. M. Starrett, 1988, *Foundations of Public Economics.* Cambridge University Press, Cambridge, Mass.

J. E. Stiglitz, 1977, The Theory of Local Public Goods. In Martin Feldstein and Robert P. Inman (eds.) *The Economics of Public Services.* Macmillan, London, 274–333.

J.-F. Thisse and D. E. Wildasin, 1992, Public Facility Location and Urban Spatial Structure: Equilibrium and Welfare Analysis. Journal of Public Economics 48, 83–118.

C. M. Tiebout, 1956, A Pure Theory of Local Public Goods. *Journal of Political Economy* 64, 416–424.

M. H. Wooders, 1978, Equilibria, the Core, and Jurisdiction Structures in Economies with a Local Public Good. *Journal of Economic Theory* 18, 328–348.

11
Specialisation and Trade

In this chapter we investigate the reasons behind the formation of diverse cities by identical individuals when there are more than a single private good—a structure that leads to specialisation and trade. We derive a model that combines Wilson (1987) with Gilles and Scotchmer (1997, 1998), and then reduce it to a simple optimisation problem with two variables: the relative prices of the private goods and the urban population size. This reduced form sheds new light on the advantages and disadvantages of choosing a diversified city system with specialisation and trade over a system of identical autarchic cities. Which of the two is preferable depends on the size of gains from trade in a diversified system relative to gains from an efficient city size in a system of identical autarchies. We present several examples to illustrate different possible outcomes of the model. The issue of decentralisation is discussed in Papageorgiou and Pines (1998).

11.1 Introduction

As we have explained in section 10.1.1, when the population consists of identical individuals and there is a *single private good*, it is optimal to form *identical cities* of *optimal size* for undertaking collective activities such as the production of goods or the consumption of collective goods. The optimal size of each city depends on the trade–off between the advantage of sharing the cost of producing the collective good and the disadvantage of congested production or consumption. Since there is only one tradeable private good (collective goods are not traded), cities are autarchic. But once there is more than one tradeable good, there are several reasons for specialisation and trade among cities.

The standard approach for specialisation and trade in urban economics argues that the main agglomeration advantage is coming from localisation economies (scale economies internal to the firm or the industry)—rather than urbanisation economies resulting from positive inter–industry externalities.[1] If this were the case, it would be always more efficient to establish single–product cities or, when firms have strong input–output relationships and the shipment of intermediate goods is costly, a single industrial complex.[2] The production of different goods by different cities implies specialisation and trade.

In this chapter we present another argument for specialisation and trade based on the theory of clubs, which has markedly different implications than the argument based on localisation economies. We develop a framework that integrates Wilson (1987) with Gilles and Scotchmer (1997, 1998) who extended models akin to that of section 10.1.1 by introducing more than a single private good. We are thus able to account simultaneously for the effect of production (a key element in Wilson) and for the effect of an impure collective good (a key element in Gilles and Scotchmer). Along with these authors, we find that the optimal urban structure for *identical individuals* and more than a single private good is not necessarily a system of identical autarchies, but may form diversified city types differing from each other by size, production mix and consumption.

The underlying premises of the model we use to explain the advantage of specialisation and trade differ from those of Henderson (1997). In contrast to Henderson, where production exhibits scale economies, production here exhibits constant returns to scale and diminishing labour productivity. As in chapter ten, the main agglomeration advantage arises from the collective use of the public facilities and the cost–sharing it affords.

The main difference between the implications of the model presented in this chapter and Henderson's (1997) model can be illustrated by using his example. Suppose we begin with an autarchic city which produces both food and steel. Based on the standard reasoning for specialisation and trade, Henderson will recommend that this city be replaced by two separate, *smaller* than the original cities that specialise on a single type of product. Based on club and local public good theories, Wilson (1987) and Gilles and Scotchmer (1997, 1998) will also recommend that this city be replaced by two separate specialised cities. However, unlike Henderson, one of these cities should be *smaller* while the other should be *larger* than the original.

---

[1] See Isard (1956) and Henderson (1988, 1997). Another explanation comes from the new economic geography, which uses the differentiated products approach to represent scale economies (see, for example, Krugman (1991, 1995). As explained in chapter nine, since under this approach every individual consumes all varieties and since production of these varieties is dispersed among cities, each city specialises on a segment of the interval that determines the range of the differentiated product. Therefore each city imports varieties it does not produce and exports varieties it produces.

[2] As Henderson (1997, p.594) expresses it, "...if we take two cities, one specialized in food production and the other in iron and steel and combine them, everyone loses."

11.2 An Extended Club–LPG Model

We use a version of the model proposed by Stiglitz (1977), generalised to incorporate a crowding effect both on the utility and on the provision cost of the collective good.[3] There is a large number of potential sites where cities can locate, each site extending over one unit of land. A smaller number κ of these potential sites have actually been developed into cities accommodating a total population $\bar{N}$ of identical individuals.[4] The welfare of an individual in city $j = 1, ..., \kappa$ is determined by a well-behaved utility function $u\left[Z_{1j}, Z_{2j}, \Upsilon_j, \bar{M}_j\right]$ where Z_{ij} $(i = 1, 2)$ denotes the *individual consumption* of the two private goods; Υ_j is the *amount of a congested local public good supplied*; and $\bar{M}_j$ is the *population size of the city.* All goods are essential. The public good is *congested* in the sense that the benefit from and the cost of its provision changes with the city's population size. It is *local* in the sense that only residents of city j can benefit from its provision. Finally, excessive crowding is intolerable because $\lim_{\bar{M}_j \to \infty} \partial u / \partial \bar{M}_j = -\infty$.

Each of the two private goods is produced using labour and land according to a linear homogeneous production function $f_i\left[\bar{M}_{ij}, L_{ij}\right]$ where $\bar{M}_{ij}$ and L_{ij} denote the *labour* and *land inputs used to produce private good* i in city j. The two production functions are distinct in the sense that their isoquant slopes differ for the same input combination. The total quantity available from the first private good in each city is used either for direct consumption or for the production of the local public good. Since Z_1 is the numéraire, the *cost of producing the public good* is measured in units of Z_{1j} as $C_j = c\left[\Upsilon_j, \bar{M}_j\right]$.[5] As in the case of utility, excessive crowding is intolerable because $\lim_{\bar{M}_j \to \infty} C_j / \bar{M}_j = \infty$.

A *feasible allocation* $\left(Z_{1j}, Z_{2j}, \Upsilon_j, \bar{M}_j, \bar{M}_{1j}, L_{1j}\right)$ satisfies the population and material balance constraints:

$$\sum_{j=1}^{\kappa} \bar{M}_j = \bar{N} \qquad (a)$$

[3] As in section 10.1, geographical space does not play a rôle in this chapter. Since we do not need to distinguish between those variables that vary with location and those which do not, we simplify notation. Therefore absence of a bar in this chapter does not imply that the corresponding variable varies with location.

[4] We refrain from the more general case of an heterogeneous population with differentiated tastes and initial endowments, as in Gilles and Scotchmer (1997, 1998), in order to highlight the specific implications of introducing multiple private goods for specialisation and trade. A model with heterogeneous individuals can blur the surprising implication that trade may be advantageous even among identical individuals (gains from trade among heterogeneous individuals is never surprising).

[5] Observe that in this formulation we have simplified the model of chapter ten by assuming that the frequency of visits to the collectively used facility is fixed.

$$\begin{aligned}\sum_{j=1}^{\kappa}\left(\bar{M}_j Z_{1j} + C_j - f_1\left[\bar{M}_{1j}, L_{1j}\right]\right) &= 0 \quad (b)\\ \sum_{j=1}^{\kappa}\left(\bar{M}_j Z_{2j} - f_2\left[\bar{M}_j - \bar{M}_{1j}, 1 - L_{1j}\right]\right) &= 0. \quad (c)\end{aligned} \tag{11.1}$$

A *feasible equal–treatment allocation* is a feasible allocation in which utility is equalised across cities. Finally an *optimal allocation* is a feasible equal–treatment allocation that maximises the common utility level.

We assume that an optimal urban system can contain either one city type (as in the case of identical autarchies in section 10.1.1) or, at most, two types. In the former case the urban system is *uniform*, in the latter *diversified*. If we denote by a and b the *city types* in a diversified system, and by κ_a and κ_b the corresponding *total number of cities* then we can define an *optimum* as a pair of allocations $(Z^{\circ}_{1j}, Z^{\circ}_{2j}, \Upsilon^{\circ}_j, \bar{M}^{\circ}_j, \bar{M}^{\circ}_{1j}, L^{\circ}_{1j})$ for $j = a, b$ and a partition of the κ cities into types $(\kappa^{\circ}_a, \kappa^{\circ}_b | \, \kappa^{\circ}_a + \kappa^{\circ}_b = \kappa)$ that maximise the common utility level $\bar{U}$ subject to

$$\begin{aligned} u\left[Z_{1j}, Z_{2j}, \Upsilon_j, \bar{M}_j\right] &= \bar{U} \text{ for } j = a, b \quad (a)\\ \sum_{j=a,b} \kappa_j \bar{M}_j &= \bar{N} \quad (b)\\ \sum_{j=a,b} \kappa_j\left(\bar{M}_j Z_{1j} + C_j - f_1\left[\bar{M}_{1j}, L_{1j}\right]\right) &= 0 \quad (c)\\ \sum_{j=a,b} \kappa_j\left(\bar{M}_j Z_{2j} - f_2\left[\bar{M}_j - \bar{M}_{1j}, 1 - L_{1j}\right]\right) &= 0 \quad (d)\end{aligned} \tag{11.2}$$

where (11.2(a)) represents the equal–utility constraints, (11.2(b)) is the population constraint and (11.2(c), (d)) are the material balance constraints.

The first–order conditions of (11.2) imply that all cities must be self–sufficient at the optimum:

$$\bar{M}^{\circ}_j Z^{\circ}_{1j} + C^{\circ}_j - f_1\left[\bar{M}^{\circ}_{1j}, L^{\circ}_{1j}\right] + \left(\bar{M}^{\circ}_j Z^{\circ}_{2j} - f_2\left[\bar{M}^{\circ}_j - \bar{M}^{\circ}_{1j}, 1 - L^{\circ}_{1j}\right]\right) P^{\circ} = 0. \tag{11.3}$$

where P is the implicit price of the second private good so that the optimal $P^{\circ} \equiv \text{MRS}\langle Z^{\circ}_{1j}, Z^{\circ}_{2j}\rangle$. They also imply the Samuelson rule

$$\bar{M}^{\circ}_j \cdot \text{MRS}\left\langle Z^{\circ}_{1j}, \Upsilon^{\circ}_j\right\rangle = \frac{\partial c^{\circ}}{\partial \Upsilon_j} \tag{11.4}$$

and the Henry George rule

$$\text{ALR}^{\circ} = C^{\circ}_j - \bar{M}^{\circ}_j\left(\frac{\partial c^{\circ}}{\partial \bar{M}_j} - \bar{M}^{\circ}_j \cdot \text{MRS}\left[Z^{\circ}_{1j}, \bar{M}^{\circ}_j\right]\right) \tag{11.5}$$

for $j = a, b$.[6] In this version, the Henry George rule says that the provision cost net of the warranted user charges (which represents the deficit associated with the provision of the local public good in the city) is equal to the aggregate land rent in that city at the optimum.

Given our initial assumptions, a solution to this model can represent either a uniform system of autarchies or a diversified system with two distinct city types. On the one hand an uniform urban system will emerge if either $\kappa_a^\circ \kappa_b^\circ = 0$ or $\left(Z_{1a}^\circ, Z_{2a}^\circ, \Upsilon_a^\circ, \bar{M}_a^\circ\right) = \left(Z_{1b}^\circ, Z_{2b}^\circ, \Upsilon_b^\circ, \bar{M}_b^\circ\right)$. We assume that the optimal allocation is unique and that the maximum utility level at any feasible urban population size has a $\cap$-shape around the optimal population size. On the other hand a diversified urban system is characterised by $\kappa_a^\circ > 0, \kappa_b^\circ > 0$ and $\left(Z_{1a}^\circ, Z_{2a}^\circ, \Upsilon_a^\circ, \bar{M}_a^\circ\right) \neq \left(Z_{1b}^\circ, Z_{2b}^\circ, \Upsilon_b^\circ, \bar{M}_b^\circ\right)$. Since there is only one type of optimal autarchy, cities in a diversified system must trade Z_1 for Z_2.

11.3 A Reduced Form of the Optimal Allocation Problem

In order to illuminate the underlying conditions for the establishment of a uniform or a diversified urban system at the optimum, we reduce the optimal allocation problem to a model which specifies the maximum attainable common utility level at any price P and urban size $\bar{M}$. This will allow us to compare the common utility levels achieved in uniform and feasible diversified urban systems, and to determine under what circumstances one is preferable to the other. We begin by defining a reduced utility function from which the public good is eliminated:

$$\begin{aligned} V\left[\mathbb{Z}_1, Z_2, \bar{M}\right] &= \max_{Z_1, \Upsilon} u\left[Z_1, Z_2, \Upsilon, \bar{M}\right] \quad (a) \\ &\text{subject to} \qquad\qquad (11.6) \\ Z_1 + \frac{C}{\bar{M}} &= \mathbb{Z}_1. \qquad (b) \end{aligned}$$

Having combined private consumption of Z_1 together with the public good Υ in $\mathbb{Z}_1$, we render $V[\cdot]$ compatible to the example provided by Gilles and Scotchmer (1997, 1998). We further reduce $V[\cdot]$ to an indirect utility function

[6] In deriving the Henry George rule, we evaluate per-unit land rent using the value of the marginal product of land:

$$\frac{\partial f_1}{\partial L_{1j}} = P \cdot \frac{\partial f_2}{\partial L_{2j}}.$$

However, each of these expressions is also representing DLR since the quantity of land in each city is unitary.

by defining [7]

$$v\left[P,\bar{M}\right] = \left\{\begin{array}{l} \max\limits_{\mathbb{Z}_1, Z_2, \bar{M}_1, L_1} V\left[\mathbb{Z}_1, Z_2, \bar{M}\right] \\ \text{subject to} \\ \bar{M}\mathbb{Z}_1 - f_1\left[\bar{M}_1, L_1\right] + \\ \left(\bar{M}Z_2 - f_2\left[\bar{M} - \bar{M}_1, 1 - L_1\right]\right) P = 0 \end{array}\right\} =$$

$$\max_{Z_2, \bar{M}_1, L_1} V\left[\frac{1}{\bar{M}}\left(f_1\left[\bar{M}_1, L_1\right] + \left(f_2\left[\bar{M} - \bar{M}_1, 1 - L_1\right] - \bar{M}Z_2\right)P\right), Z_2, \bar{M}\right]. \tag{11.7}$$

The solution to (11.7) yields $Z_2\left[P,\bar{M}\right], \bar{M}_1\left[P,\bar{M}\right], L_1\left[P,\bar{M}\right]$ and, therefore, . Furthermore, using (11.6), we can also derive $\mathbb{Z}_1\left[P,\bar{M}\right]$ and $\Upsilon\left[P,\bar{M}\right]$. With this information one can determine

$$v\left[P,\bar{M}\right] = V\left[\mathbb{Z}_1\left[P,\bar{M}\right], Z_2\left[P,\bar{M}\right], \bar{M}\right] \tag{11.8}$$

which specifies the maximum utility level a self–sufficient city can attain for any combination of price and population size.

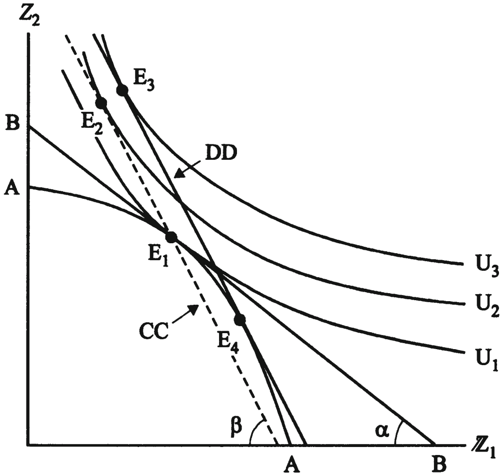

FIGURE 11.1. Gains from Trade.

The mapping $v\left[P,\bar{M}\right]$ can be explained in terms of figure 11.1. Given some $\bar{M} = \hat{M}$, the per–capita production possibility frontier of the city is uniquely determined and represented by the locus AA in that figure. Suppose that the

[7] A standard indirect utility function is defined over prices and income. In this formulation the supply of land is fixed, hence output and income are fully determined by P and $\bar{M}$.

city confronts a price level $\hat{P} = -1/\tan\beta$. Then the consumption possibility frontier is the locus DD, where E_3 represents the optimal consumption bundle and E_4 the optimal production bundle. Therefore $v\left[\hat{P}, \hat{M}\right]$ is the utility level which corresponds to the indifference curve U_3.

In order to characterise $v\left[P, \bar{M}\right]$, we first discuss combinations of P and $\bar{M}$ which imply an autarchy. The utility–maximising consumption and production bundles of an autarchic city with population size $\hat{M}$ is E_1, which allows for a utility level that corresponds to the indifference curve U_1. The consumption and production bundles E_1 will also be chosen if a city with the same population size $\hat{M}$ confronts a price level $\bar{P} = -1/\tan\alpha$ and is free to trade. For, in that case, its consumption possibility frontier becomes BB so that the optimal production and consumption bundles coincide at E_1. Thus, given any $\bar{M}$, we have just determined $\bar{P}$ which satisfies

$$\begin{aligned} \bar{M}\mathbb{Z}_1\left[\bar{P}, \bar{M}\right] &= f_1\left[\bar{M}_1\left[\bar{P}, \bar{M}\right], L_1\left[\bar{P}, \bar{M}\right]\right] & (a) \\ \bar{M}Z_2\left[\bar{P}, \bar{M}\right] &= f_2\left[\bar{M} - \bar{M}_1\left[\bar{P}, \bar{M}\right], 1 - L_1\left[\bar{P}, \bar{M}\right]\right]. & (b) \end{aligned} \quad (11.9)$$

We denote this mapping from $\bar{M}$ to P by $\bar{P} = \bar{P}\left[\bar{M}\right]$ and refer to it as an *autarchy line* on the P—$\bar{M}$ plane. The slope of the autarchy line depends on whether $\mathbb{Z}_1$ and $\bar{M}$ are substitutes or complements, and whether $\mathbb{Z}_1$ is labour–intensive relative to Z_2. For example, if $\mathbb{Z}_1$ and $\bar{M}$ are neither substitutes nor complements and $\mathbb{Z}_1$ is labour–intensive, an increase in labour supply $\bar{M}$ reduces its relative price, that is, it increases $\bar{P}$. Moreover if $\mathbb{Z}_1$ is neither labour– nor land–intensive but $\mathbb{Z}_1$ and $\bar{M}$ are substitutes, an increase in $\bar{M}$ reduces the $\text{MRS}\langle Z_2, \mathbb{Z}_1\rangle$—hence it increases once again $\bar{P}$. In both examples the autarchy line slopes upwards on the P—$\bar{M}$ plane. Since, in the present context, only the absolute value of the slope of the autarchy line matters, we shall assume that $\bar{P}\left[\bar{M}\right]$ is upward–sloping unless stated otherwise.

We next discuss how gains from trade can arise in the context of figure 11.1. Suppose that, given $\bar{M}$, the price decreases from $-\tan a$ to $-\tan\beta$, which allows for an achievable utility level that corresponds to the higher indifference curve U_3. This increase follows from two effects. First, keeping the production unchanged, consumption is adjusted from E_1 to E_2 on CC which allows for a move from a lower indifference curve U_1 to a higher indifference curve U_2. A further improvement can be accomplished by adjusting production from E_1 to E_4 which raises the consumption possibility frontier from CC to DD and, consequently, allows a move to a higher indifference curve U_3. Since the consumption bundle at E_3 is different from the corresponding production bundle at E_4, the city is no longer autarchic. More precisely, starting from an autarchy, if the city is allowed to trade Z_1 for Z_2 at a price *lower* than $\bar{P}$ then the achievable utility can be increased. Likewise, starting once again from an autarchy, it can be seen that if the city is allowed to trade Z_2 for Z_1 at a price *higher* than $\bar{P}$ then the achievable utility can also be increased.[8] Hence, keeping $\bar{M}$ fixed, any deviation from $\bar{P}$ allows gains from trade. This is precisely the case of a consumer who owns a given amount of two goods (or gets his or her income in kind). Since $\bar{P}$ is the consumer's $\text{MRS}\langle Z_2, \mathbb{Z}_1\rangle$ at the endowment point, any other price ratio allows the consumer to be better–off. It follows that the autarchy line on the P—$\bar{M}$ plane is a trough line.

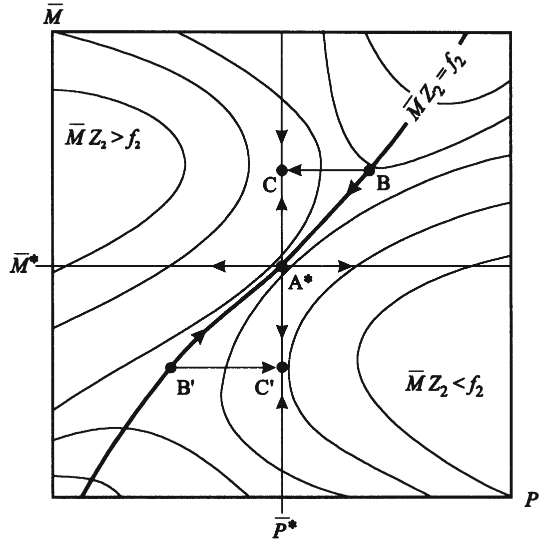

FIGURE 11.2. Conditions for a Diversified Urban System.

Although gains from trade are necessary for a specialised urban system, they are not sufficient. We shall see that this happens because specialisation necessitates an urban size other than the optimal size of an autarchy. Since utility decreases away from the optimal size, the choice between a uniform and a diversified urban system hinges upon whether or not the losses from inefficient population size are greater than the corresponding gains from trade. A case where gains from trade dominate over losses from inefficient size is shown in figure 11.2. This figure includes a map of $v\left[P, \bar{M}\right]$ and the autarchy line on the P—$\bar{M}$ plane. The autarchy line slopes upward and is a trough line as explained above. Arrows point toward higher achievable utility. The optimal autarchy corresponds to $\left(\bar{P}^*, \bar{M}^*\right)$, where utility is maximised over the autarchy line. Since autarchic utility $v\left[\bar{P}\left[\bar{M}\right], \bar{M}\right]$ has a $\cap$–shape around its optimal size $\bar{M}^*$, we conclude that $\left(\bar{P}^*, \bar{M}^*\right)$ is a saddle point.

Along the autarchy line (11.9) is satisfied. Starting from any point on the autarchy line, if we decrease $\bar{P}$ while keeping population fixed then production

[8]To see this just exchange the notation on the axes such that $P^* = -\tan[1/\alpha]$ and the price increases to $-\tan[1/\beta]$.

and consumption will adjust to $\bar{M}\mathbb{Z}_1 < f_1[\cdot]$ and $\bar{M}Z_2 > f_2[\cdot]$ as it has been discussed in conjunction with figure 11.1. Similarly, a movement of the relative price toward the opposite direction implies $\bar{M}\mathbb{Z}_1 > f_1[\cdot]$ and $\bar{M}Z_2 < f_2[\cdot]$. Therefore any combination $(P, \bar{M})$ *to the right* of the autarchy line in figure 11.2 implies trading Z_1 for Z_2, and any combination *to the left* of the autarchy line implies trading Z_2 for Z_1. Consequently, specialisation and trade among cities requires that one city type has a population larger than the optimal autarchy size $\bar{M}^*$ while the other's population is smaller. This discussion clarifies the notion of a trade–off between gains from trade and losses from inefficient city size mentioned above.

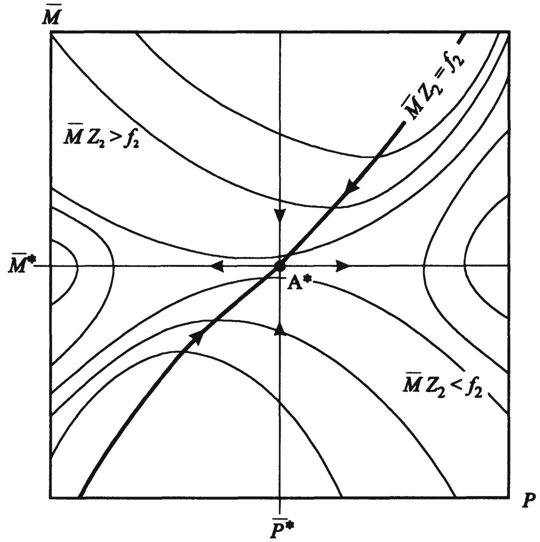

FIGURE 11.3. Conditions for a Uniform Urban System.

In order to establish the possibility of a diversified urban system we need to further assume that, for any given P, $\lim_{\bar{M}\to 0} v\left[P, \bar{M}\right] = \lim_{\bar{M}\to\infty} v\left[P, \bar{M}\right] < v\left[\bar{P}^*, \bar{M}^*\right]$.[9] Then it is true that, as you go from north to south or vice–versa on figure 11.2, you encounter at least one ridge line. If, further, $v\left[\bar{P}^*, \bar{M}\right]$ assumes a local minimum at $\left(\bar{P}^*, \bar{M}^*\right)$ as it does in figure 11.2 then gains from trade dominate and a diversified system does emerge. The reason is that, under these circumstances, there exists some *price range* $[P_{\min}, P_{\max}]$ containing $\bar{P}^*$ such that, for any price in that range, $v\left[P, \bar{M}\right]$ attains local maxima both at some finite city size *above* the line of autarchy and at some positive city size *below* it. For $P = \bar{P}^*$, these local maxima are represented by points C and C′ in figure 11.2. Therefore, over $[P_{\min}, P_{\max}]$, there are two ridge sections with elevations

[9]These assumptions can be justified if one takes into account that a minimal level of collective good, say health care, is necessary; and if the negative crowding effect dominates when the population increases beyond any bound.

higher than $v\left[\bar{P}^*, \bar{M}^*\right]$, one located above and the other below the autarchy line. This property can be used to determine the optimal city sizes $\bar{M}_a^\circ, \bar{M}_b^\circ$ and the optimal implicit price P° in the case of a diversified urban system. By contrast, if $v\left[\bar{P}^*, \bar{M}\right]$ assumes a local maximum at $\left(\bar{P}^*, \bar{M}^*\right)$, as it does in figure 11.3, trade is unprofitable and we observe a uniform system of identical autarchies.

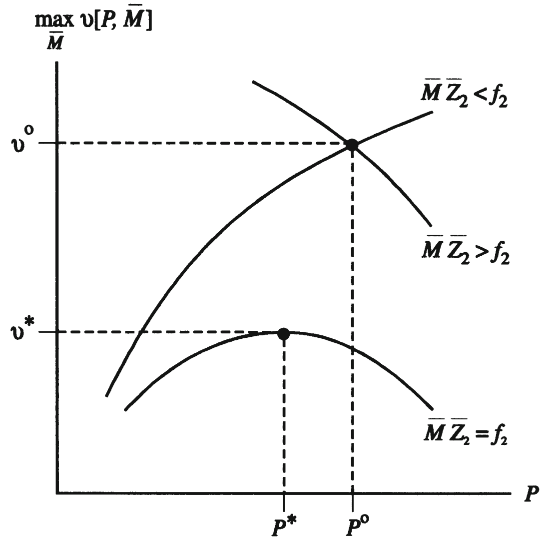

FIGURE 11.4. Determination of a Diversified–System Equilibrium.

Figure 11.4, which corresponds to figure 11.2, illustrates the determination of $\bar{M}_a^\circ, \bar{M}_b^\circ$ and P°, where $\max_{\bar{M}} v\left[P^\circ, \bar{M}\right] = v\left[P^\circ, \bar{M}_a^\circ\right] = v\left[P^\circ, \bar{M}_b^\circ\right]$. Denote the population size on the autarchy line, corresponding to any arbitrary P, by $\bar{P}^{-1}[P]$. Then imagine that, starting from $P_{\min}$, we increase price gradually over its entire range $[P_{\min}, P_{\max}]$. Let $\bar{M}_a$ be the maximand of $v\left[P, \bar{M}\right]$ subject to $\bar{M}_a > \bar{P}^{-1}[P]$ and $\bar{M}_b$ be the maximand of $v\left[P, \bar{M}\right]$ subject to $\bar{M}_b < \bar{P}^{-1}[P]$. These constrained maximisation problems generate two ridge lines. Suppose one represents $\bar{M}_a$ above the autarchy line such that $\bar{M}_a Z_{2a} > f_2[\cdot]$, and the other represents $\bar{M}_b$ below the autarchy line such that $\bar{M}_b Z_{2b} < f_2[\cdot]$. Accordingly, in figure 11.4, the line labelled $\bar{M}Z_2 > f_2[\cdot]$ denotes the achievable utility in city a, and the line labelled $\bar{M}Z_2 < f_2[\cdot]$ denotes the achievable utility in city b. The utility achievable under autarchy is labelled $\bar{M}Z_2 = f_2[\cdot]$. By inspecting figure 11.2 we observe that, when P is increased from $P_{\min}$ to $P_{\max}$, the constrained maximum utility level for type b increases from a level below $v\left[P^*, \bar{M}^*\right]$ to one above $v\left[P^*, \bar{M}^*\right]$, while the constrained maximum utility level for type a decreases from a level above $v\left[P^*, \bar{M}^*\right]$ to one below $v\left[P^*, \bar{M}^*\right]$. Therefore the two loci representing the constrained maxima must intersect at some (P°, v°). There, $\bar{M}_a^\circ$ and $\bar{M}_b^\circ$ are obtained as the two solutions for $\bar{M}$ of $v^\circ = v\left[P^\circ, \bar{M}\right]$.

In general, a pair of trading cities as defined above do not satisfy the material balance constraints of an optimal allocation. However, once the optimal levels of $\bar{M}_a^\circ, \bar{M}_b^\circ$ and P° have been determined, we can calculate the total population of an optimal economy that will allow feasible trade of Z_2. Following Gilles and Scotchmer (1997, 1998), we solve for the ratio κ_a/κ_b in

$$\sum_{j=a,b} \kappa_j \left(\bar{M}_j^\circ Z_{2j}\left[P^\circ, \bar{M}_j^\circ\right] - f_2\left[\bar{M}_j - \bar{M}_{1j}\left[P^\circ, \bar{M}_j^\circ\right], 1 - L_{1j}\left[P^\circ, \bar{M}_j^\circ\right]\right]\right) = 0. \tag{11.10}$$

If this ratio is a rational number then we can use it in conjunction with $\kappa_a^\circ + \kappa_b^\circ = \kappa$ to obtain κ_a° and κ_b°. It follows that an urban system with population $\bar{N}^\circ = \kappa_a^\circ \bar{M}_a^\circ + \kappa_b^\circ \bar{M}_b^\circ$ will allow feasible trade of Z_2. Since the market for Z_1 is also cleared by the Walras law, $\bar{N}^\circ, \kappa_a^\circ$ and κ_b° together with $\bar{M}_a^\circ, \bar{M}_b^\circ$ and P° determine an optimal urban system.

11.4 Gains from Trade Versus the Loss from Inefficient City Size

In the previous section we argued that if $v\left[\bar{P}^*, \bar{M}\right]$ assumes a local minimum at $\left(\bar{P}^*, \bar{M}^*\right)$ as in figure 11.2 then the optimal system is diversified, and if it assumes a local maximum as in figure 11.3 then it is uniform. The question is, when does $v\left[\bar{P}^*, \bar{M}\right]$ assume a maximum and when does it assume a minimum? Using the envelope theorem, one realises that $\left(\bar{P}^*, \bar{M}^*\right)$ is either a local extremum or an inflection point of $v\left[\bar{P}^*, \bar{M}\right]$. In what follows we disregard the second possibility and investigate the factors that determine whether $\left(\bar{P}^*, \bar{M}^*\right)$ is a maximum or a minimum of $v\left[\bar{P}^*, \bar{M}\right]$.

Consider a shift from A to C in figure 11.2, which represents a departure from the unique optimal autarchy to a more populated city that can trade Z_2 at a price $\bar{P}^*$. This shift can be decomposed in two: first, a shift from A to B which still leaves the city autarchic but with a larger population $\bar{M}_C$; second, a shift from B to C which leaves the population unchanged but reduces the relative price from P_B to $\bar{P}^*$. The shift from A to B causes a decline in $v\left[\bar{P}, \bar{M}\right]$ because of a deviation from the optimal autarchic size. We define this decrease as *the loss from inefficient city size.* The shift from B to C, which we define as *gains from trade,* causes an increase in $v\left[\bar{P}, \bar{M}\right]$.[10]

Beginning with the loss from inefficient city size, we can distinguish three effects associated with the shift from A to B. First, the utility is directly affected by the increased crowding. The size of this effect is equal to $\left(\partial u/\partial \bar{M}\right)\Delta\bar{M}$, where the derivative is evaluated somewhere between A and B. Second, the average product of labour in producing both Z_1 and Z_2 is reduced as labour intensity rises. Third, the per-capita cost of the public good may increase or

[10] Similar observations apply to the shift from A to C′.

decrease, depending on how net crowding affects the average provision cost $C/\bar{M}$.[11]

We now turn to gains from trade. These are associated with the shift from B to C in figure 11.2 and they follow from adjusting production and consumption to the change in P.[12] The total benefit (the aggregate consumer and producer surplus), derived from a marginal change in the relative price of Z_2, is obtained by differentiating (11.7) with respect to P:

$$\frac{\partial v}{\partial P}=\frac{\partial V}{\partial \mathbb{Z}_1}\left[P,\bar{M}\right]\left(\frac{f_2\left[\cdot\right]}{\bar{M}}-Z_2\left[P,\bar{M}\right]\right). \tag{11.11}$$

The LHS of (11.11) vanishes at B on the autarchy line and it is negative between B and C. Thus gains from trade can be calculated by integrating the negative value of the RHS of (11.11) over the interval CB as [13]

$$\begin{aligned}-\int_{\bar{P}*}^{P_{\mathrm{B}}}\frac{\partial v}{\partial P}dP &= \int_{\bar{P}*}^{P_{\mathrm{B}}}\frac{\partial V}{\partial \mathbb{Z}_1}\left[P,\bar{M}_{\mathrm{C}}\right]\left(Z_2\left[P,\bar{M}_{\mathrm{C}}\right]-\frac{f_2\left\langle\bar{M}_{\mathrm{C}}\right\rangle}{\bar{M}_{\mathrm{C}}}\right)dP \\ &> 0\end{aligned} \tag{11.12}$$

since the RHS of (11.12) is always negative between C and B. However, the net effect of losses and gains cannot be ascertained a priori.

11.5 Examples

11.5.1 Consumption and Production Groups with a Pure Local Public Good

An important case in which gains from trade always dominate over the loss from inefficient city size is provided by Wilson (1987). According to his basic specification, the crowding effect on the cost function of the collective good is either ignored or assumed to be small. As a consequence, the crowding effects associated with a shift from A to B in figure 11.2 are restricted to a decreasing average productivity of labour and, perhaps, a *moderate* decline in utility. These adverse effects can always be more than offset by the gains from trade associated with the shift from B to C. In order to demonstrate this assertion,

[11] Observe that no crowding effect is imposed on the utility or on the per capita provision cost. The only requirements are that the three crowding effects combined establish $\bar{M}^*$ as a unique optimum to $v\left[\bar{P}^*\left[\bar{M}^*\right],\bar{M}^*\right]$ and that their net impact on the utility of an autarchic jurisdiction is positive for populations smaller than $\bar{M}^*$ and negative for populations larger than $\bar{M}^*$.

[12] Recall that gains form trade result from the *change* in P rather than from its *decrease*. For gains from trade are also realised when the price increases from B′ to C′ in figure 11.2.

[13] See Varian (1984, pp. 263–266) for a discussion about the change in consumer surplus associated with a change in the price of a single commodity.

let us evaluate the outcome of a direct shift from A to C. As long as there is no full specialisation, the Rybzynski theorem of international trade tells us that such a shift leaves both the wage rate $\left(\partial f_2/\partial \bar{M}\right) P$ and the aggregate land rent $(\partial f_2/\partial L_2) P$ unchanged. It follows that, under Wilson's specification, the gains from trade associated with modifying the production mix allow the same consumption pattern $\left(Z_1\left[\bar{P}^*, \bar{M}^*\right], Z_2\left[\bar{P}^*, \bar{M}^*\right], \Upsilon\left[\bar{P}^*, \bar{M}^*\right]\right)$ at C as at A in figure 11.2.[14] However, the attainable utility at C must exceed the corresponding utility at A. This is true because the consumption bundle $(Z_1^*, Z_2^*, \Upsilon^*)$ is inefficient for a population of size $\bar{M}_C$. The reason is that Samuelson's rule (11.4), which holds for the optimal autarchic population $\bar{M}^*$, must be violated for the larger population $\bar{M}_C$. Hence, starting from the bundle $(Z_1^*, Z_2^*, \Upsilon^*)$, a preferable allocation can be achieved by adjusting consumption to satisfy Samuelson's rule. This explains why the homogenous community structure in Wilson (1987) can never be efficient, and why an efficient allocation does require that each community specialises in the production of one private good only.[15] Nevertheless, once we remove the restrictions about crowding on the utility and cost functions, Wilson's results change so that partial specialisation, or even autarchy, become admissible structures for an optimal allocation.

11.5.2 Consumption and Production Groups with an Impure Local Public Good

Introducing impure local public goods can alter Wilson's (1987) results because the Rybzynski theorem is no longer sufficient to guarantee the superiority of a diversified system of cities. This theorem asserts that, under our assumptions, both the aggregate land rent and the wage rate do not change with population size. If the public good were pure, we know from (11.5) that the aggregate land rent would be sufficient to finance it for different city sizes. Then, as population changes, it would still be feasible to support the same level of private consumption with the fixed wage rate and the same level of the local public good with the fixed aggregate land rent in both type–a and type–b cities. By contrast, with an impure local public good, although the aggregate land rent and the wage rate remain the same as the population changes, there are other effects which imply that the original utility level may no longer be feasible. In particular, the larger city suffers from increased congestion while the smaller city enjoys some relief from congestion. It follows that the simple reasoning of the previous section, which was based on the Rybzynski theorem, does not apply any longer. But although we cannot specify theoretically the nature of an optimal urban system using the framework of section 11.5.1, the numerical example of table 11.1 indicates that if the public good is impure then both uniform and diversified systems are candidates for an optimum.

---

[14] Recall that there is no significant crowding effect on the provision cost.

[15] As long as specialisation is incomplete, the Rybzynski theorem applies so that further improvements are possible.

TABLE 11.1. Example with an Impure Local Public Good

$u[\cdot]$	$f_1[\cdot]$	$f_2[\cdot]$	$c[\cdot]$
$Z_1^{1/3} Z_2^{1/3} (\Upsilon - 10)^{1/3}$	$\bar{M}_1$	$\left(\bar{M} - \bar{M}_1\right)^{1/2}$	$\Upsilon + .01\bar{M}^2$

In this example Z_1 is produced only from labour while the production of Z_2 uses all the land available, so that $f_2\left[\bar{M} - \bar{M}_1, 1 - L_1\right] = f_2\left[\bar{M} - \bar{M}_1, 1\right]$. There are two opposing effects on utility as the urban population grows. On the one hand, utility increases with increasing population because the per–capita burden of the local public good $\Upsilon/\bar{M}$ decreases and because ten units of public good are required at a minimum. On the other hand, utility decreases with increasing population because of diminishing returns in the production of Z_2 and because the cost of producing the public good increases. Our calculations show that these two opposing effects produce a $\cap$–shaped autarchic utility schedule $v\left[\bar{P}\left[\bar{M}\right], \bar{M}\right]$ as expected, and that the function $v\left[P, \bar{M}\right]$ of this example corresponds to figure 11.3.[16] Thus $\bar{M}^*$ does not only represent the optimal size of an autarchy, but it also maximises $v\left[\bar{P}^*, \bar{M}\right]$ for aggregate income achievable through trading at the optimal imputed autarchic price $\bar{P}^*$. We conclude that, in this case, a uniform urban system is optimal. A diversified urban system emerges if the impurity of the public good (reflected by the exponent of $\bar{M}$ in the cost function) is sufficiently reduced. For example, if we re–define the cost function as $c\left[\Upsilon, \bar{M}\right] = \Upsilon + .01\bar{M}^{1.2}$ and leave everything else unchanged then the $\cap$–shaped autarchic utility still holds, but the function $v\left[P, \bar{M}\right]$ of the new model now corresponds to figure 11.2. Thus although $\bar{M}^*$ still maximises $v\left[\bar{P}\left[\bar{M}\right], \bar{M}\right]$, it minimises $v\left[\bar{P}^*, \bar{M}\right]$ so that, in this case, a diversified urban system becomes optimal.

11.5.3 Consumption Groups with a Crowding Effect

Gilles and Scotchmer's (1997, 1998) specification refers to a standard club model, where the production of private goods is replaced by a fixed per–capita endowment $\left(\hat{Z}_1, \hat{Z}_2\right)$.[17] Therefore, while the crowding effect in Wilson (1987) is mainly caused by a diminishing marginal productivity, crowding in Gilles and Scotchmer has a direct impact on utility (which may be attributed to the impurity of an underlying local public good). It turns out that this difference

[16] We can provide our calculations upon request.

[17] At first glance, it appears that their third example does not describe a club economy since there is no collective good. However, as we have shown in section 11.3, a club economy can be reduced to one where only two private goods are consumed.

TABLE 11.2. Examples with a Crowding Effect

	Example 1	Example 2
$\bar{\alpha}$	.50	.50
$\bar{A}$	5000.00	100.00
$\bar{B}$	1.00	1.00
η	2.00	2.00
$\bar{\rho}$	.95	.10
$\hat{Z}_1$	1000.00	1000.0
$\hat{Z}_2$	1000.00	1000.00

is not crucial and, as in Wilson's paper, gains from trade versus the loss from inefficient city size combine once again to determine whether or not a uniform system structure is superior to a diversified one.

We illustrate the above points by using two examples of economies defined in Gilles and Scotchmer (1997, 1998). Both cases use

$$\begin{aligned} u\left[Z_1, Z_2, \Upsilon\right] &= \left(\left(Z_1^{\alpha}\left(\Upsilon - \bar{A}\right)^{1-\bar{\alpha}}\right)^{\bar{\rho}} + Z_2^{\bar{\rho}}\right)^{1/\bar{\rho}} & (a) \\ c\left[\Upsilon, \bar{M}\right] &= \Upsilon + \bar{B}\bar{M}^{\eta}. & (b) \end{aligned} \tag{11.13}$$

The parameters chosen for the examples appear in table 11.2. Observe that those two differ from one–another mainly because their elasticities of substitution are different: whereas in the second example the elasticity of substitution is close to unitary (1.1), in the first it is much higher (20.0). This higher elasticity allows gains from trade to mitigate the crowding effect of increased population size by substituting greater amounts of Z_1 for Z_2 through trade. It also allows agents to tolerate a higher level of crowding relative to that experienced in the optimal autarchy by substituting Z_2 for Z_1 when the population decreases. This explains why, in the first example, specialisation and trade are optimal. In the second example, where the elasticity of substitution is not sufficiently high, the optimal configuration is a system of identical autarchies.[18]

[18]The autarchy lines generated by both examples are decreasing on the P–$\bar{M}$ plane, rather than increasing as in figures 11.1 and 11.2. In those figures, the slope of the autarchy line is dominated by the higher labour intensity of producing Z_1 relative to that of Z_2, which more than offsets the complementarity effect in consumption. Here, where economies have no production, the slope of the autarchy line only reflects the complementarity between population size and Z_1. Therefore, as population increases, the marginal rate of substituting Z_2 for Z_1 also increases.

11.6 References

R. P. Giles and S. A. Scotchmer, 1997, Decentralization in Replicated Club Economies with Multiple Private Goods. *Journal of Economic Theory* 72, 363–387.

R. P. Giles and S. A. Scotchmer, 1998, Decentralization in Club Economies: How Multiple Private Goods Matter. In D. Pines, E Sadka and I. Zilcha (eds.) *Topics in Public Economics.* Cambridge University Press, New York, 89–120.

J. V. Henderson, 1988, *Urban Development.* Oxford University Press, Oxford.

J. V. Henderson, 1997, Medium Size Cities. *Regional Science and Urban Economics* 27, 583–612.

W. Isard, 1956, *Location and Space Economy.* MIT Press, Cambridge, Mass.

P. Krugman, 1991, Increasing Returns and Economic Geography. *Journal of Political Economy* 99, 483–499.

P. Krugman, 1995, *Development, Geography and Economic Theory.* Cambridge University Press, Cambridge.

Y. Y. Papageorgiou and D. Pines, 1998, Specialization and Trade: The Perspective of Club and Local Public Good Theories. Tel Aviv University, (unpublished).

J. E. Stiglitz, 1977, The Theory of Local Public Goods. In M. Feldstein and R. P. Inman (eds.) *The Economics of Public Services.* Macmillan, London, 274–333.

H. R. Varian, 1987, *Micro Economic Analysis.* Norton, New York.

J. D. Wilson, 1987, Trade in a Tiebout Economy. *American Economic Review* 77, 431–441.

12

Externalities, Nonconvexity and Agglomeration

In this chapter we discuss and clarify a number of conceptual problems associated with urban externalities, some of which have prevented a meaningful exchange of ideas in the literature about the distribution of population among cities. We shed new light on an old debate about the direction of market bias caused by external economies and diseconomies, and we emphasise the crucial rôle of nonconvexity in urban resource misallocation. We also demonstrate that whenever we abandon the neat framework of club theory, where it is always possible to distribute the total population among optimal cities, any claim about the direction of market bias is vulnerable because small changes in model specification can induce big changes in model outcomes.

12.1 Introduction

So far, our analysis concerning the distribution of population among cities has utilised a Tiebout–type allocation. This implies that the optimal distribution of population among cities can be achieved by allocating individuals to an urban system where each city has an optimal size—given the type of production it specialises and the socio–economic group to which it provides LPGs. Such an allocation is based on the strong premise the cities are perfectly *divisible* and *replicable*. By divisibility we mean that the urban population can be divided into an integer number of optimal–size cities. By replicability we mean that the number of potential cities is not constrained, and that the minimum resources required to provide a given level of utility to $2\bar{N}$ individuals is no more than twice as much as those required to provide that utility to $\bar{N}$ individuals. Following Lösch (1940) and Tiebout (1956), we have also assumed that the optimal population size for each type of city is finite. Hence the optimal (Tiebout–type) allocation is sustainable by a competitive price system as described in the preceding chapters.

In this chapter we discuss two main causes for the failure of the market to achieve such an optimal allocation. The first cause is the standard resource misallocation resulting from external economies and diseconomies. A classical case for this type of market failure is the inability to collect the toll required to internalise the external diseconomies associated with traffic congestion. The second cause is the infeasibility of establishing the 'right' number of cities due to either indivisibility or non–replicability or both, which we refer to as *nonconvexity*.

If the markets are distorted by externalities and if the appropriate decision–maker (entrepreneur or urban system co–ordinator) is not equipped with all the policy instruments necessary for implementing the optimal allocation then, by definition, the Tiebout–type allocation is suboptimal as illustrated in chapter seven. This however will not concern us here. The main issue of this chapter is about whether or not, given the prevailing market structure and the constrained menu of policy instruments available to the planner, *the allocation can be improved by an urban policy which induces the reallocation of people among cities.* Since the discussion of this issue in the literature is very confused, we first aim to clear some of the conceptual problems involved. In what follows we distinguish several main sources of difficulty, and we elaborate on them in subsequent sections. In section 12.2 we explore the very concept of an externality. This concept is not well–defined in the literature or, more precisely, the economic discipline has not succeeded so far to adopt a standard definition of the term. In consequence, different authors use the term to signify quite different things. In section 12.3 we discuss *external scale diseconomies* and *economies,* and their effect on the distribution of population among cities. Firstly, we define such externalities in terms of a gap between the net *marginal social cost* MSC and the net *marginal private cost* MPC of providing an individual the equilibrium utility. A *positive* gap is defined as external scale *diseconomies* and a *negative* gap as external scale *economies.* Secondly, we show that unpriced transportation congestion results in external scale diseconomies. This is far from tautological since it is not self–evident that, when urban markets are distorted by external diseconomies (such as congested roads and pollution), this distortion is necessarily reflected in a positive gap between the MSC and the MPC of *city size.* Thirdly, we show that even when urban diseconomies (economies) are reflected in external *scale* diseconomies (economies), it is not clear whether or not such external scale diseconomies (economies) induce excessive (insufficient) agglomeration in big cities. This depends on the relationship between the gap and city size, that is, whether the gap increases, decreases or remains unchanged with city size. Finally, in section 12.4, we argue that externalities may not be the only, or even the main, cause of urban market failure. Nonconvexity is far more important than previously thought.

12.2 What is an Externality?

In reviewing the debate about the definition of externalities, Maurice Lagueux (1998, p. 120) states that 'externality' "... is still one of the most vaguely defined concepts of economic theory and one which seems to be a source of considerable embarrassment for economists." This statement applies forcefully in the context of inter–city resource allocation where there is no common language denominator—a prerequisite for any meaningful deliberation of what is perhaps the most important urban policy issue! For example, consider a statement by Krugman (1995, p. 51; italics are ours.): "Suppose that we think of positive local external economies, which tend to *promote concentration* of production, as being opposed by other effects—congestion or land cost—that tend to *promote dispersal.*" Similar views are also expressed by other leading urban economists.[1] Now compare this statement with analogous statements in the well–known debate, lucidly represented by Tolley and Grihfield (1987) on the one hand and Mills and Hamilton (1984) on the other, about inter–city resource allocation. In contrast to Krugman who states that congestion promotes dispersal, Tolley and Grihfield maintain that congestion is bound to *encourage agglomeration*; and in Mills and Hamilton, as well as in Helpman (1998), we learn that external economies are expected to *deter agglomeration*—not to promote concentration as in Krugman. Such inconsistencies support the criticism of Lagueux (1998, p. 120) that economics as a discipline can be accused for its "... inability to define its basic concepts ..." and in which we find an "... extreme variety of incompatible meanings attributed to the notion of 'externality' by eminent economists of various economic orientations."

The above statements about the effect of externalities on agglomeration are in conflict because they are based on different approaches to defining the concept of an 'externality'. Since consistency is important, we shall explicitly adopt a single definition of 'externality' and base all our discussion on this definition. In particular, we shall adopt an approach represented by Arrow (1970, p. 2), who maintains that externalities "... are relative to the mode of economic organization ...", and that the "... problem of externalities is ... the failure of market to exist." (Ibid, p. 17.) A similar approach is also implied by Varian (1978, p. 203) where an externality arises "... when the action of one agent affects the environment of another agent other than by affecting prices." According to the definition of Varian, the very existence of an externality depends on the price system which, in turn, depends on the very existence of a market as asserted by Arrow.[2]

---

[1] Fujita and Thisse (1996) and Anas, Arnott and Small (forthcoming) use analogous terms when discussing the positive effect of external economies on agglomeration.

[2] A second, well known approach defines an externality in technological terms alone: "An externality is said to exist if some of the variables which affect one decision-maker's utility or profits are under the control of another decision-maker." (Gravelle and Rees (1981, p. 509).) Thus an externality arises whenever the choice of an action by one agent appears as an argument in the objective function of another agent. The two approaches differ with respect to

Let us now apply this definition of 'externality' in the context of agglomeration. Assume that the production function of firm i in a given city is $\mathcal{Z}_i = \bar{M}_i \sum_{j \in \bar{m}} \bar{M}_j$, where $\mathcal{Z}_i$ is the *quantity of composite good produced* by firm i, $\bar{M}_i$ is the *number of employees* in firm i and $\bar{m}$ is the *number of firms* in the city. The aggregate output is

$$\sum_{i \in \bar{m}} \mathcal{Z}_i = \left(\sum_{j \in \bar{m}} \bar{M}_j \right)^2 , \tag{12.1}$$

which certainly exhibits *scale economies.* But does production exhibit *external economies*? According to our definition of an externality, no answer can be given before the market structure is specified. We consider two cases. On the one hand, if the production is carried out by a single multi–plant producer, there is no externality involved. On the other hand, if competitive firms take the aggregate employment as a parameter outside their control, there is a positive externality reflected by the difference between the marginal social productivity and the corresponding productivity as perceived by the firms, that is,

$$\frac{\partial}{\partial \bar{M}_i} \left(\sum_{j \in \bar{m}} \bar{M}_j \right)^2 - \sum_{j \in \bar{m}} \bar{M}_j = \sum_{j \in \bar{m}} \bar{M}_j > 0. \tag{12.2}$$

In this case production is characterised by *scale economies* which are *external to the firms.*

Next, consider the volume of equilibrium employment for the two cases above. In order to abstract from the issue of monopoly we assume that, although the demand price decreases with increasing output, producers in both cases take the price as given. In the first case, which is free of externality, the marginal labour productivity is $2 \sum_{j \in \bar{m}} \bar{M}_j$ and, therefore, the marginal cost is $\omega / 2 \sum_{j \in \bar{m}} \bar{M}_j$ where ω is the *wage rate.*[3] Let the demand price be given as $P \left[\sum_{j \in \bar{m}} \bar{M}_j \right]$. Equating price to marginal cost we have

$$P \left[\sum_{j \in \bar{m}} \bar{M}_j \right] = \omega / 2 \sum_{j \in \bar{m}} \bar{M}_j . \tag{12.3}$$

In the second case, which involves an externality, the perceived marginal productivity is $\sum_{j \in \bar{m}} \bar{M}_j$ and the perceived marginal cost is $\omega / \sum_{j \in \bar{m}} \bar{M}_j$. Therefore,

the rôle market structure plays in determining whether or not an externality does exist. While the proponents of the first approach maintain that information about the market structure is essential, the proponents of the second believe that knowledge of the technology involved (utilities, production functions) is sufficient to decide about the existence of an externality.

[3] We impose the constraint $\omega > 25$ which ensures that the slope of the demand price is steeper than the slope of the marginal cost, so that equating marginal cost to price is a stable equilibrium.

equating once again price to marginal cost yields

$$P\left[\sum_{j\in\bar{m}} \bar{M}_j\right] = \omega / \sum_{j\in\bar{m}} \bar{M}_j. \qquad (12.4)$$

Since the demand price declines with increasing output and since output increases with total employment, it must be true that total employment in the first case (without externality) is larger than total employment in the second case (with externality). Employment here signifies agglomeration. We conclude that agglomeration in both cases examined is generated by *scale economies.* But in the second case, where there are positive *external economies*, the agglomeration is smaller than in the first case where there are no external economies. Thus we also conclude that external economies *deter agglomeration*, which agrees with the terminology of Mills and Hamilton (1984) and of Helpman (1998).[4] In the case of external diseconomies, the same approach leads to a conclusion consistent with Tolley and Grihfield (1987) who maintain that congestion *encourages agglomeration.*[5]

12.3 The Population Distribution Among Cities

12.3.1 External Scale Diseconomies and Economies in an Old Debate

We define the externalities associated with *city size*, that is, *external scale diseconomies and economies*, as the gap between the *net* MSC of accommodating the population in the city at the equilibrium utility level and the *net* MPC of a bundle which allows the individual to achieve the same utility.[6] The *net* MSC is the amount of resources the city must forego in order to accommodate the additional individual, that is, the difference between the MSC and the *marginal social product* MSP. It is equal to the negative value of the marginal surplus of the city (see Hochman (1978, 1981)). The *net* MPC is equal to the difference between the *market value* MV of the consumption bundle required for achieving the equilibrium utility level and the wage ω earned by the individual in the city.

[4]Suppose that $P[\cdot] = 5 - 20\sum_{j\in\bar{m}} \bar{M}_j$ and $\omega = .25$. Then the stable equilibrium output under one producer (which is also an optimum) equals .222 while, under unco-ordinated producers, it equals .181. The difference between the two, .041, specifies the degree to which the positive externality deters employment.

[5]Krugman's (1995) statement at the beginning of this section seems to be consistent with the alternative definition of an externality in footnote 2. Indeed if external economies are identified with increasing returns to scale then they tend to *promote concentration.* Similarly, external diseconomies tend to *promote dispersal.*

[6]In these definitions we represent the concepts formulated, implicitly or explicitly, in the debate between Mills and Hamilton (1984) and Tolley and Grihfield (1987) about the effect of positive and negative externalities on agglomeration

Thus the gap is given by

$$\underset{\text{social cost}}{\overset{\text{net marginal}}{(\mathrm{MSC}-\mathrm{MSP})}} - \underset{\text{private cost}}{\overset{\text{net marginal}}{(\mathrm{MV}-\omega)}} = (\mathrm{MSC}-\mathrm{MV}) - (\mathrm{MSP}-\omega)\,. \tag{12.5}$$

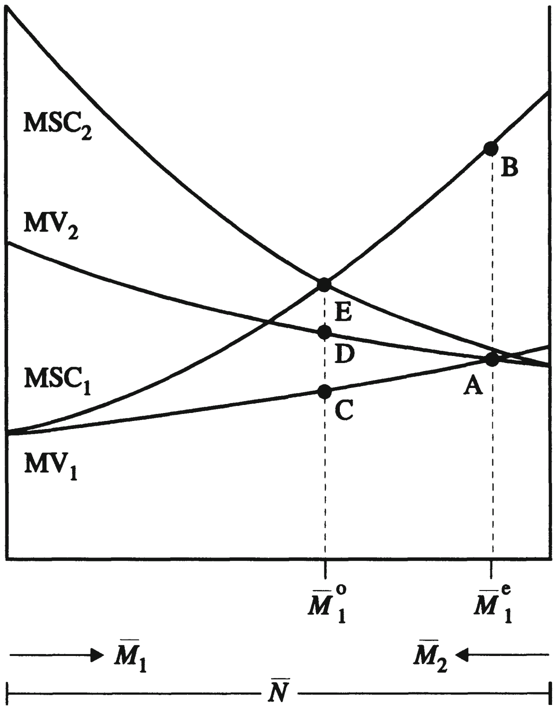

FIGURE 12.1. Partitioning of Population Between Two Cities.

Based on the above definitions, figure 12.1 illustrates the distribution of population in a system of two cities and the associated gaps. The total population $\bar{N}$ in that figure is partitioned between the two cities, where $\bar{M}_i$ $(i = 1, 2)$ denotes the city populations so that $\bar{M}_1 + \bar{M}_2 = \bar{N}$. The population of city one is measured from left to right, and the population of city two from right to left. We assume constant returns to scale in production and a competitive wage, so that the wage earned by an individual equals the MSP. In this case the gap between net MSC and net MPC, which represents an externality, simply equals MSC − MV. As shown in figure 12.1, the gap increases with population in both cities. Since the MSC exceeds the MV, there are *external scale diseconomies* by definition.

Figure 12.1 can be used to describe the basic arguments in the debate about the effect of externalities on agglomeration, as represented by Tolley and Grihfield (1987) on the one hand and by Mills and Hamilton (1984) on the other. The figure represents the argument of Tolley and Grihfield. The equilibrium partition of the population between the two cities is where the two MVs of the bundle that yields the equilibrium utility are equalised. This happens at A, where $\bar{M}_1^e$ individuals live in city one and the rest in city two. However this allocation is inefficient because, at A, $MSC_1 > MSC_2$. Thus, by moving an individual from the large city one to the small city two, we can save resources. The efficient allocation is at E, where $MSC_1 = MSC_2$ and where $\bar{M}_1^\circ$ individuals live in city one and the rest in city two. In this manner Tolley and Grihfield conclude that market forces in the presence of external scale diseconomies induce excessive concentration of people in the big city. The criticism of Mills and Hamilton on this argument is that there are also external economies associated with agglomeration which generate gaps in favour of the net market benefit. In essence, whereas Tolley and Grihfield focus on the difference between MSC and MV of the gap, Mills and Hamilton focus on the difference between MSP and ω. The debate is about which one of the two components that determine the gap, (MSC − MV) or (MSP − ω), is likely to dominate. As we have seen, if the former dominates, external scale diseconomies induce excessive concentration of people in the big city. By contrast, if the latter dominates, external scale economies cause insufficient concentration of people in the big city.[7] Since no one knows the relative size of the two components of the gap, no conclusion can be made based on this analysis according to Mills and Hamilton. But as we have already stated in the introduction, even more fundamental issues must be settled before the analysis based on figure 12.1 can become meaningful.[8] These can be summarised as follows.

1. Why don't we have *two* efficient cities of type one or, even better, why don't we have an *infinite number* of type–one cities each of infinitely small size? The answer must be that either the cities are not replicable or that they are indivisible.[9] Thus the difficulty is not just external scale effects, but such externalities in conjunction with nonconvexity.

2. If the cities differ from one–another due to nonconvexity, why should we expect that the gap will depend only on population size? And if it depends on other factors, why is the gap of the smaller city smaller than the gap of the larger city?

[7] We can use a modification of figure 12.1 in order to illustrate the effect of external scale economies. Under these circumstances the net MV exceeds the net MSC. Now replace MSCs in figure 12.1 with net MVs, and MVs with net MSCs. Then the equilibrium partition is at E where the two net MVs are equalised, while the efficient partition is at A where the two net MSCs are equalised.

[8] In discussing these issues we refer to external scale diseconomies although, as we shall illustrate, they apply equally well to external scale economies.

[9] As we have shown in chapter eleven, we can have cities of different size—even under perfect replicability. But if we introduce externalities in chapter eleven, the issue of efficient city size changes and the question is now whether or not a city of given type is optimal. In this context we can obtain a result that every city, small and large, can be either too small or too large. This complicated issue is beyond the scope of our book.

3. How does the gap between MSC and MV reflect the externalities that distort resource allocation inside the city? In particular, does unpriced transportation congestion generate external *scale* diseconomies?

4. If the answer to the previous question is on the affirmative and, indeed, the gap depends only on population size, does the gap increase as shown in figure 12.1? This issue is important because, otherwise, the assertion that the gap of the larger city is larger than that of the smaller city is baseless.

The first two issues are destructive because they can render the debate, as represented in figure 12.1, meaningless. We avoid both of them by assuming nonconvexity and that the gap depends only on city size. In the following two sections we discuss the last two issues as they relate to the gaps MSC $-$ MV, and MSP $- \omega$.

12.3.2 The Gaps MSC – MV and MSP – ω

The arguments of Tolley and Grihfield (1987) in the previous section were based on the premise that, under constant returns to scale in production, external diseconomies associated with pollution and unpriced transportation congestion imply MSC > MV for any city size. Although this seems plausible, it may not be so in the context of an urban model where the mere complexity of interdependencies among the different elements that constitute a city prevents straightforward intuition. Thus, strictly speaking, the claim 'MSC > MV under external diseconomies' needs a proof in the context of an urban model. To our knowledge, such proof does not exist in the literature. In the appendix we provide a partial justification to that claim, which also indicates to the reader that these matters are not as simple and straightforward as tacitly believed. In particular, we show that the internal unpriced transportation congestion generates a positive gap between the marginal cost of accommodating an additional individual in the city and the market value of the consumption bundle which allows for the predetermined utility level.[10]

The justification is based on result 7.3 in chapter seven. On the one hand, MSC is the shadow value of the consumption bundle plus the marginal social effect of the marginal individual on transport cost:

$$\text{MSC} = Z^e + \lambda_2^\circ H^e + \int_0^x \left(Q^e \frac{\partial \psi}{\partial Q} \langle Q^e, L_t^\circ \rangle - \Lambda_1^\circ \frac{\partial \psi}{\partial Q} \langle Q^e, L_t^\circ \rangle \right) dx. \qquad (12.6)$$

The first two terms on the RHS of (12.6) represent the social cost of the consumption bundle at any given location. The last term is the social transport cost. The first component in parenthesis is the marginal cost of a commuter, while the second is the social 'credit' for the cost which substitutes the missing

[10] Since our result applies to the *gross* values, it is meaningful only under constant returns to scale in production.

congestion toll (see section 7.3.2). On the other hand, MV corresponds to the same bundle evaluated by market prices plus the average transport cost borne by the individual:

$$\text{MV} = Z^e + R^e H^e + \int_0^x \Psi dx. \tag{12.7}$$

Hence the gap is given by

$$\text{MSC} - \text{MV} = (\lambda_2^\circ - R^e)\, H^e + \int_0^x \left((Q^e - \Lambda_1^\circ) \frac{\partial \psi}{\partial Q} \langle Q^e, L_t^\circ \rangle - \Psi \right) dx. \tag{12.8}$$

Since the gap is the same for all locations, we evaluate it at the centre where there are no transportation costs and, therefore, where we must examine only the difference between the shadow value of housing and its market value. We know from result 7.3 that, at the centre, the shadow price of land exceeds the market price, which implies MSC > MV.[11]

The gap between MSP and ω can be derived directly from the specification of the production function and the perceptions of the producers (see Dixit (1973), Henderson (1986) and Helpman (1998)). Given that the production function of firm i in city j is $\bar{M}_j f\left[\bar{N}\right]$ where $df/d\bar{N} > 0$ and $\bar{N} = \sum_{j \in \bar{m}} \bar{M}_j$, the wage ω is equal to the perceived marginal product $f\left[\bar{N}\right]$.[12] Therefore

$$\begin{aligned} \text{MSP} - \omega &= \frac{\partial}{\partial \bar{M}_j} \left(\sum\nolimits_{j \in \bar{m}} \bar{M}_j f\left[\bar{N}\right] \right) - f\left[\bar{N}\right] \\ &= \bar{N} \frac{df}{d\bar{N}}. \end{aligned} \tag{12.9}$$

Although it was not explicitly specified, we believe that this is the gap referred to by Mills and Hamilton (1984).

12.3.3 Variation of Gaps with City Size

So far, we have explained the contribution of unpriced transport congestion to external scale diseconomies reflected in the gap MSC –MV, and the contribution of external economies in production to the external scale economies reflected in the gap MSP $-\ \omega$. However, what determines how these scale externalities affect agglomeration is whether the total gap (MSC − MV) − (MSP − ω) is increasing or decreasing. In the first case the scale externalities enhance agglomeration as claimed by Tolley and Grihfield (1987). In the second case they deter agglomeration, a possibility raised by Mills and Hamilton (1984). But this observation implies that, in contrast to Mills and Hamilton, knowing the size of the gaps is not sufficient to infer what is the combined effect of the scale externalities on

[11] Notice that, for our calculations, we use the primitive elaborated model with unpriced transportation congestion to derive the reduced values of MSC − MV.

[12] This is a more general alternative with respect to the specification of scale economies than that used by Helpman (1998).

agglomeration. To illustrate, suppose that MSC – MV = 0, so that what matters is only MSP – ω. This satisfies the requirement of Mills and Hamilton for insufficient agglomeration. But does this actually mean that the big city is not big enough? The answer is negative. For if this gap depends only on population size and it decreases as population increases, the large city is too large—not too small!

Of course, the variation of the gap with city size can be partitioned into the variation of MSC – MV and the variation of MSP – ω. If, on the one hand, the first gap is increasing while the second is non–increasing, or if the first gap is non–decreasing while the second is decreasing, we get the result of Tolley and Grihfield. If, on the other hand, the first gap is non–increasing while the second is increasing, or if the first gap is decreasing while the second is non–decreasing, the possibility pointed out by Mills and Hamilton is obtained. If however both gaps are either increasing or decreasing then we have to compare the rates of variation with city size. Our main conclusion is that the problem is more complicated than Mills and Hamilton believed. In what follows we show that there is no a priori information about how these gaps change with city size. Beginning with MSC – MV, we show in the appendix that in the special case of Leontief preferences, the gap MSC – MV indeed increases as expected by Tolley and Grihfield, and it is far from negligible. However, we lack a general proof that this must be the case, and we conjecture that it need not be true in general. Indeed, the behaviour of the gap with city size crucially depends on the average transportation cost function and whether, as the traffic volume increases, the ratio (marginal minus average)/average cost increases or decreases. This relation can also depend on other urban characteristics like, for example, the land–supply function.

Perhaps stronger is our reservation about the implicit assumption of Mills and Hamilton that the gap MSP – ω is increasing with city size. As we have seen, this gap is actually increasing in the example of section 12.2. In this model, $f\left[\bar{N}\right] = \sum_{j\in\bar{m}} \bar{M}_j$ and the aggregate production in a given city is $\bar{N}\sum_{j\in\bar{m}} \bar{M}_j$. If the firms take the aggregate employment outside of their control then the gap $\bar{N}df/d\bar{N}$ is given by (12.2) as $\sum_{j\in\bar{m}} \bar{M}_j \equiv \bar{N}$. We conclude that, in this specific case, the gap MSP $-\,\omega$ increases with city size. However, in the more general case where the production function of firm i is $\bar{M}_i f\left[\bar{N}\right]$, this conclusion may not apply. The aggregate output is $\sum_j \bar{M}_j f\left[\bar{N}\right]$ and, as shown in the preceding section, the gap between the marginal social output and the perceived marginal output is $\bar{N}df/d\bar{N}$. Thus the variation of the gap with city size is given by $df/d\bar{N} + \bar{N}\ d^2f/d\bar{N}^2$. The sign of the first expression is, indeed, positive; but the sign of the second is ambiguous and, therefore, can even dominate the first expression to produce a gap *decreasing* with city size. For example, $f\left[\bar{N}\right] = \log\bar{N}$ implies $df/d\bar{N} + \bar{N}\ d^2f/d\bar{N}^2 = 0$, so that the gap remains constant with city size. Another admissible example is given by $f\left[\bar{N}\right] = a(1-\exp(-\bar{N}))$ which implies a gap equal to $a\bar{N}\exp\left(-\bar{N}\right)$. The variation of this gap with distance is given by $a\left(1-\bar{N}\right)\exp\left(-\bar{N}\right)$ which is negative for all $\bar{N}>1$.[13]

12.4 The Rôle of Nonconvexity

In this section we discuss the last main issue presented at the beginning of this chapter. Namely, what is the rôle of nonconvexity in distorting the population distribution among cities. We shall argue that, under nonconvexity, *external scale economies* may be consistent with *excessive concentration* of people in big cities—rather than with *insufficient concentration* as maintained by Mills and Hamilton (1984) and Helpman (1998).

12.4.1 The Utility–Size Configuration Revisited

A key concept in the analysis of this section is the *utility–size configuration* $v\left[\bar{M}_i,\bar{N}\mid\mathcal{E}\right]$, which describes the dependence of the utility achieved in city i on its *population* $\bar{M}_i$ and on the *total urban population* $\bar{N}$ for a given *economy* $\mathcal{E}$. It is convenient for our purposes to distinguish in the description of an economy between *fundamentals* $\mathcal{F}$ and *market structure* $\mathcal{M}$, so that $\mathcal{E}\equiv(\mathcal{F},\mathcal{M})$. The fundamentals specify utilities, production functions and initial aggregate endowments in the economy. The market structure specifies the decision–making process which includes a list of agents, their objective functions, the instruments under their control and the distribution of the initial endowments among them.[14]

The reader should keep in mind that, here, the shape of the utility–size configuration depends not only on city size but on the total urban population as well. This differs from the utility–size configuration in chapters ten and eleven, where it was defined for autarchic cities. Furthermore, the structure of $v\left[\bar{M}_i,\bar{N}\mid\mathcal{F},\mathcal{M}\right]$ implies that the utility–size configuration may depend not only on the resource allocation in city i, but also on the resource allocation in the other cities. Take for example land ownership in a system of cities inhabited by identical individuals. If land is equally owned by everyone, the aggregate land rent is equally distributed to individuals irrespectively of their residential location. This, in turn, implies inter–city resource transfers which depend on the resource allocation in all cities. Another example is provided by the new economic geography in the tradition of Krugman (1991). There, utility depends

[13] More generally, in standard textbooks, the average output function is drawn as a ∩–shaped curve. This implies that, beyond some output, the gap between marginal and average output declines. Hence an assumption that the gap is increasing cannot be taken as self–evident.

[14] Within the list of agents we may include local governments which supply LPGs and finance them by local taxes. We may also consider a *co–ordinated* economy where a central government affects the inter–city resource transfers, other than the transfers implied by the distribution of initial endowments and the decision making of the remaining agents.

on the real wage which, in turn, depends on local wage and the price index. Since consumption includes imports from other cities, it follows that their prices affect local utility. As a result of this dependence the utility–size configuration need not assume the standard ∩–shape. Moreover, the shape itself changes with the total population size.

For the rest of this section we consider a two–city system. We use specific examples of this system in order to illustrate how and to what extent an economy affects the utility–size configuration. We base our discussion on a prototype model of Helpman (1998), which retains the main implications of his full model. We present this model immediately below, and we modify it subsequently in different ways to produce other examples.

First Example (Helpman)

The *fundamentals* in this example include (1) a Cobb–Douglas utility function $Z_i^\beta H_i^{1-\beta}$ for city $i = 1, 2$; (2) a production function $\bar{M}_{ij}^\delta \sum_{k\in\bar{m}_i} \bar{M}_{ik}$ of the composite good produced in city i by firm j, where $\bar{M}_{ij}$ is the *number of employees* in firm j and $\bar{m}_i$ is the *number of firms* in city i; (3) an urban housing stock of total size $\bar{H}_i$ for city i, the supply of which is perfectly inelastic; and (4) an initial endowment of one labour unit for every individual so that the *aggregate labour supply* in city i is $\bar{M}_i$. The *market structure* includes the following information. (5) Each employer perceives $\bar{M}_i$ as a given parameter, so that the firm's output is proportional to the number of workers it employs. (6) The markets are perfectly competitive: each factor receives its (perceived) value of marginal product and the markets are cleared, which implies

$$\begin{aligned}
\sum_i \bar{M}_i Z_i &= \left(\sum_i \bar{M}_i\right)^{1+\delta} && (a) \\
\bar{M}_i H_i &= \bar{H}_i && (b) \\
\bar{M}_i &= \textstyle\sum_{k\in\bar{m}_i} \bar{M}_{ik}\,. && (c)
\end{aligned} \tag{12.10}$$

(7) The land in each city is owned by the entire population in equal share, so that each individual is entitled $1/\bar{N}$ of the aggregate rent in the urban system.

Using the above specification of fundamentals and market structure we can derive the utility–size configuration as follows. With an equilibrium wage rate of $\bar{M}_i^\delta$, the aggregate wage bill in the system is $\bar{M}_i^{1+\delta} + \left(\bar{N} - \bar{M}_i\right)^{1+\delta}$. Furthermore, with a Cobb–Douglas utility function the aggregate land rent is equal to $(1-\beta)\left(\bar{M}_i^{1+\delta} + \left(\bar{N} - \bar{M}_i\right)^{1+\delta}\right)/\beta$. It follows that per capita income in city i, which includes the wage and the share of aggregate land rent, is given by

$$\bar{Y}_i = \bar{M}_i^\delta + (1-\beta)\left(\bar{M}_i^{1+\delta} + \left(\bar{N} - \bar{M}_i\right)^{1+\delta}\right) / \left(\beta\bar{N}\right). \tag{12.11}$$

Accordingly, with a Cobb–Douglas utility function, the equilibrium utility–size configuration can be written as

$$v\left[\bar{M}_i, \bar{N} \mid \mathcal{F}, \mathcal{M}\right] = \left(\beta \bar{Y}_i\right)^{\beta} \left(\frac{\bar{H}}{\bar{M}_i}\right)^{1-\beta}. \tag{12.12}$$

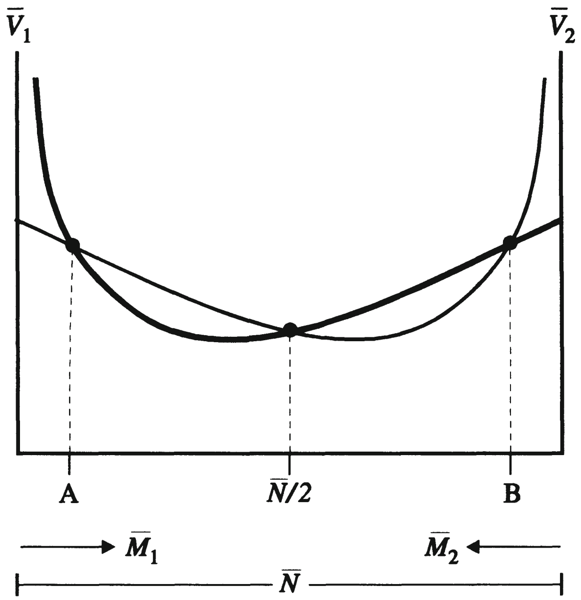

FIGURE 12.2. A Utility–Size Configuration for the First Example.

Figure 12.2 shows the graph of the utility–size configuration (12.12) for $\beta = .75, \delta = .5, \bar{H}_i = 1$ and $\bar{N} = 50$. The thick line denotes the achievable utility as a function of the population in city one (measured by the distance from the left axis). The thin line corresponds to city two. In this example scale diseconomies dominate scale economies for small population size, and the opposite happens for large population size.

Second Example

In this example we simply replace the Cobb–Douglas utility function of the first example with a Stone–Geary utility function $(Z_i - A)^{\beta}(H_i - B)^{1-\beta}$. With this simple change in the fundamentals, we obtain for $i, j = 1, 2$

$$\bar{Y}_i = \frac{b_i a_{jj} - b_j a_{ij}}{a_{ii} a_{jj} - a_{ij} a_{ji}} \tag{12.13}$$

where

$$\begin{aligned} a_{ii} &= 1 - \frac{(1-\beta)\bar{M}_i}{\bar{N}\left(1-A\beta\bar{M}_i\right)}; \quad a_{ij} = -\frac{(1-\beta)\bar{M}_j}{\bar{N}\left(1-A\beta\bar{M}_j\right)} \\ a_{ji} &= -\frac{(1-\beta)\bar{M}_i}{\bar{N}\left(1-A\beta\bar{M}_i\right)}; \quad a_{jj} = 1 - \frac{(1-\beta)\bar{M}_j}{\bar{N}\left(1-A\beta\bar{M}_j\right)} \\ b_i &= \bar{M}_i - (1-\beta)B\left(\frac{\bar{M}_i}{(1-A\beta\bar{M}_i)} + \frac{\bar{M}_j}{(1-A\beta\bar{M}_j)}\right)/\bar{N} \\ b_j &= \bar{M}_j - (1-\beta)B\left(\frac{\bar{M}_i}{(1-A\beta\bar{M}_i)} + \frac{\bar{M}_j}{(1-A\beta\bar{M}_j)}\right)/\bar{N}. \end{aligned} \tag{12.14}$$

The equilibrium utility–size configuration becomes

$$v\left[\bar{M}_i, \bar{N} \mid \mathcal{F}, \mathcal{M}\right] = \left(\frac{\beta\left(\bar{Y}_i - B\right)\left(1 - A\bar{M}_i\right)}{1 - A\beta\bar{M}_i}\right)^{\beta}\left(\frac{1}{\bar{M}_i} - A\right)^{1-\beta}. \tag{12.15}$$

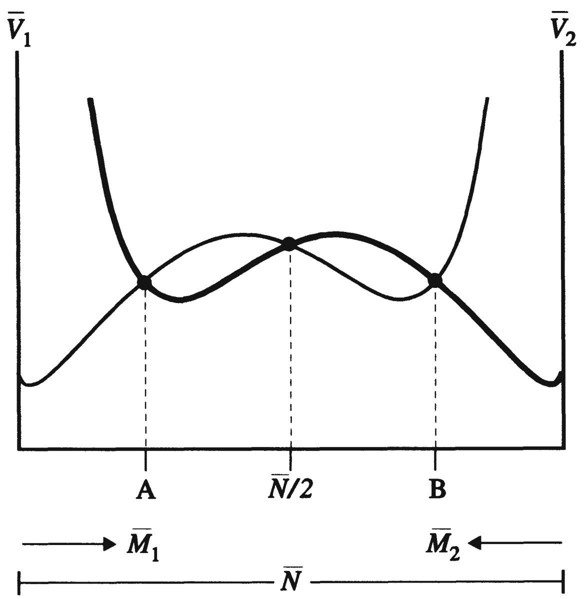

FIGURE 12.3. A Utility–Size Configuration for the Second Example.

Figure 12.3 shows the graph of the utility–size configuration (12.12) for $A = B = .1$, $\beta = .75$, $\bar{H}_i = 1$ and $\bar{N} = 7.65$. By introducing a minimum consumption level for each good, we forced the utility level in this example to decline with population size beyond some level.

12.4.2 Equilibrium, Stability and Market Bias

Before we examine the implications of the above two examples in more detail, it is useful to define the concept of *migration equilibrium*, *stability* of the equilibrium and the *characteristics of market failure* associated with the alternative specifications of $v\left[\bar{M}_i, \bar{N} \mid \mathcal{F}, \mathcal{M}\right]$. We assume that migration is costless, so that the equilibrium population partition between the two cities must equalise utility levels—provided that both cities are populated. Thus $(\bar{M}_1 = 0, \bar{M}_2 = \bar{N}), (\bar{M}_1 = \text{A}, \bar{M}_2 = \bar{N} - \text{A}), (\bar{M}_1 = \bar{M}_2 = \bar{N}/2), (\bar{M}_1 = \text{B}, \bar{M}_2 = \bar{N} - \text{B})$ and $(\bar{M}_1 = \bar{N}, \bar{M}_2 = 0)$ are the equilibria of both examples in figures 12.2 and 12.3. Of these, $(\bar{M}_1 = 0, \bar{M}_2 = \bar{N})$, $(\bar{M}_1 = \bar{N}, \bar{M}_2 = 0)$ and $(\bar{M}_1 = \bar{M}_2 = \bar{N}/2)$ in both examples are *unstable* equilibria because a migration from city one to city two implies higher utility for city two, and vice–versa. The remaining equilibria $(\bar{M}_1 = \text{A}, \bar{M}_2 = \bar{N} - \text{A})$ and $(\bar{M}_1 = \text{B}, \bar{M}_2 = \bar{N} - \text{B})$ in both examples are *stable.*

We say that 'the market fails' if any stable equilibrium is Pareto–dominated by any other feasible allocation which can be achieved by redistributing income.[15] We also say that 'a market allocation is biased toward concentration' if any stable equilibrium partition of the population between the two cities is Pareto–dominated by another equilibrium partition in which the big city is smaller; and that 'a market allocation is biased toward dispersion' if any stable equilibrium partition of the population between the two cities is Pareto–dominated by another equilibrium partition in which the big city is bigger. In the first example, the stable equilibria $(\bar{M}_1 = \text{A}, \bar{M}_2 = \bar{N} - \text{A})$ and $(\bar{M}_1 = \text{B}, \bar{M}_2 = \bar{N} - \text{B})$, which represent partial agglomeration, are Pareto–dominated by $(\bar{M}_1 = 0, \bar{M}_2 = \bar{N})$ and $(\bar{M}_1 = \bar{N}, \bar{M}_2 = 0)$ which represent full agglomeration. Consequently, the market allocations in the first example are biased toward dispersion. By contrast, in the second example, the stable equilibria $(\bar{M}_1 = \text{A}, \bar{M}_2 = \bar{N} - \text{A})$ and $(\bar{M}_1 = \text{B}, \bar{M}_2 = \bar{N} - \text{B})$, which represent partial agglomeration as before, are Pareto–dominated by $(\bar{M}_1 = \bar{M}_2 = \bar{N}/2)$ which represents full dispersion. Thus the market allocations in the second example are biased toward concentration.

12.4.3 Externalities, Nonconvexity and the Direction of Market Bias

In both the first and second examples of section 12.4.1, the production functions of the composite good exhibit *external scale economies* where the external effect, as represented by the gap MSC − MV, increases with city size. In both examples there are no external scale diseconomies because housing is a standard private good. Under these circumstances Mills and Hamilton (1984) would expect that the market is biased toward dispersion. This is precisely the

[15] Notice that, in this comparison, we disregard other allocations that can be achieved by appropriate inter city transfers. Thus our definition of failure refers to sufficient, rather than necessary, conditions.

result obtained in the first example, the prototype of Helpman's (1998) model. However, in the second example, which introduces another source of scale diseconomies for housing but not *external* scale diseconomies, the market is biased toward agglomeration. This allows us to infer

> **Result 12.1**: External scale economies, in conjunction with nonconvexity, may lead to excessive rather than insufficient agglomeration. This can be true even if the external effect, as reflected by MSP $-\omega$, increases with city size.

Third Example

If in the second example we change the aggregate production function from $\bar{M}_i^{1+\delta}$ to $\bar{M}_i \log \bar{M}_i$, we obtain a model in which the external scale economies are constant with city size because $\partial\left(\bar{M}_i \log \bar{M}_i\right)/\partial \bar{M}_i - \log \bar{M}_i = 1$. We have already noted that, in this case, the external economies have no effect. More precisely, a perfect internalisation scheme will not change the equilibrium because the lump–sum Pigouvian subsidy will be perfectly offset by a lump–sum tax necessary to finance it. Hence this case is equivalent to one without external economies.

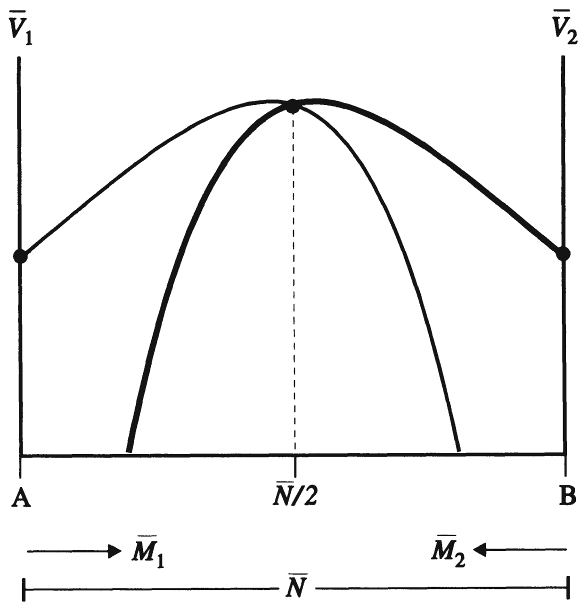

FIGURE 12.4. A Utility–Size Configuration for the Third Example.

If we let $\beta = .9$, $\bar{N} = 10$ and keep all the other parameters associated with figure 12.3 unchanged, we obtain figure 12.4. In this figure the two stable equilibria $(\bar{M}_1 = 0, \bar{M}_2 = \bar{N})$ and $(\bar{M}_1 = \bar{N}, \bar{M}_2 = 0)$ are Pareto–dominated by the allocation ($\bar{M}_1 = \bar{M}_2 = \bar{N}/2$), which implies that the market is biased toward agglomeration.[16] We conclude that, in this case, a market which is equivalent to one without any external economies fails to distribute the population optimally between the two cities. In this example market failure is caused by nonconvexity alone.

> **Result 12.2**: Nonconvexity in itself is a source of market failure, and the market bias depends not only on whether the economy is characterised by scale economies or diseconomies, but also upon the combined effects of the fundamentals and the market structure on the utility–size configuration.

One might object that the evidence we presented above in order to support result 12.2 is incomplete, since all our examples are based on variations in the fundamentals but we have no variation in market structure. Thus it is not clear that market structure can affect the direction of market bias as claimed in result 12.2. However, it is easy to construct examples in which changes in market structure affect the utility–size configuration. To see this, in the context of Helpman's prototype model, assume that city governments tax away all land rents in their jurisdiction and redistribute them equally among the local population. In this case market structure changes because the two cities become autarchic; and there is no change in the fundamentals. If we take into account that, under these conditions, individuals must consume an amount of composite good equal to their equilibrium wage rate $\bar{M}_i^{\delta}$, we can write the equilibrium utility–size configuration directly as

$$\begin{aligned} v\left[\bar{M}_i, \bar{N} \mid \mathcal{F}, \mathcal{M}\right] &= \left(\bar{M}_i^{\delta}\right)^{\beta}\left(\frac{\bar{H}}{\bar{M}_i}\right)^{1-\beta} \\ &= \bar{H}^{1-\beta}\bar{M}_i^{\beta(1-\delta)-1}, \end{aligned} \tag{12.16}$$

which does not depend on total population since the cities are autarchic. Hence $dv/d\bar{M}_i > (=,<)\ 0 \Leftrightarrow \beta(1-\delta) > (=,<)\ 1$.[17] In this example, unlike in all others, the equilibrium utility–size configuration behaves monotonically with increasing population size.

We close this section with some recent findings in more sophisticated models that combine Krugman's (1991) differentiated product approach with the monocentric model, or with a simplified version of it. Hadar (1997) has modified Helpman's (1998) full model by considering a CES utility function, rather than

---

[16] If in the same model we suppress Stone–Geary's minimum consumption levels by setting $A = 0$, we can also produce an example where the market is biased toward dispersion. With the remaining parameters equal to $\beta = .7, \bar{H}_i = 1$ and $\bar{N} = 12$, the population partition ($\bar{M}_1 = \bar{M}_2 = \bar{N}/2$) becomes a stable equilibrium which is dominated by full agglomeration ($\bar{M}_1 = 0, \bar{M}_2 = \bar{N}$) and ($\bar{M}_1 = \bar{N}, \bar{M}_2 = 0$).

[17] Notice that this specification is at odds with club theory except when $\beta(1-\delta) = 1$.

the standard Cobb–Douglas utility function commonly used by those who adopt the specification of Dixit and Stiglitz (1977). With this modification, Hadar obtains figure 12.3, rather than 12.2 as in the original model of Helpman.[18] Precisely the same configuration is also obtained by Tabuchi (forthcoming) in whose model scale diseconomies stem from urban crowding, rather than from the need to provide immobile farmers with manufacturing goods as in Krugman. In a sense, the housing component of Helpman's model is a reduced form of Tabuchi, so that Tabuchi and Hadar are closely related. Since in both models there are external scale economies but no external scale diseconomies, these models show once again that external scale economies are consistent with excessive agglomeration not only in the case of the simple examples we presented in this chapter, but also in the full context of the new economic geography.

12.5 Concluding Comments

This chapter demonstrates how vulnerable can be any conclusions in the literature about the way externalities affect market bias. It becomes clear that the direction of bias crucially depends on the specification of the fundamentals and the market structure. Since the results are not robust, one should be very careful in drawing conclusions and should specify with excessive care the underlying premises. This is especially relevant in modelling frameworks such as the new economic geography, where specific functional forms are extensively used. As we have pointed out, choosing a specific utility function is almost assuming the outcome of the analysis.

Our discussion has shed new light on an old debate. If market failure in the context of inter–city population distribution was just caused by external economies and diseconomies, the right policy would be to internalise these externalities by appropriate Pigouvian taxes or subsidies. For why is it then necessary to prevent the excessive growth of Paris due to congestion and pollution, as implied by Tolley and Grihfield (1987)? If such externalities were the only cause of market failure, wouldn't it be much better to charge congestion tolls and let the market do the rest, including the right distribution of population among cities? The message of this chapter is that such standard policy prescriptions may well be insufficient. In the presence of nonconvexity, more comprehensive urban policies must be designed to encourage the efficient distribution of population among cities.

[18] Figure 12.2 applies both to the prototype model of Helpman, which we have presented in our first example, and to his full model. In Hadar's model, figure 12.3 can be obtained for parameter values $\bar{H} = .4, \bar{\gamma} = 1.61, \bar{N} = 2.87$ and for an elasticity of substitution $1/(1 - \bar{a}) = .28$, where we have used the same notation as in section 9.2.4 of chapter nine.

12.6 Appendix: Gap Calculations

In this appendix we derive the gap between MSC and MV for a standard urban model with unpriced transportation congestion which was elaborated in chapter seven. We use the same notation.

Recall that in chapter seven we maximised $\bar{U}$ subject to (7.10(a) – (f)) and (7.13). Consider here the dual problem of minimising

$$\int_{\mathcal{X}} \left(nz\left[R,\bar{U}\right] + Q\psi\left[Q,L_t\right]\right)dx + \Theta\left\langle\bar{x}\right\rangle\bar{R} \tag{12.17}$$

subject to (7.10(a) – (e)) and (7.13), where $\bar{U}$ is given. Denote the Lagrangean of the dual problem as $\mathcal{L}^{\mathrm{D}}\left[L_t,\mathbf{\Phi};\boldsymbol{\lambda}\right]$. By inspection, one can verify that this Lagrangean satisfies

$$\mathcal{L}^{\mathrm{D}}\left[L_t,\mathbf{\Phi};\boldsymbol{\lambda}\right] = \frac{\bar{U}}{\bar{\lambda}_7^{\mathrm{SB}}} - \frac{\mathcal{L}^{\mathrm{SB}}\left[L_t,\mathbf{\Phi};\boldsymbol{\lambda}\right]}{\bar{\lambda}_7^{\mathrm{SB}}}, \tag{12.18}$$

where $\bar{U}$ in (12.18) equals the equilibrium utility level attained by the second–best allocation in chapter seven and superscript 'SB' denotes that the designated expressions refer to the second–best problem of that chapter. It follows that all the first–order conditions (7.41) apply, and so does the analysis in section 7.4.5 of chapter seven.

We next evaluate the gap at the centre of the city where transport cost is zero. Using the envelope theorem and the first–order conditions (7.41) we obtain

$$\begin{aligned}
\mathrm{MSC} &= \frac{\partial\mathcal{L}^{\mathrm{D}}}{\partial\bar{N}} \\
&= z\left[R^e\langle 0\rangle,\bar{U}\right] + \lambda_2^{\circ}\langle 0\rangle\, h\left[R^e\langle 0\rangle,\bar{U}\right] \\
&= z\left[R^e\langle 0\rangle,\bar{U}\right] + R^e\langle 0\rangle\, h\left[R^e\langle 0\rangle,\bar{U}\right] \\
&+ \left(\lambda_2^{\circ}\langle 0\rangle - R^e\langle 0\rangle\right) h\left[R^e\langle 0\rangle,\bar{U}\right] \\
&= e\left[R^e\langle 0\rangle,\bar{U}\right] + \left(\lambda_2^{\circ}\langle 0\rangle - R^e\langle 0\rangle\right) h\left[R^e\langle 0\rangle,\bar{U}\right] \\
&= \mathrm{MV} + \left(\lambda_2^{\circ}\langle 0\rangle - R^e\langle 0\rangle\right) h\left[R^e\langle 0\rangle,\bar{U}\right] \\
&> \mathrm{MV},
\end{aligned} \tag{12.19}$$

where the inequality follows because of result 7.3.

We are unable to determine in general how the gap behaves with population size. We can however corroborate the conjecture that the gap increases with population size in the special case where the utility function exhibits zero elasticity of substitution between housing and the composite good, that is,

$u[Z,H] = \min\{Z,H\}$. Under these conditions, the *shadow* value of housing at the centre is equal to the *social* transport cost of an individual at the boundary,

$$\mathrm{MSC} = \int_0^{\bar{x}^e} \left(\psi [Q^e, L_t^\circ] + Q^e \frac{\partial \psi}{\partial Q} \langle Q^e, L_t^\circ \rangle \right) dx \tag{12.20}$$

(see Arnott and MacKinnon (1978)), while the *market* value of housing at the centre is equal to the *private* transport cost of an individual at the boundary:

$$\mathrm{MV} = \int_0^{\bar{x}^e} \psi [Q^e, L_t^\circ] \, dx. \tag{12.21}$$

Consequently, the gap between the shadow and the market value of land at the centre is

$$\mathrm{MSC} - \mathrm{MV} = \int_0^{\bar{x}^e} Q^e \frac{\partial \psi}{\partial Q} \langle Q^e, L_t^\circ \rangle \, dx. \tag{12.22}$$

The intuition behind (12.22) is straightforward. Given the utility level, the consumption bundle is completely determined. The only way an additional individual can be accommodated in the city (while enjoying the given utility level) is to be located at the boundary and be provided there by the same consumption bundle as everywhere else in the city. Using (12.22), since both Q^e and $\bar{x}^e$ increase with population size, the gap *in this case* also increases with population size as conjectured by Tolley and Grihfield (1987).

Further insight can be gained by using Vickrey's specification for the marginal resource cost of travel, that is,

$$\psi [Q, L_t] = \left(\frac{Q}{L_t} \right)^{\xi}. \tag{12.23}$$

In this case we have

$$Q^e \frac{\partial \psi}{\partial Q} \langle Q^e, L_t^\circ \rangle = \bar{\xi} \cdot \psi [Q^e, L_t^\circ]. \tag{12.24}$$

Hence

$$\begin{aligned} \frac{(\lambda_2^\circ \langle 0 \rangle - R^e \langle 0 \rangle) h \left[R^e \langle 0 \rangle, \bar{U} \right]}{R^e \langle 0 \rangle h \left[R^e \langle 0 \rangle, \bar{U} \right]} &= \left(\begin{array}{c} \int_0^{\bar{x}^e} (1+\xi) \psi [Q^e, L_t^\circ] \, dx \\ - \int_0^{\bar{x}^e} \psi [Q^e, L_t^\circ] \, dx \end{array} \right) \\ &\div \int_0^{\bar{x}^e} \psi [Q^e, L_t^\circ] \, dx \\ &= \bar{\xi}. \end{aligned} \tag{12.25}$$

With due reservations, this example allows us to evaluate the gap. With an empirically observed range of $\bar{\xi}$ between 2 and 4, 20–30 percent for the share of housing in consumption expenditure, 50 percent for the share of land in housing cost and ignoring the agricultural rent, we find that the gap MSC – MV ranges between 20 and 60 percent of the market value of the urban consumption bundle—which is substantial.

One can argue that, with the exception of the Leontief preferences, the gap is smaller than the external transportation effect.[19] Does this imply that our estimate is upwardly biased? The answer is negative. With H and Z sensitive to housing price, not only the *shadow value* of housing is smaller than the external transportation effect, but also the *market value* of housing is smaller than the differential private transportation cost. Consequently, the ratio between the gap and the housing market value at the centre may be even larger than in the case of Leontief utility. We therefore rely on the above illustration to state that the gap between MSC and MV is likely to be substantial.

12.7 References

A. Anas, R. J. Arnott and K. A. Small, forthcoming, Urban Spatial Structure. *Journal of Economic Literature.*

R. J. Arnott and J. M. MacKinnon, 1978, Market and Shadow Rent with Congestion. *American Economic Review* 68, 588–700.

K. J. Arrow, 1970, Political and Economic Evaluation of Social Effects and Externalities. In J. Margolis (ed.) *The Analysis of Public Output.* Columbia University Press, New York, 1–23.

A. K. Dixit, 1973, The Optimum Factory Town. *Bell Journal of Economics and Management Science* 4, 637–651.

A. K. Dixit and J. E. Stiglitz, 1977, Monopolistic Competition and Optimum Product Diversity. *American Economic Review* 67, 297–308.

M. Fujita and J.-F. Thisse, 1996, Economics of Agglomeration. *Journal of the Japanese and International Economics* 10, 339–378.

H. Gravelle and R. Rees, 1981, *Microeconomics.* Longman, London.

Y. Hadar, 1997, The Formation of Cities. Tel–Aviv University, (unpublished).

E. Helpman, 1998, The Size of Regions. In D. Pines, E. Sadka, and I. Zilcha (eds.) *Topics in Public Economics.* Cambridge University Press, New York, 33–54.

J. V. Henderson, 1986, Efficiency of Resources Usage and City Size. *Journal of Urban Economics* 19, 47–70.

[19] For positive elasticity of substitution between housing and the composite good, the gap is smaller than the external effect $Q^e \partial\psi/\partial Q^e - \Psi$ generated by a commuter who lives at the boundary. This can be verified by using (12.8) in conjunction with the analysis of section 7.3.2, and it is reflected in the numerical calculations of Arnott and MacKinnon (1978).

O. Hochman, 1978, A Two Sector Model with Several Externalities And Their Effect on Urban Setting. *Journal of Urban Economics* 5, 198–218.

O. Hochman, 1981, Land Rents, Optimal Taxation, and Local Public Independence in an Economy with Local Public Goods. *Journal of Public Economics* 15, 59–85.

P. Krugman, 1991, Increasing Returns and Economic Geography. *Journal of Political Economy* 99, 483–499.

P. Krugman, 1995, *Development, Geography and Economic Theory.* Cambridge University Press, Cambridge.

M. Lagueux, 1998, Learning from the Debate on Externalities. In R. E. Backhouse, D. M. Houseman, U. Maki and A. Salanti (eds.) *Economics and Methodology.* MacMillan, London, 120–147.

A. Lösch, 1940, *Die Raumliche Ordnung der Wirtschaft.* Translated in English as *The Economics of Location* by W. H. Woglom and W. F. Stopler. Yale University Press, New Haven, 1954.

E. S. Mills and B. W. Hamilton, 1984, Urban Economics. Harper Collins College Publishers, New York.

T. Tabuchi, forthcoming, Urban Agglomeration and Dispersion: A Synthesis of Alonso and Krugman. *Journal of Urban Economics.*

C. M. Tiebout, 1956, A Pure Theory of Local Public Goods. *Journal of Political Economy* 64, 416–424.

G. Tolley and J. Grihfield, 1987, City Size and Place as Policy Issue. In E.S. Mills (ed.) *Handbook of Regional and Urban Economics* II*: Urban Economics.* North–Holland, Amsterdam, 1285-1311.

H. R. Varian, 1987, *Micro Economic Analysis.* Norton, New York.

Author Index

Index

Q

R

S

T