Theories of urban externalities

Kanemoto, Yoshitsugu

University of Tokyo

1980

THEORIES OF URBAN EXTERNALITIES

Yoshitsugu Kanemoto
University of Tokyo

The Author's Note: This book was originally published as *Theories of Urban Externalities* by North-Holland in 1980. The book became out of print and the copyright was returned to me. The electronic version of the book is now offered, for free, to people who are interested in studying urban economics. I would like to thank Mrs. Akiko Nishiyama and Mrs. Miyabi Okamoto for the excellent and laborious job they did creating the electronic version from the original hard copy.

To my parents

PREFACE

In this monograph several aspects of externalities in cities are analyzed using extensions of a standard residential land use model. Topics covered are optimal and market city sizes, local public goods, traffic congestion, externalities between different types of households, and the growth of a system of cities.

The monograph grew out of the Ph.D. dissertation I submitted to Cornell University in 1977, which contained several original contributions to theoretical urban economics. I have made an effort to integrate recent theoretical development, and have added appendices on the envelope property and on optimal control theory to make the exposition self-contained.

Although the monograph is written primarily for researchers in the profession, it is designed to be accessible for graduate students who have finished a first year graduate microeconomics course. Mathematically oriented undergraduate students should be able to understand the materials after careful reading of the appendices on the envelope property and on optimal control theory.

I am grateful to the members of my dissertation committee, Walter Isard, who served as chairman. Henry Y. Wan Jr., and Richard E. Schuler, for their comments, criticisms, and suggestions. I wish to express my deepest gratitude to Robert M. Solow who acted as my adviser while the dissertation was being written and made possible an extremely fruitful year at the Department of Economics, Massachusetts Institute of Technology. My interest in this field was initially stimulated by his earlier works on theoretical urban economics. I also benefited greatly from his comments on an earlier draft.

I started doing research on some of the topics in this monograph when I was still a student at the University of Tokyo. I am grateful to Koichi Hamada, Takashi Negishi, Yasuhiko Oishi, Yukihide Okano, and Isao Orishimo for their guidance and suggestions.

I am indebted to Richard Arnott, who read an earlier version of the first five chapters and offered me many valuable comments, and to Masahisa Fujita, who gave me useful comments on the first two chapters.

I owe an enormous intellectual debt to many other people who have worked on urban economics, but I do not list them here. Acknowledgement of prior contributions are gathered in the *Notes* at the end of each chapter.

David Robinson provided editorial assistance. His contribution goes, however, beyond the usual editorial work. He made a great contribution to making the manuscript readable, and, being an aspiring economist himself, spotted many errors in earlier versions.

I would like to thank May McKee and Hilary Wilson for the excellent job they did typing the camera-ready copy of the manuscript. May typed Chapters I and II; Hilary, Chapters III and IV; Virginia Tabak, most of Appendices; and I typed the rest of the book.

INTRODUCTION

Cities are concentrations of people, and the essence of urban life is the presence, for better and for worse, of many other people. It could be argued that the essence of urban economics is therefore the analysis of externalities. Traffic congestion, discrimination, pollution, and public services all involve externalities, and all are important matters of public policy. To design better policies, the implications of externalities within a decentralized market system must be understood

The kind of interactions most often analyzed in economics are transactions of ordinary private goods which are bought and sold by individuals at a market price. This type of interaction always involves flows in two directions: a payment is made whenever a good is transferred. This book is concerned with interactions of a different kind - with externalities and public goods, in particular.

Externalities arise when an agent does not compensate others for the effect of his actions. Smokers who do not, for example, pay for cleaning windows, or for the damage they may do to others' health, or for the discomfort they may cause, produce a variety of externalities. Urban life, in fact, is filled with examples of externalities, some of which we consider in this book: firms often prefer to locate in larger cities because of the presence of other firms; individuals sometimes choose to avoid certain neighbourhoods because of the presence of certain ethnic groups; commuters find their travel costs increased because others choose to travel at the same time.

Public goods are goods that are consumed jointly by many individuals. A private good has the property that consumption is exclusive: if an individual eats an apple, nobody else can eat the apple. In the case of a public good, such as national defense, consumption of the good by one individual does not prevent others from consuming the good at the same time. As it turns out, it is difficult to achieve an efficient supply of a public good through the market, and most public goods are provided by the government.

There are different degrees of publicness in different public goods. At one extreme is the pure public good which is consumed by all individuals in the economy simultaneously and which it is impossible to prevent anyone consuming once it is supplied. The classic example is national defense. Most public goods are not pure in this sense, however. In this book we consider public goods which are jointly consumed but only by those who live closer to the place of supply. Parks, street lighting, or sidewalks are typical of such local public goods.

With the exception of Chapter VI, the book is concerned with the *normative* aspect of externalities and public goods, or with what *should* be done if there are externalities or public goods. There are two major issues in normative analysis: *efficiency* and *equity.* The aspect of efficiency is usually represented by the concept of Pareto optimality. *An allocation is called Pareto optimal if nobody can be made better off without making somebody else worse off.* Pareto optimality ignores distributional equity, however: the allocation with only one individual obtaining all the wealth and the

rest of the population starving to death may well be Pareto optimal. Although the problem of equity is extremely important, Chapter VI is the only chapter that deals with the problem of income distribution, and our analysis there is *descriptive* rather than normative.

For the sake of simplicity, we usually restrict our analysis to the case in which all households obtain the same utility level, and then examine the optimum at which the common utility level is maximized. Using this procedure, the income distribution is necessarily the one that yields equal utilities. Since we are interested in the properties of an efficient allocation in general, and not the properties of this particular income distribution, it is fortunate that many of the results in the equal-utility case either apply directly to more general cases, or approximate the results in the general case at a Pareto optimum.

The book therefore deals primarily with the efficiency aspect of externalities and public goods. The best starting point for the analysis of efficiency is the *Fundamental Theorem of Welfare Economics.* The Theorem examines the optimality of competitive equilibrium, where competitive equilibrium is, roughly speaking, the allocation at which supply equals demand for all goods, with all agents taking prices as given. Assuming that all goods are private goods and that no externalities exist, it has been shown that a competitive equilibrium is Pareto optimal under some mild regularity conditions, and that under the additional assumption of convex preferences and a convex production possibility set a Pareto optimal allocation can be achieved as a competitive equilibrium. Thus in the sense of Pareto, competitive equilibrium is optimal. This result, however, breaks down if there are externalities or public goods.

In making decisions, individuals who generate externalities do not take into account the external effect on others. Their decisions therefore must be corrected to include the external effects. Introducing a Pigouvian tax/subsidy is one way of modifying individual decisions in order to achieve an efficient allocation. When, for example, there is traffic congestion producing an externality among travelers, a Pigouvian tax on congestion can be imposed. An efficient allocation results if the tax each traveler pays is equal to the marginal cost she imposes on others by traveling. The problem with the Pigouvian tax/subsidy is that it usually requires very high administrative costs. Chapter II considers the case of a special kind of Marshallian externality, and explores the possibility of internalizing the externality through the ownership of land.

Schemes for making agents pay all the costs they impose on others are sometimes too costly. Policy makers may then want to achieve the best allocation possible when relative prices are distorted by an unpriced externality. This is the *second best problem.* The second best problem turns out to be much more complicated than the first best. In Chapter V, we examine an example of a second best problem - deciding how much road to build when congestion tolls cannot be levied.

A public good is supplied efficiently if the marginal cost is equal to the sum of the marginal benefits received by all individuals who consume the good. It is, however, extremely difficult for the supplier of the public good to know how much people benefit. In Chapter III, we examine whether it is possible to devise a competitive system that achieves an efficient allocation of *local* public goods.

We analyze externalities and local public goods within extensions of a standard

residential-land-use model. The basic features of our cities are as follows. A city is built on a flat featureless plain. All residents in the city work in the central business district (CBD) at the center. People in our model travel only between their homes and the CBD. Travel is equally costly in all directions, so that the only spatial characteristic of any location that matters is the distance from the city center. We can therefore treat the city as if it were one-dimensional.

The city may be closed, in which case the population of the city is fixed; or open, in which case migration into and out of the city is allowed. We often consider the extreme case of an open city which is small compared with the rest of the world, so that the utility level of the residents is fixed from outside. We also consider an economy consisting of many cities.

We consistently assume that commuting costs are the only transportation costs incurred in the economy. This assumption is a reasonable first approximation, since transporting human beings is much more costly than transporting most commodities. The way transportation costs are introduced marks the boundary between conventional location theory and the approach based on residential land use which was called the new urban economics by Mills and MacKinnon (1973). In location theory, there are no commuting costs, nor even workers, but transporting goods is costly.

One of the advantages of our approach is that we can assume without being logically inconsistent that producers are perfectly competitive, since if there are no transportation costs for products, they compete each other directly in the world market. In location theory a producer has monopoly power in the market area surrounding his factory because more distant producers have higher transportation costs. Competition occurs only at the boundary between different producers, and if a producer raises the price of the product, his market area becomes smaller but, in contrast to what happens in the case of perfect competition, demand for his product does not fall to zero. Since we avoid the complications arising from the monopolistic element, we can introduce other complications, such as externalities, without making the analysis intractable.

It is not our purpose to elaborate a comprehensive theory of urban externalities. Rather, we isolate each particular kind of externality in a very simple model, and focus on its special properties. We often concentrate on polar cases to obtain clear-cut results. In those cases the results should not be taken too literally: they simply illustrate the directions of basic forces which operate in more general cases.

This strategy reflects our belief that the only way to understand a very complex real world is to construct simple imaginary world, each of which includes one, or a few, important aspects of the real world, and to study their workings. Once we understand the simple models, they can be made more complicated by combining them or by introducing more realistic elements, and eventually we may understand all the important aspects of the real world. This view was eloquently expressed by R.M. Solow (1973) :

> Simplifying assumptions are not an excrescence on model-building; they are its essence. Lewis Carroll once remarked that a map on the scale of one-to-one would serve no purpose. And the philosopher of science Russell Hanson noted that if you progressed from a five-inch balsa wood model of a Spitfire airplane to a 15-inch model without moving parts, to a half-scale model, to a full-size entirely accurate one, you would end up not with a model of a Spitfire but with a Spitfire. He then remarked that if you equipped the Spitfire with illuminated tubing in red, blue and green to illustrate the fuel, ignition and hydraulic systems, it would again be a kind of model but mainly by virtue of its differences from the real thing.

Our analysis is confined to the economic aspect of cities. Sociological and political aspects enter our analysis only as an environment which is taken as given. Narrowing our scope allows us to use some of the more powerful tools from the economist's tool kit. We hope that the precision we gain justifies the generality we lose.

As in standard economic theory, we assume that rational individuals act according to consistent preferences which can be represented by a well-behaved utility function. Although we do not believe that all people behave rationally all the time, it is clearly a better assumption than that people are always fools, for example, or that every decision is made by flipping a coin. The limits of the assumption, however, must be borne in mind.

The organization of this book is as follows. In Chapter I, we formulate a simple residential land use model which serves as the basis for later developments. The model captures the trade-off between commuting costs and lot size in the simplest possible form. In order to compensate for a rise in commuting costs, the lot size must increase with distance from the center, which is made possible by a fall in land rent. We introduce the concepts of a closed city and a small open city, and analyze both competitive equilibria and optimal allocations.

We develop a model of an economy consisting of many cities in Chapter II, and analyze the optimum and market city sizes. Two cases are considered: one is the case of scale economy internal to a firm and the other is the Marshallian externality case with scale economy external to a firm but internal to a city.

Local public goods are introduced in Chapter III. We examine how the optimal supply of local public goods is achieved in a decentralized market system.

Traffic congestion and land use for transportation are introduced in Chapter IV. The optimal allocation requires that congestion tolls be levied and that roads are built to equate the marginal saving in transportation costs from widening the road with the land rent. Because of huge administrative costs, however, it is usually impossible to levy the optimal congestion tolls. In the absence of congestion tolls, the investment criterion of roads must also be modified. In Chapter IV we compare the optimal allocation with the market equilibrium where congestion tolls are not levied and roads are built according to the usual benefit-cost criterion. Since the usual benefit-cost criterion of comparing the saving in transportation costs with land rent is misleading when congestion tolls are not levied, we, in Chapter V, explore the second best allocation in which roads are built optimally under the constraint that congestion tolls are impossible.

In Chapter VI externalities between different types of individuals are introduced. Assuming that one type, called discriminators, suffers external diseconomy from the presence of the other type, called nondiscriminators, in their neighbourhood. We examine what kind of spatial pattern emerges given the externality. Using the model we analyze the possibility of a so-called cumulative decay process of a city.

Capital accumulation is introduced in Chapter VII and optimal growth of a system of identical cities is analyzed. The major question asked in the chapter is whether the city size increases in the process of capital accumulation.

There are four appendices after the main text. Appendix I analyzes a problem that arises in Chapter I. In Chapter I, it is found that households receive different utility levels at the Benthamite optimum. We will explore the reason why utility levels are different even though the Benthamite social welfare function is egalitarian. Appendix II extends the analysis of local public goods in Chapter III to a more general model. Appendices III and IV develop two useful mathematical tools. In Appendix III, the *Envelope Theorem* is explained and properties of the *indirect utility function* and the *expenditure function* are derived as applications of the Theorem. Appendix IV gives a brief review of *optimal control theory,* which is used extensively in this book.

It is probably useful to note here that equations from preceding chapters are referred to by adding the chapter number: for example, Equation (2.1) in Chapter I is called Equation (I.2.1) in other chapters.

REFERENCES

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

Solow, R.S., (1973), "Rejoinder to 'A Comment on Some Uses of Mathematical Models in Urban Economics'," *Urban Studies* 10, 267.

CHAPTER I
THE BASIC MODEL

The simple residential land use model developed in this chapter will be used later to analyze urban externalities. It is helpful, however, to examine competitive equilibrium and optimal allocation in the basic model first, as we do in sections 1 and 2 respectively.

The size and form of a city are at least partially determined by the market decisions of households which buy or rent housing. The decisions involve hundreds of factors such as the size of a lot, the size of a house, distance to the workplace, neighbourhood characteristics, the quality of the schools, the property tax rate and so on. Although all of these factors are important, in this chapter we concentrate on one of the most important: the trade-off between accessibility and lot size. Our households are constantly asking "shall we live in a town-house near work or on a larger lot in the suburbs?".

To avoid unnecessary complications, we make the following assumptions:

(a) In our city *the central business district (CBD) is the only center.* All city residents work in the CBD and commute from the surrounding residential area. This assumption does not, as it turns out, affect the residential pattern: the qualitative results are essentially the same in a multi-centered model.[1]

(b) *All households are identical.* They have the same preferences and the same number of workers. For simplicity, we assume that each household has one worker. All the workers are assumed to have the same skill. These assumptions are important in deriving some of the results. The assumption of the same skill can be easily relaxed, but it is difficult to obtain clear-cut results in a model with different preferences unless the difference in preferences is of a particularly simple nature.

(c) *The only transportation costs incurred are the costs of commuting to the CBD,* either to work or to shop. *The value of commuting time is constant* for any amount of commuting time and the same for all households. Time costs are included in the pecuniary costs of transportation. These assumptions are easily relaxed.[2]

(d) *An individual may reside at only one location.* This assumption eliminates, for example, households with an apartment in the city and a house in the suburbs. The actual number of such households is so small that they can safely be ignored. As will be seen in Appendix I on equality and the

---

[1] However, it is not easy to determine the number, locations and sizes of centers. Once they are determined, the residential patterns are obtained in essentially the same way as in a monocentric model.

[2] Henderson (1977), for example, uses a model with time costs.

Benthamite function, this assumption introduces nonconvexity, and is a major departure from the standard neoclassical theory.

(e) *Housing capital can be instanteously adjusted.* Although housing is in reality a durable good, we assume that all the characteristics of houses such as the size of a lot and the size of a house can be changed instantaneously. Ours is, therefore, a city at the imaginary long-run stationary state, in which the capital-land ratio is always perfectly adjusted. Analysis is simplified by this assumption, yet many of the results obtained in the simple polar case carry over to more complex cases. Even if different results are obtained, it serves as a useful reference point and illustrates the basic mechanism. Furthermore, the comparative static results of long-run equilibria suggest the direction of change of an urban economy to policy changes.
If we further assume that the relative prices of housing capital (buildings) and other consumer goods do not change, then by Hicks' Aggregation Theorem houses can be treated as part of the consumer good.[3] The assumption allows us to concentrate on the amount of land used for housing.

(f) *Transportation requires no land input.* We also assume away traffic congestion so that commuting costs are simply a function of the distance from the CBD. This assumption will be relaxed in Chapters IV and V.

(g) *There are no externalities and no public goods.* This assumption will also be relaxed in later chapters. Externalities among producers will be examined in Chapter II; local public goods in Chapter III; traffic congestion in Chapters IV and V; and externalities between different types of individuals in Chapter VI.

1. Market Cities

In this section we analyze competitive equilibrium of a city. The equilibrium spatial structure is examined in subsection 1.1. It is assumed that all residents receive the same income. Because everyone is assumed to have the same utility function, the utility level must be the same everywhere in the city. Land rent, thus, declines with distance from the CBD to offset an increase in commuting costs. As the relative price of land falls, consumption of land increases while consumption of the consumer good decreases. It follows that population density declines with distance from the center, as observed in most cities in the world. Furthermore, if the commuting cost is a linear or concave function of distance, the rent function must be a convex function of distance.

We consider different income classes in subsection 1.2 although we continue to assume that households are identical in all other respects: all households have the same preferences and transportation costs. Under these assumptions, richer households live farther from the center than poorer households if land is a normal good. This result follows from the fact that richer households have a flatter rent curve at the boundary. The rent must fall with distance from the center in order to offset an increase in commuting costs, but the required fall is smaller for richer households since under the normality assumption they consume more land, and therefore benefit more from the same fall in rent.

---

[3] See Hicks (1946, pp. 312-313).

In subsections 1.1 and 1.2, the utility levels and the incomes of residents are left undetermined. Two ways of determining these variables are introduced in subsections 1.3 and 1.4. The more popular formulation is that of a *closed city,* which assumes that the population of a city is given. This type of model may be interpreted as dealing with a time period long enough to attain an equilibrium within a city, but too short to allow migration between cities. Since it takes a long time to change the housing stock, this interpretation is somewhat schizophrenic.

It is more consistent to interpret the closed city model as the long-run stationary equilibrium of a closed homogeneous economy with given population, a given number of identical cities and an insignificant rural sector. The population of a single city is then given by simple division.

As a natural extension of this interpretation, we can take the number of cities as a variable. A non-urban sector such as an agricultural sector can also be introduced so that migration between urban and nonurban sectors can be analyzed. These extensions are considered in the next chapter on city sizes.

In subsection 1.4 we examine a *small "open" city,* where openness means that migration of households and transportation of products between cities are costless and otherwise unrestricted. In an open city, commodity prices and the utility level of residents are equal to those in the rest of the economy. When an open city is small compared with the entire economy, any change in allocation within the city will spread over the whole economy and local prices and utility level will not be affected significantly. Prices and the utility level may, therefore, be taken as given for the city.

This model is appropriate when the long-run allocation of a city is the focus. A city administrator, for example, may want to adopt this model to analyze the long-run effects of his policies. The model may also be applied to cities in developing countries with surplus labour, or to cities in a small country which allows free migration.

In both open and closed cities we have to distinguish between the *"absentee-landlord"* case, in which land is owned by absentee landlords who spend their incomes outside the city, and the case of *"public ownership"*. In our treatment of public ownership a city government rents the land from agricultural landowners at the agricultural rent and sublets it to households at the market rent, using the net revenue to subsidize city residents equally.

1.1. The Spatial Structure of a Residential City

Consider a city in a featureless agricultural plain. To simplify exposition, we assume that production does not require space, so that the CBD is just a point.[4] The residential zone extends to distance $\bar{x}$ from the CBD. The analysis may be applied to any shape, but it is often easiest to imagine dealing with a circular city. In any ring between radius x and $x+dx$, there are $\theta(x)dx$ units of land available, out of which $L_H(x)dx$ units are used for housing. The structural component of housing is included in the composite consumer good. At the edge of the residential zone the residential

[4] It is not difficult to introduce land use for urban production. See Appendix II for this extension in the context of local public goods.

rent must be equal to the rural rent.

One person from each household commutes to the CBD. The commuting costs, $t(x)$, for a household at a radius *x*, are assumed to be an increasing function of distance from the center:

$$t'(x) > 0 \quad . \tag{1.1}$$

Consumption of the composite consumer good, which includes buildings, and consumption of land for housing are denoted by $z(x)$ and $h(x)$ respectively. Transporting the consumer good is

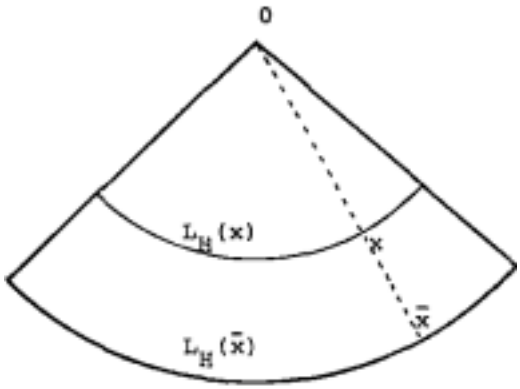

Figure 1. The residential zone

costless. All households have the same quasi-concave utility function,

$$u = u(z,h) \quad . \tag{1.2}$$

We assume that the utility function is appropriately differentiable, although it is not necessary for all the results that follow.

The budget constraint for a household at x is

$$I(x) \equiv y - t(x) = z(x) + R(x)h(x), \tag{1.3}$$

where $I(x), y,$ and $R(x)$ are respectively the income net of the commuting costs, the gross income, and the residential land rent. The rent function, $R(x)$, provides the rent for a unit area of land at any given radius. The gross income is assumed to be the same for every household. How the income level is determined will be specified later. Note that the consumer good is taken as the numeraire.

A household maximizes the utility function, (1.2), subject to the budget constraint, (1.3). The first order condition for this maximization problem is

$$\frac{u_h}{u_z} = R(x), \tag{1.4}$$

where subscripts *h* and *z* denote partial derivatives with respect to *h* and *z*. This is the

familiar condition that the price ratio and the marginal rate of substitution are equal. From this first order condition and the budget constraint, demands for the consumer good and land can be written as functions of the net income, $I(x) \equiv y - t(x)$, and land rent, $R(x)$:

$$z(x) = \hat{z}(I(x), R(x)) , \tag{1.5}$$

$$h(x) = \hat{h}(I(x), R(x)) . \tag{1.6}$$

Since these functions describe the levels of demand obtained at a fixed income level, they are nothing but *uncompensated* (*or Marshallian*) *demand functions.* By substituting (1.5) and (1.6) into the utility function, we obtain the *indirect utility function,*

$$v(I(x), R(x)) \equiv u[\hat{z}(I(x), R(x)), \hat{h}(I(x), R(x))] , \tag{1.7}$$

which describes the maximum utility level available to consumers, given the net income, $I(x)$, and land rent, $R(x)$.[5]

The demand functions satisfy the following useful relationships obtained by differentiating the budget constraint (1.3):

$$h + R\hat{h}_R + \hat{z}_R = 0 \quad , \tag{1.8}$$

$$R\hat{h}_I + \hat{z}_I = 1 \quad , \tag{1.9}$$

where subscripts R and I denote respectively partial derivatives with respect to $R(x)$ and $I(x)$. Using these equations, we can see that the indirect utility function satisfies *Roy's Identity*[6]*:*

$$v_R = -v_I h \quad . \tag{1.10}$$

Since households are identical, in equilibrium the utility level must be the same everywhere in the city. Otherwise, households at a place of lower utility level have an incentive to relocate, and the allocation cannot be a market equilibrium. Thus the land

---

[5] See Section 3 of Appendix III on the envelope property for discussions of the indirect utility function in conjunction with the Envelope Theorem.

[6] Roy's Identity is derived in the following way. From (1.7), partial derivatives of $v(I, R)$ are given by

$$v_R = u_z\left(\hat{z}_R + \frac{u_h}{u_z}\hat{h}_R\right)$$

$$v_I = u_z\left(\hat{z}_I + \frac{u_h}{u_z}\hat{h}_I\right).$$

In view of (1.8) and (1.9), substitution of (1.4) into these equations yields

$$v_R = -v_I h \quad .$$

See Section 3 of Appendix III for a more elegant way of deriving Roy's Identity which makes use of the Envelope Theorem.

rent must satisfy

$$v(y - t(x), R(x)) = u = \text{const.} \quad , \tag{1.11}$$

which can be solved for $R(x)$ to yield

$$R(x) = R(y - t(x), u) \quad . \tag{1.12}$$

This function is called the *bid rent function.* It describes the maximum rent which a household can pay at a particular distance from the center if it is to receive the given utility level. If the utility level and the income level are known, the bid rent function gives the equilibrium rent. This is merely a result of the rational behaviour of households. If, for example, the actual rent were lower than the bid rent, it would be possible to achieve a higher utility level, and a rational household would not fail to do so. The actual rent cannot be higher than the bid rent simply because it is impossible to pay any higher rent and achieve the given utility level. The bid rent function is extremely useful in a model with one type (or a few types) of households, since in each type the income and the utility level must be the same at any distance from the center. The bid rent function summarizes, in a single function, the rent profile that is compatible with the given income and utility levels.

At the edge of the city, where $x = \bar{x}$, the residential rent must equal the rural rent R_a:

$$R(\bar{x}) = R_a \quad . \tag{1.13}$$

Given the levels of income and utility, (1.12) and (1.13) completely determine the rent profile. Once the rent profile is determined, the allocation of a city is fully characterized, since (1.5) and (1.6) give the consumption of the consumer good and of land for housing at each location.

In this simple model, the transportation cost function and the utility function completely determine the spatial structure of the city as Figure 2 illustrates. Consider any two locations, x_1, and x_2, where x_1 is closer to the center than x_2. Inspection of the budget constraint (1.3) shows that a budget line intersects the vertical axis at $y - t(x)$. Since the utility level is maximized under the budget constraint, the budget line must be tangent to an indifference curve at the optimum. If the utility level is the same everywhere in the city, households are on the same indifference curve, *u*, at any location *x*. The budget line is thus fully determined and the consumption of the consumer good and land can be read off.

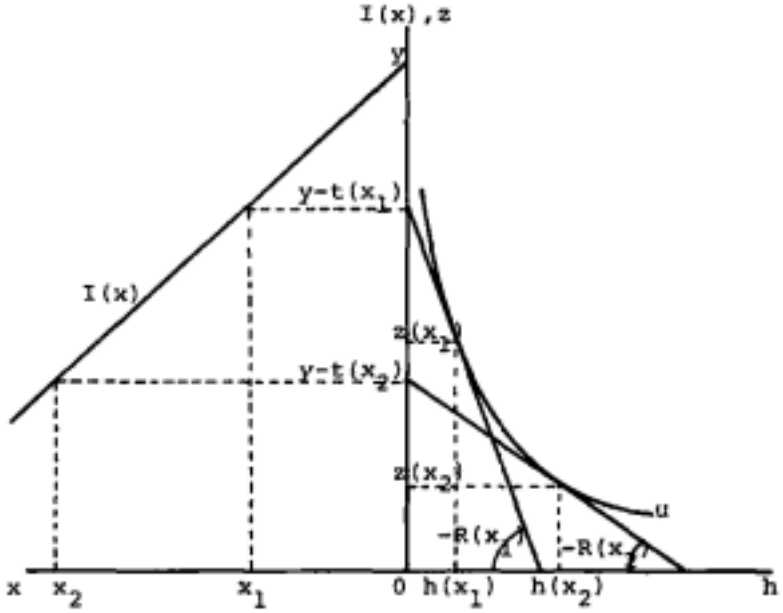

Figure 2. Allocation in the basic model

The bid rent is given by the slope of the budget line. Convexity of indifference curves implied by quasi-concavity of the utility function ensures that the bid rent is lower at x_2 than at x_1. That is, the bid rent curve, $R(I(x),u)$, is a decreasing function of distance x from the center. Furthermore, the lot size increases and the consumption of the consumer good decreases with distance from the center, as households substitute land for the consumer good.

More precise properties can be derived by using calculus. From (1.11) and Roy's Identity (1.10), the rent profile satisfies the following simple relationships:

$$R_I = 1/h(x) \tag{1.14}$$

$$R_u = -1/v_I h(x) \quad . \tag{1.15}$$

Thus, demand for land is a reciprocal of the partial derivative of the bid rent function with respect to income. Differentiating (1.12) and substituting (1.14) yields

$$R'(x) = -t'(x)/h(x) < 0\,, \tag{1.16}$$

which shows that *the land rent declines with distance from the center.*

If demand functions are obtained for a given utility level instead of a given income level, we have *compensated* (*or Hicksian*) *demand functions:*[7]

$$z(x) = z(R(x),u) \tag{1.17}$$

$$h(x) = h(R(x),u) \tag{1.18}$$

[7] See Section 3 of Appendix III for a derivation of the compensated demand function and its properties from the expenditure function.

The compensated demand functions are useful since the signs of partial derivatives are unambiguous:

$$z_R \geq 0 \tag{1.19}$$

$$h_R \leq 0\,. \tag{1.20}$$

The first inequality is a result of the fact that if there are only two goods, they are always net substitutes. The second inequality represents the elementary property that the (own) substitution effect is negative.

The slopes of $z(x)$ and $h(x)$ are obtained from (1.16), (1.19) and (1.20):

$$z'(x) = z_R R'(x) = -\frac{t'(x)}{h(x)} z_R \leq 0 \tag{1.21}$$

$$h'(x) = h_R R'(x) = -\frac{t'(x)}{h(x)} h_R \geq 0\,. \tag{1.22}$$

The consumption of the consumer good is a nonincreasing function and the lot size a nondecreasing function of distance. The latter property is used by urban economists to explain the fact that the population density declines with distance from the center in most cities.

Differentiating (1.16) again, we obtain

$$R''(x) = -\frac{t''(x)}{h(x)} + \frac{h'(x)}{(h(x))^2} t'(x) \tag{1.23}$$

From (1.22), a sufficient condition for $R''(x) > 0$ is that $t''(x)$ is nonpositive. This yields another well-known result: *if the commuting cost is a linear or concave function of distance from the CBD, the rent function is convex.*

We were able to treat $z(x)$ and $h(x)$ as choice variables because we assumed that housing capital is extremely cooperative. We have ignored a very important aspect of the housing market: the durability of the housing stock. The model therefore describes a long-run stationary state which may never come to exist. In order to introduce durability we would have to develop a dynamic model, making analysis much more complicated.

1.2. Several Income Classes

The above analysis can be easily extended to include different types of households.[8] In this section we consider the case where there are two income classes. For simplicity, and in accordance with empirical observations, land is assumed to be a *normal good:*

[8] Although everybody is assumed to have the same skill, households can have different incomes since they may own different shares of firms and land.

$$\hat{h}_I[I(x), R(x)] > 0 .$$

Assuming normality, we can show that *there is segregation by income*: the residential zone is divided into two rings, each occupied by one income class. Moreover, we can show that *the richer group lives in a ring farther from the center*, which agrees with the actual residential pattern in most American cities. The argument is quite direct.

Space is occupied by those who are willing to pay the highest rent for it. In other words, the equilibrium rent at any point is simply the highest of the bid rents at that point. Now, the bid rents are functions of income and utility levels, and the rich have higher incomes than the poor: $y^r > y^P$.

At some radius x^*, rich and poor living in the same city must live side by side. This radius is the boundary between two rings of households with different incomes. At this location the two income groups must pay the same rent. From (1.16), the bid rent function is steeper for the lower income group since t' is the same for both groups, and by the normality assumption the lower income group consumes a smaller amount of land. It follows that the richer income group has the higher bid rent outside x^* and lives there. Thus the equilibrium residential pattern is complete segregation with the richer income class living in the outer ring.[9]

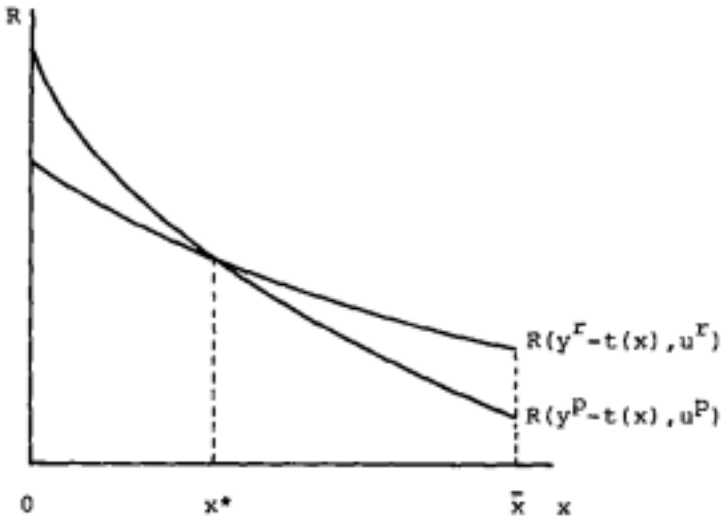

Figure 3. Two income classes

The flatter bid rent curve of the rich can be understood as follows. Suppose that as a poor household moved outwards, the loss of utility due to increased commuting costs was just offset by an increase in utility arising from increased land use. Clearly, this is possible only if the rent on a unit of land falls. But since richer households have larger lot sizes, the same decline in rent allows them larger savings in the total

[9] For arbitrary utility levels, it is possible that the bid rent of one income class is higher than that of the other everywhere in the city. In such a case only one income class lives in the city. The utility levels must be adjusted in order for both groups to live in the city.

expenditures on land. Richer households, therefore, benefit more from the same fall in rent and would be willing to accept a smaller decline in rent as they move outward. The bid rent curve of the rich thus falls less rapidly with distance from the center.

This result has been used to explain the residential pattern observed in the United States. However, it crucially depends on the assumption that all income classes have the same commuting costs. Since time costs constitute a large portion of commuting costs, richer households may live closer to the center if their value of time is much higher than poorer households'. This may explain why the opposite spatial pattern is observed in most cities in Europe, Latin America and Japan, as well as the existence of high-rent luxury apartments near the center of most cities. According to an empirical study (1977) by Wheaton, if time costs are taken into account, the tendency of wealthier households to move to the periphery is weak even in American cities. This suggests that the observed pattern is mainly caused by other factors, such as the concentration of older houses in central cities.

1.3 A Closed City

In the previous subsection, important variables such as incomes and utility levels were left undetermined. In this and the following subsections, different ways of determining them are introduced. For simplicity, we consider cities with only one income class.

The analysis in subsection 1.1 shows that the allocation of a city is completely determined by utility maximization of households and spatial arbitrage, if the utility level, the income level and the size of the city are specified. Since we already have condition (1.13) as one of the three equations required to determine these variables, only two more equations must be specified.

In this subsection, we consider a *closed city;* immigration into and out of the city is impossible and therefore the population is *fixed.* For convenience, the population is identified with the number of households. Denoting the total population of the city by P, the *population constraint* is

$$P = \int_0^{\bar{x}} N(x)dx \tag{1.24}$$

where $N(x)dx$ is the number of households living between x and $x+dx$. Recalling that $L_H(x)$ and $h(x)$ denote respectively the total land available for housing and the lot size at radius x, we can write $N(x)$ as

$$N(x) = \frac{L_H(x)}{h(x)} \tag{1.25}$$

The *aggregate production function* is

$$Y = F(P), \tag{1.26}$$

where all factors other than labor are assumed to be fixed and suppressed. If a city resident is paid the value of the marginal product of labor, the wage rate is given by $w = F'(P)$. If city residents collectively own firms and factors other than labor, a city resident will receive the average product, $F(P)/P$. In either case wages are a fixed

amount w if the population is fixed.

Income may differ from wages depending on the treatment of land rent. We consider only the two polar cases; the "*absentee-landlord*" case, and the "*public-ownership*" case. Intermediate cases are left to the reader. In the absentee-landlord case, land is owned by landlords who do not live in the city, and the rent is spent outside the city. The income of a resident is simply the given wage rate:

$$y = w . \tag{1.27}$$

(1.24) and (1.27) give our missing two equations and the allocation of the city is completely determined.

The absentee-landlord case is used more often in descriptive analysis to avoid an artificial institutional arrangement. If the optimality of an allocation is a major issue, however, the absentee-landlord case is not convenient because the welfare of absentee landlords has to be taken into account, forcing us to compare utilities of landlords and tenants. We shall therefore adopt the public-ownership framework in normative analysis.

For the public-ownership case we assume the following rather artificial institutional arrangement. The city residents form a government which rents the land for the city from rural landlords. We assume that landlords cannot obtain any monopolistic power, so that the city government needs only to pay the rural rent R_a. The city government, in turn, subleases the land to city residents at the competitively determined rent, $R(x)$. The net revenue is divided equally among households.

There is $\theta(x)dx$ of land between x and $x+dx$, out of which the city sublets $L_H(x)dx$ to city residents and uses the rest for public purposes such as roads and parks. The net revenue of the government is then given by

$$\int_0^{\bar{x}} \left[R(x)L_H(x) - R_a\theta(x)\right]dx \quad .$$

The income of a household is the sum of wages and the "*social dividend*" it receives from the city government:

$$y = w + \frac{1}{P}\int_0^{\bar{x}} \left[R(x)L_H(x) - R_a\theta(x)\right]dx \tag{1.28}$$

We temporarily assume that the entire land is rented to city residents for residential use:

$$L_H(x) = \theta(x) \qquad\qquad 0 \le x \le \bar{x} \tag{1.29}$$

We shall relax this assumption in Chapter IV when we introduce land for transportation use.

(1.28) describes how factor incomes are allocated. If we consider how the goods are allocated, the following constraint is obtained:

$$Pw = \int_0^{\bar{x}} \left\{\left[z(x) + t(x)\right]N(x) + R_a\theta(x)\right\}dx \tag{1.30}$$

The city residents collectively command Pw units of the consumer good, which are consumed or spent on commuting costs and the payment of the rural rent. This

constraint is a *resource constraint* that the city faces and will be used in the optimization framework. The equivalence of (1.28) and (1.30) can be readily derived by using the budget constraint (1.3).

1.4 A Small and Open City

A perfectly closed city is one where migration in and out is impossible. It is useful to consider the case in which migration is possible. We assume that migration of households and transportation of products between cities are completely costless. We further assume that the city is so small that any change within the city does not affect the outside world. Prices and the utility level within the city, therefore, equal world levels and may be taken as given.

Since the population size is endogenous in an open city, wages cannot in general be taken as exogenous.[10] Therefore, the income of a household is

$$y = w(P)$$

in the absentee-landlord case, and

$$y = w(P) + \frac{1}{P}\int_0^{\bar{x}} \left[R(x)L_H(x) - R_a\theta(x)\right]dx$$

in the public-ownership case. Either of these equations, if coupled with (1.24), determines the population size and the income level, and thereby completely specifies the resource allocation in the city.

Although it is possible that the city government would be controlled by old residents who treat newcomers differently, as in some of the club theory literature, for example, McGuire (1974), we shall not pursue this line here. We assume that newcomers receive all privileges of citizenship including a share of net city revenue.

If a city is not small but open, a case intermediate between a closed city and a small city is obtained. Given the total population of the economy, the population of the rest of the economy can be expressed in terms of the population, P, of the city. When households leave the city, the marginal product of labour rises in the city and falls elsewhere, as a result of diminishing returns. Since migration is free, equilibrium will be reached when the utility level outside the city, $V(P): V'(P) > 0$, equals the utility level in the city:

$$u = V(P).$$

This condition replaces the fixed-population constraint in a closed city and the fixed-utility constraint in a small city. This more general formulation will be used in Chapter VI. Note that the polar cases of $V'(P) = 0$ and $V'(P) = \infty$ yield a small city and a closed city respectively.

---

[10] If, however, constant returns to scale are assumed and a resident receives the average product, w is constant. This assumption is quite often made (at least implicitly) in the literature.

2. Optimum Cities

To obtain an optimal allocation, an objective, or criterion, function must be specified. Probably the most natural one is a Benthamite social welfare function which is the sum of the utilities of individual households,

$$\int_0^{\bar{x}} u(z(x),h(x))N(x)dx\,. \tag{2.1}$$

Note that the Benthamite social welfare function requires that utility be cardinal.[11] In addition it is commonly assumed that the marginal utility of income decreases as income increases. This is a cardinal property and it is represented by the assumption that the utility function is concave.

We can imagine the Benthamite optimum being achieved as follows.[12] Let an individual choose the optimal resource allocation, including income distribution, based on her own selfish preferences. Decisions must be made, however, "behind the veil of ignorance": she must not know which of the residents she will become. If she has an equal chance of becoming any of the residents, her expected-utility maximization is equivalent to maximizing the Benthamite social welfare function.

It turns out that at the Benthamite optimum the utility level varies with the distance from the center. When land is a normal good, the utility level rises with distance from the CBD. It also turns out that for an appropriate unequal income distribution the corresponding competitive equilibrium exactly replicates the optimum solution.

Theorists have been intrigued to find that the optimal utility levels differ among locations even though the social welfare function is egalitarian. This result is surprisingly robust, at least among additive social welfare functions. It can be explained as follows. Because of the difference in commuting costs, identical households at different locations have different capability to generate utility from the same amount of resource. The Benthamite optimum, therefore, is attained if more resource is allocated to the more efficient households.

As Appendix I shows, the difference in the efficiency with which households realize utility from their commodity bundles arises from the most fundamental properties of our spatial allocation problem. We assumed that a household cannot live at more than one location. Each household, therefore, must choose one location, and every location has an associated commuting cost. Identical households with equal incomes, once they choose different locations and hence different consumption bundles, are in effect no longer identical. If households are able to divide their time among two or more residences, however, every household faces the same opportunity set and the inequality of utility levels will disappear.

---

[11] If utility is merely ordinal, any monotonic transformations of a utility function are considered as equivalent. A monotonic transformation can, however, yield a different Benthamite optimum. In order to obtain the same Benthamite optimum, we must assume that utility functions are equivalent only up to linear transformations, i.e., utility is cardinal.

[12] See, for example, Arnott and Riley (1977).

Even if the social welfare function is made more egalitarian by taking a concave transformation of the utility function - that is, if a new social welfare function,

$$\int_0^{\bar{x}} \phi[u(z(x),h(x))]N(x)dx,$$

is adopted - the optimal allocation continues to have unequal utility levels. This conclusion follows immediately from the observation that even if we redefine the utility function as $U(\cdot) = \phi(u(\cdot))$, our assumptions on the original utility function still hold for the new one.

The only way of obtaining an equal utility level with an additive social welfare function is to take a limit coinciding with the Rawlsian welfare function, which maximizes the minimum utility level. For example, Dixit (1973) considered the welfare function

$$\int_0^{\bar{x}} -u(z(x),h(x))^{-m} N(x)dx$$

and obtained a uniform utility level by taking the limit as $m \to \infty$. Appendix I contains a detailed discussion of why utility levels differ between different locations except in the limit.

Some economists prefer the Benthamite welfare function on the grounds that the Rawlsian welfare function has the undesirable property of ignoring the welfare of all but the poorest individual. Although the Rawlsian function is the only *additive* social welfare function that yields equal utility, there are other nonadditive functions that will do. As shown in Appendix I, equal utility requires social welfare indifference curves to have sufficiently strong kinks on the line where utility levels are equal.

Except in this section we will consider only cases where utility levels are equal for identical households. The reason is twofold. First, this case is mathematically more tractable, and easier to compare with the market equilibrium. Second, readers might object to giving different utility levels to households which differ *only* in the location of their residences.

2.1 A Closed City

In this subsection, we consider optimal allocation of a closed city. Only the public-ownership case is analyzed because in the absentee-landlord case the welfare of absentee landlords must be taken into account, which destroys the simple structure of our problem. The total amount, *Y*, of the consumer good produced in the city is used for direct consumption, transportation, and the payment of the rural rent. The *resource constraint* for the city is then

$$Y = \int_0^{\bar{x}} [(z(x) + t(x))N(x) + R_a\theta(x)]dx \tag{2.2}$$

which corresponds to (1.30) in the previous section. The city

also faces the *population constraint,* (1.24), and the *land constraint,*

$$\theta(x) = N(x)h(x), \quad 0 \le x \le \bar{x} \tag{2.3}$$

The land constraint is obtained by combining (1.25) and (1.29).

The objective function is the Benthamite social welfare function (2.1). The Lagrangian for this problem is

$$\Lambda = \int_0^{\bar{x}} u(z(x),h(x))N(x)dx + \delta\left\{Y - \int_0^{\bar{x}}\left[(z(x)+t(x))N(x) + R_a\theta(x)\right]dx\right\}$$
$$+ \gamma\left[P - \int_0^{\bar{x}} N(x)dx\right] + \int_0^{\bar{x}} \mu(x)\left[\theta(x) - N(x)h(x)\right]dx\,, \qquad (2.4)$$

where δ, γ and $\mu(x)$ are respectively Lagrange multipliers associated with (2.2), (1.24) and (2.3). δ can be interpreted as the shadow price of the consumer good, γ the shadow 'price' of a household (with the total production in the city fixed), and $\mu(x)$ the shadow rent of land, all in utility terms. The shadow 'price' of a household may sound peculiar, but it naturally appears in our problem because an increase in population changes the maximum value of the Benthamite social welfare function. The choice variables are $z(x)$, $h(x)$, $N(x)$, and $\bar{x}$, where $z(x)$, $h(x)$, and $N(x)$ are chosen at each x between 0 and $\bar{x}$.

As shown in section 4 of the appendix on optimal control theory, control theory may be applied to this problem and the following first order conditions are immediately obtained:

$$u_Z(z(x),h(x)) = \delta\,, \qquad 0 \le x \le \bar{x}\,, \qquad (2.5a)$$

$$u_h(z(x),h(x)) = \mu(x)\,, \qquad 0 \le x \le \bar{x}\,, \qquad (2.5b)$$

$$u(x) = \delta\left[z(x)+t(x)\right] + \mu(x)h(x) + \gamma\,, \qquad 0 \le x \le \bar{x}\,, \qquad (2.5c)$$

$$\left[u(\bar{x}) - \delta(z(\bar{x})+t(\bar{x})) - \gamma\right]N(\bar{x}) = \delta R_a\theta(\bar{x})\,. \qquad (2.5d)$$

Using (2.5c), (2.5d) can be written

$$\mu(\bar{x}) = \delta R_a \qquad (2.5d')$$

(2.5a) and (2.5b) require that the marginal utility of the consumer good equal its shadow price, and that the marginal utility of land equal the shadow rent at each radius. (2.5c) means that the utility level of a household equals the shadow value of its consumption bundle plus the shadow 'price' of a household. A household at x contributes to the social welfare by $u(x)$, but consumes resources whose value is $\delta\left[z(x)+t(x)\right] + \mu(x)h(x)$. The difference is the marginal social value of a household, or the shadow 'price' of a household, γ. According to (2.5d'), the shadow rent of the city equals the rural rent times the shadow price of the consumer good at the optimum.

If the utility function is concave and land is a normal good, we can also show that *the utility level rises with distance from the center at the Benthamite optimum.* Differentiating (2.5c) with respect to x and substituting (2.5a) and (2.5b) yields

$$\mu'(x) = -\delta t'(x)/h(x) < 0\,. \qquad (2.6)$$

Thus the shadow rent is a decreasing function of distance from the center. The desired result follows if the optimal utility level is a decreasing function of the shadow rent.

Implicit differentiation of (1.3) and (1.4) yields the income derivative of the uncompensated demand function for land:

$$\hat{h}_I(I,R) = \frac{u_Z}{D}(u_{hZ}u_Z - u_h u_{ZZ}), \tag{2.7}$$

where

$$D \equiv 2u_{hZ}u_Z u_h - u_h^2 u_{ZZ} - u_Z^2 u_{hh}. \tag{2.8}$$

Since D is nonnegative when the utility function is quasi-concave, (strong) normality of land, $\hat{h}_I > 0$, implies that

$$u_{hz}u_z - u_h u_{zz} > 0. \tag{2.9}$$

From (2.5a) and (2.5b), $z(x)$ and $h(x)$ can be written as functions of $\mu(x)$ and δ: $\tilde{z}(\mu(x),\delta)$ and $\tilde{h}(\mu(x),\delta)$, and the optimal utility level as

$$u^*(x) = u\left[\tilde{z}\mu(x),\delta),\tilde{h}(\mu(x),\delta)\right] \equiv \tilde{u}(\mu(x),\delta).$$

Differentiating (2.5a) and (2.5b), we obtain

$$\frac{\partial \tilde{z}}{\partial \mu} = -\frac{u_{zh}}{u_{zz}u_{hh} - (u_{hz})^2},$$

$$\frac{\partial \tilde{h}}{\partial \mu} = \frac{u_{zz}}{u_{zz}u_{hh} - (u_{hz})^2}.$$

From these equations, we get

$$\frac{\partial \tilde{u}}{\partial \mu} = u_z \frac{dz}{d\mu} + u_h \frac{dh}{d\mu} = \frac{u_h u_{zz} - u_z u_{zh}}{u_{zz}u_{hh} - (u_{zh})^2}. \tag{2.10}$$

This is negative since the denominator is nonnegative when the utility function is concave and the numerator is negative from (2.6). Therefore, from (2.9) we obtain

$$\frac{du^*}{dx} = \frac{\partial \tilde{u}}{\partial \mu}\mu'(x) > 0. \tag{2.11}$$

Thus, the optimal utility level rises with distance from the center.

Next, we examine whether the optimal allocation is attained as a competitive equilibrium. An allocation is a competitive equilibrium in our model if the following conditions are satisfied:

(i) Each household maximizes the utility level with respect to z and h subject to the budget constrain and taking the land rent, $R(x)$, as given.

(ii) No household has an incentive to move to other locations.

(iii) Demand for land equals supply of land.

(iv) Demand for the consumer good equals the supply of the consumer good.

(v) The rent at the edge of the city equals the rural rent.

Defining $R(x) \equiv \mu(x)/\delta$ and $y(x) \equiv (u(x)-\gamma)/\delta$, (2.5a) through (2.5c), (2.5d') and (2.6) can be rewritten

$$u_z(z(x),h(x)) = \text{const.}, \qquad 0 \le x \le \bar{x}, , \tag{2.12a}$$

$$u_h / u_z = R(x), \qquad 0 \le x \le \bar{x}, \tag{2.12b}$$

$$y(x) = z(x) + R(x)h(x) + t(x), \qquad 0 \le x \le \bar{x}, \tag{2.12c}$$

$$R(\bar{x}) = R_a, \tag{2.12d}$$

$$R'(x) = -t'(x)/h(x), \qquad 0 \le x \le \bar{x} \tag{2.12e}$$

Condition (i) is satisfied at the Benthamite optimum since (2.12b) is the first order condition for the problem of maximizing the utility function, $u(z,h)$, subject to the budget constraint, $y(x) = z + R(x)h + t(x)$, with respect to z and h.

Condition (ii) is satisfied if a household living at any radius x^* achieves its maximum utility at x^*, that is, a household with income $y = y(x^*)$ maximizes the indirect utility function, $v(y - t(x), R(x))$, with respect to x at x^*. The first order condition for the maximization is

$$\frac{dv}{dx} = -v_I\left[R'(x)\hat{h}(y - t(x), R(x)) + t'(x)\right] = 0, \tag{2.13}$$

where we used Roy's Identity (1.10), and $\hat{h}(\cdot)$ is the uncompensated demand for land (1.6). (2.12e) ensures that (2.13) is satisfied at the Benthamite optimum. The second order condition is

$$\frac{d^2v}{dx^2} = -v_I\left[t''(x) + R''(x)h(x) + R'(x)(-\hat{h}_I t'(x) + \hat{h}_R R'(x)\right] \le 0. \tag{2.14}$$

Since (2.13) is satisfied at each x if $y = y(x)$, we have

$$R'(x)\hat{h}(y(x) - t(x), R(x)) + t'(x) = 0, \qquad 0 \le x \le \bar{x}\ 0. \tag{2.15}$$

Differentiating this equation with respect to x yields

$$t''(x) + h(x)R''(x) + R'(x)[\hat{h}_I(y'(x) - t'(x)) + \hat{h}_R R'(x)] = 0. \tag{2.16}$$

Using this equation, the second order condition becomes

$$\frac{d^2v}{dx^2} = v_I \hat{h}_I R'(x) y'(x) \le 0 \qquad 0 \le x \le \bar{x}, \tag{2.17}$$

which is satisfied at the Benthamite optimum since from (2.11) we have

$$y'(x) = u'(x)/\delta > 0 \tag{2.18}$$

if $\hat{h}_I > 0$. This also shows that the income level rises with distance from the center in market equilibrium, and corresponds to the result in subsection 1.2 that if land is normal, richer households live farther away from the center than poorer households.

Conditions (iii), (iv) and (v) are guaranteed by (2.3), (2.2) and (2.12d). Thus the

Benthamite optimum is attained as a competitive equilibrium for a suitable choice of income distribution.

Now we add the constraint that the utility level be equal everywhere in the city, and maximize this equal utility level. Thus, our problem is one of maximizing

$$\int_0^{\bar{x}} uN(x)dx \tag{2.19}$$

subject to the resource constraint (2.2), the population constraint (1.24), the land constraint (2.3), and the constraint that the utility level be equal everywhere in the city,

$$u = u(z(x), h(x)), \qquad 0 \le x \le \bar{x}. \tag{2.20}$$

The Lagrangian for this problem is

$$\begin{aligned}\Lambda = &\int_0^{\bar{x}} uN(x)dx + \int_0^{\bar{x}} \nu(x)\left[u(z(x), h(x)) - u\right]dx \\ &+ \delta\left\{Pw - \int_0^{\bar{x}}\left[(z(x) + t(x))N(x) + R_a\theta(x)\right]dx\right\} \\ &+ \gamma\left[P - \int_0^{\bar{x}} N(x)dx\right] + \int_0^{\bar{x}} \mu(x)\left[\theta(x) - N(x)h(x)\right]dx\end{aligned} \tag{2.21}$$

The only new Lagrange multiplier is $\nu(x)$ which can be interpreted as the weights that have to be attached to the utilities of households at different locations if all households are to obtain equal utility levels.

As shown in section 4 of the appendix on optimal control theory, the first order conditions are

$$\nu(x)u_z - \delta N(x) = 0 \qquad 0 \le x \le \bar{x} \tag{2.22a}$$

$$\nu(x)u_h - \mu(x)N(x) = 0 \qquad 0 \le x \le \bar{x} \tag{2.22b}$$

$$u - \delta\left[z(x) + t(x)\right] - \gamma = \mu(x)h(x) \qquad 0 \le x \le \bar{x} \tag{2.22c}$$

$$\left[u - \delta(z(\bar{x}) + t(\bar{x})) - \gamma\right]N(\bar{x}) = \delta R_a\theta(\bar{x}) \tag{2.22d}$$

$$\int_0^{\bar{x}} N(x)dx = \int_0^{\bar{x}} \nu(x)dx \tag{2.22e}$$

The difference from the Benthamite case mainly lies in (2.22a). Here, the marginal utility of the consumer good does not need to be equal at different locations, while the utility level is equal. In the Benthamite case, the marginal utility is equal but the utility level is not.

Defining $R(x) \equiv \mu(x)/\delta$ and $y \equiv (u - \gamma)/\delta$, (2.22a) through (2.22e) can be written

$$u_h / u_z = R(x), \qquad 0 \le x \le \bar{x} \tag{2.23a}$$

$$y = z(x) + R(x)h(x) + t(x), \qquad 0 \le x \le \bar{x} \tag{2.23b}$$

$$R(\bar{x}) = R_a\,, \tag{2.23c}$$

$$\frac{1}{\delta}\left[\int_0^{\bar{x}} N(x)dx\right] = \int_0^{\bar{x}} \frac{l}{v_1} Nx)dx \quad . \tag{2.23d}$$

Comparison of these equations with market equilibrium conditions in section 1 shows that the optimal solution exactly coincides with the market allocation of a closed city. (2.23d) shows that the reciprocal of the social value of the numeraire is equal to the average of reciprocals of marginal utilities of income.

2.2 A Small and Open City

In a small, open city it is meaningless to maximize the utility level of city residents because the level is determined independently of the allocation within the city. Under some circumstances, however, maximizing the net product of the city may be of interest: a mining company, for example, building a townsite on its own land would maximize the total product of the city minus the cost of maintaining the utility level required to attract a work force. The profit for such a producer would be

$$F\int_0^{\bar{x}} N(x)dx - \int_0^{\bar{x}} [(z(x)+t(x))N(x) + R_a\theta(x)]dx \tag{2.24}$$

Labour costs do not include the land rent that workers pay since it is paid to the company.

The net product (2.24) is maximized under the land constraint (2.3) and the utility constraint,

$$u(z(x), h(x)) = \bar{u} \qquad 0 \le x \le \bar{x}. \tag{2.25}$$

where $\bar{u}$ is the exogenously given utility level. Note that since the population of the city is a choice variable in an open city, the population constraint can be ignored.

The Lagrangian for this problem is

$$\begin{aligned}\Lambda = {} & F[\int_0^{\bar{x}} N(x)dx] - \int_0^{\bar{x}} [(z(x)+t(x))N(x) + R_a\theta(x)]dx \\ & + \int_0^{\bar{x}} \nu(x)[u(z(x), h(x) - \bar{u}]N(x)dx \\ & + \int_0^{\bar{x}} R(x)[\theta(x) - N(x)h(x)]dx\,.\end{aligned} \tag{2.26}$$

The first order conditions become, after simple manipulations,

$$u_h / u_z = R(x)\,, \qquad 0 \le x \le \bar{x}\,, \tag{2.27a}$$

$$F' = z(x) + R(x)h(x) + t(x)\,, \qquad 0 \le x \le \bar{x}\,. \tag{2.27b}$$

Considering $R(x)$ as land rent, we can observe that these optimality conditions coincide exactly with the market equilibrium conditions of the absentee-landlord case of the open city if workers earn wages equal to the value of marginal productivity of labor.

Thus the market equilibrium is the optimal solution in this case as well.

Notes

The theory of residential land use which is described in this chapter was first established by Alonso (1964) following the pioneering work of Wingo (1961). Many urban economists have extended Alonso's framework. Extensive empirical research has also been carried out. These efforts have culminated in Muth (1969) and Mills (1972a,b).

The indirect utility function approach adopted in this chapter was introduced into an urban residential land use model by Solow (1973). This approach has proved to be very useful in deriving qualitative results.

The single-center assumption was relaxed by Romanes (1976) and White (1976). In a two dimensional case, introduction of subcenters gives rise to complicated partial differential equations which are very difficult to analyze.

More than one income class was introduced by Beckman (1969) and Solow (1973) among others. Beckman considered the case of Pareto income distribution. Beckman's solution was not correct since, as pointed out by Montesano (1972), he ignored boundary conditions (among other things). Our treatment of different income classes is based on Solow's. Miyao (1975) analyzed the dynamic stability of boundaries between different income classes. Empirical research on spatial residential patterns with several income classes was carried out by Wheaton (1977).

Time costs of commuting were included in Alonso's original formulation, though later studies tend to ignore time costs by considering the pecuniary cost as a surrogate. As discussed in subsection 1.2, the inclusion of time costs tends to weaken the tendency of the richer households to live farther from the center since the rich's value of time is higher than the poor's, making commuting costs for the rich greater than for the poor.

Models with durable housing stock were analyzed by Fujita (1976a,b), and Anas (1976). Since dynamic aspects must be taken into account in this case, the analysis becomes much more complicated.

Definitions of closed and open cities were introduced by Wheaton (1974) in his comparative static analysis.

The Benthamite optimal city was first analyzed by Mirrlees (1972). He discovered that utility levels are not equal at the Benthamite optimum. Riley (1973), (1974) further analyzed this property using different social welfare functions. The product of individual utilities was used in Riley (1973) as the social welfare function, and a general class of concave and additive social welfare functions in Riley (1974). He derived a result parallel to ours: when land is a normal good and when there is no preference for location *per se*, individuals further out will receive greater utility levels at the optimum. Our illustration in Appendix I of the reason why unequal utility levels are obtained at the optimum is largely based on Arnott and Riley (1977) and Levhari, Oron and Pines (1978).

The Rawlsian case was considered by Dixit (1973). The method of maximizing the utility level under the constraint that the utility level be equal everywhere in the city was adopted by Oron, Pines and Sheshinski (1973).

REFERENCES

Alonso, W., (1964), *Location and Land Use,* (Harvard University Press, Cambridge, Massachusetts).

Anas, A., (1976), "Short-run Dynamics in the Urban Housing Market", in: G.J. Papageorgiou (ed.). *Mathematical Land Use Theory,* (Lexington Books, Lexington, Massachusetts)

Arnott, R. and J.G. Riley, (1977), "Asymmetrical Production Possibilities, the Social Gains from Inequality and the Optimum Town", *Scandinavian Journal of Economics* 79, 301-311.

Beckmann, M.J., (1969), "On the Distribution of Urban Rent and Residential Density", *Journal of Economic Theory* 1, 60-67.

Dixit, A., (1973), "The Optimum Factory Town", *The Bell Journal of Economics and Management Science* 4, 637-651.

Fujita, M., (1976), "Spatial Patterns of Urban Growth: Optimum and Market", *Journal of Urban Economics* 3, 209-241.

Fujita, M., (1976), "Toward a Dynamic Theory of Urban Land Use", *Papers of Regional Science Association* 37, 133-165.

Henderson, J.V., (1977), *Economic Theory and the Cities,* (Academic Press, New York).

Hicks, J.R., (1946*), Value and Capital,* (Clarendon Press, Oxford).

Levhari, D., Y. Oron, and D. Pines, (1978), "A Note on Unequal Treatment of Equals in an Urban Setting", *Journal of Urban Economics* 5, 278-284.

McGuire, M., (1974), "Group Segregation and Optimal Jurisdiction", *Journal of Political Economy* 82, 112-132.

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

Mills, E.S., (1972), *Urban Economics,* (Scott, Foresman and Company, Glenview, Illinois).

Mirrlees, J.A., (1972), "The Optimum Town", *Swedish Journal of Economics* 74, 114-135.

Miyao, T., (1975), "Dynamics and Comparative Statics in the Theory of Residential Location", *Journal of Economic Theory* 11, 133-146.

Montesano, A., (1972), "A Restatement of Beckmann's Model on the Distribution of Urban Rent and Residential Density", *Journal of Economic Theory* 4, 329-354.

Muth, R.F., (1969), *Cities and Housing,* (University of Chicago Press, Chicago, Illinois).

Oron, Y., D. Pines and E. Sheshinski, (1973), "Optimum vs. Equilibrium Land Use Pattern and Congestion Toll", *The Bell Journal of Economics and Management Science* 4, 619-636.

Riley, J.G. (1973), ""Gammavilie": An Optimal Town", *Journal of Economic Theory* 6, 471-482.

Riley, J.G., (1974), "Optimal Residential Density and Road Transportation", *Journal of Urban Economics* 1, 230-249.

Romanos, M., (1976), *Residential Spatial Structure,* (Lexington Books, Lexington, Massachusetts).

Solow, R.M., (1973), "On Equilibrium Models of Urban Location", in: M. Parkin (ed.). *Essays in Modern Economics* (Longman, London).

Wheaton, W., (1977), "Income and Urban Residence: An Analysis of Consumer Demand for Location", *American Economic Review* 67, 620-631.

White, M.J., (1976), "Firm Suburbanization and Urban Subcenters" *Journal of Urban Economics* 3, 323-343.

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

CHAPTER II

CITY FORMATION AND CITY SIZES

Complexity generally increases more rapidly than realism in model building. Although we understand many of the principles governing city formation, we do not yet have a model which includes any large part of our knowledge of real cities and remains simple enough to work with. We can, however, extend the basic model of Chapter I in several ways, and obtain interesting results.

In a sense, the open and closed cities of Chapter I hang in mid-air. We assumed either a given population or a given level of utility, without considering how that level came about. If we are interested in how these variables are determined, we need a general equilibrium model of an economy containing cities, not just a model of a single city. In this chapter we explicitly introduce a rural sector spread over on a featureless plain. The rural sector produces an agricultural good which is consumed by households in both the rural and urban sectors. Circular cities producing an urban good are sprinkled about on the plain.

In Chapter I we assumed that commuting is costly. The commuting costs and, in fact, transportation costs in general work against city formation. Concentration of production, for example, requires the transportation of products, workers, and material inputs. To obtain cities in our model, therefore, we must assume that the concentration of economic activities results in a technological advantage which at least exceeds the transportation costs incurred. Otherwise, production will take place where there are consumers, and the consumers, who find no advantage in working together, will spread out evenly to take advantage of all the available land.

Cities will arise in our model if we assume any or all of the following:

1. concentration of immobile factors
2. increasing returns to scale or indivisibility
3. externalities or public goods.

Cities arising from *concentrations of immobile factors* are relatively easily modeled, although we only mention them here. Given an immobile and concentrated factor, like a coal bed, industries which use the factor, such as mining, locate at that point. Industries such as steel, which use the primary product intensively, tend to locate nearby to save transportation costs. Others which are related and a retail sector follow for the same reason. The neoclassical model can describe such a city: there is convexity in production technology, and there are no externalities, and therefore the market mechanism can achieve an optimal allocation.[1] A concentration of immobile

[1] As discussed in Chapter I and Appendix I, there is a concealed nonconvexity in residential land use models, since a household can choose only one location. The nonconvexity, however, does not affect

factors, however, can produce only a relatively small city, and does not seem to be an important cause of modern cities.

We will first examine cities that arise from *economies of scale.* Economies of scale are prevalent in modern technology and result from such things as the division of labor and the indivisibility of such factor inputs as machinery and buildings. If the reduction in production costs due to scale economy is greater than the increase in commuting costs, a city will emerge. Such a city is basically a factory surrounded by the residential zone of its workers and may well be called a '*factory town*'.

Modern cities are, however, too complex to have resulted from simple economies of scale. Why should industries gather in a large city, where the commuting costs for each are greater than they would be in a single industry town? The answer is that industries find it profitable to gather together for a variety of reasons: communication costs and transportation costs of intermediate inputs can be saved; there is a larger pool of skilled labour to draw on, for example, and a more sophisticated infrastructure including transportation facilities. In order to capture these elements in a simple model, we assume a variant of *Marshallian externality.* We assume externalities among firms in a city, rather than among firms in an *industry*: all firms in a *city* are assumed to benefit from an increase in the population of the city. This assumption introduces the possibility of a city consisting of many firms by allowing increasing returns to scale which are internal to a city but external to the separate firms in a city.

We assume identical cities in both the increasing-returns-to-scale and the Marshallian-externality cases. This assumption allows us to obtain clear-cut results, but obviously fails to capture the complexity of the system of cities in the real world. In the last part of this chapter we discuss some possible extensions of the model, and the associated difficulties.

One of the major theoretical issues we try to analyze in this chapter is whether the decentralized market system can achieve the optimal allocation of cities, especially the optimal city size. If it cannot, we want to know how to correct the misallocation. It is well known in Welfare Economics that competitive equilibrium is not usually Pareto optimal in the presence of increasing returns to scale or externalities: there is almost always room to make somebody better off without making any others worse off. Although the result holds in our model, it is possible to describe an institutional arrangement that leads to optimal allocation.

The major difficulty in achieving an efficient allocation of an increasing-returns-to-scale industry is that the average cost always exceeds the marginal cost. Since an efficient allocation requires that the price be set equal to the marginal cost, the total revenue does not cover the total cost and the profit of a firm is negative. It is difficult to give such a firm a subsidy to cover the loss without destroying the incentive to minimize costs.

It turns out, however, that the loss of an urban producer equals the *aggregate*

---

the optimality of competitive equilibrium.

differential rent (the competitive urban rent minus the rural rent) of a city when the number of cities is optimal. This suggests that the optimal allocation could be achieved by a system of land developers. In each city an urban producer would lease all the land necessary for a city, including the residential area, from the rural landlords at the rural rent; sublease it to households at the competitive urban rent; and maximize differential rent plus profit. We will show that this arrangement does achieve the optimum allocation.

The optimal allocation of an economy with externalities requires *Pigouvian tax-subsidies*: agents that induce external costs or benefits for other agents must pay taxes or be given subsidies. In our model the population of a city gives external benefits to urban producers, so urban residents must be given subsidies. A relationship similar to that between the profit and the differential rent in the increasing returns to scale case holds with respect to the Pigouvian subsidy and the differential rent: at the optimum number of cities the aggregate Pigouvian subsidy equals the aggregate differential rent. This might seem to suggest that the optimal solution can be attained through the market if city governments return the differential rent to city residents as an equal social dividend. Unfortunately, this is not true. Though the optimal solution is indeed a market equilibrium under this institutional arrangement, city sizes greater than the optimum can also be equilibria.

1. The Model

The basic model of Chapter I becomes a simple general equilibrium framework when the *rural sector* is explicitly introduced. Consider a flat and fertile plain over which the rural sector is spread out. Circular cities are sprinkled about on the plain as in Figure 1. The plain is so large that the cities do not overlap.

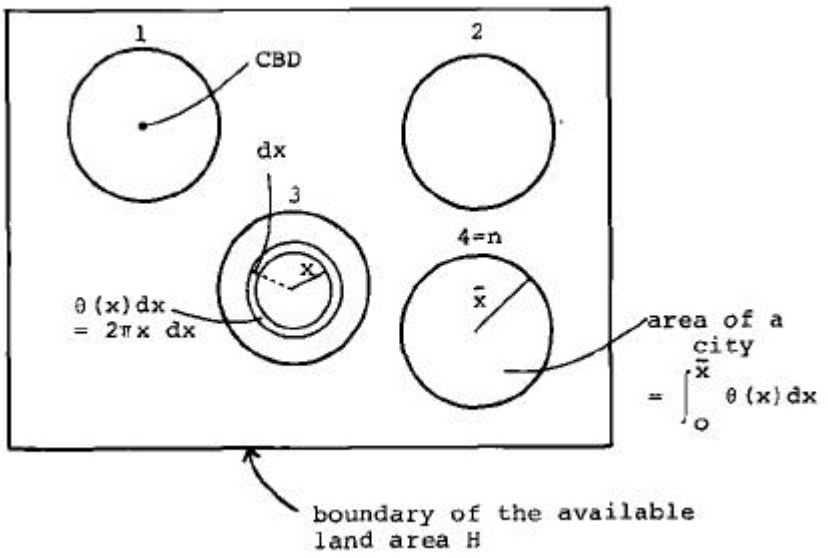

Figure 1. The Spatial Configuration of the Economy

We will imagine that each city consists of a business and production core surrounded by a residential zone. For the sake of simplicity, we will assume that

urban production takes no space and no materials, so that only labor enters the production function. As before, households are identical, and each household has a single worker who commutes to a job at the center. We use the words 'households' and 'workers' inter-changeably, and the number of households is treated as the population. Commuting involves transportation costs but we assume that both urban and rural goods can be moved costlessly.

The production function of an urban firm is

$$f(\ell, P_c), \tag{1.1}$$

where ℓ and P_c are respectively labour input to a firm and the population of the city. We assume that all firms are identical. If the number of firms is m, the total production of a city is

$$Y = mf(\ell, P_c), \tag{1.2}$$

where

$$m = \frac{P_c}{\ell}. \tag{1.3}$$

Until section 5 we will assume that $f_p(\ell, P_c) = 0$ so that firms gain no advantage from increased population. Since grouping firms will result in higher commuting costs, there will be only one firm in a city. In this case the aggregate urban production function, $F(P_c)$ and an individual firm's production function will be identical:

$$F(P_c) \equiv f(P_c, P_c). \tag{1.4}$$

The production function is differentiable, with a positive marginal product. In order to allow for the possibility of increasing returns to scale, we do *not* impose the condition that the production function be concave.

We assume that the rural sector produces some complete and useful product, say soy beans, and that it has constant returns to scale. The aggregate production function of the rural sector is

$$G(P_a, H_a)), \tag{1.5}$$

where P_a and H_a are respectively the aggregate labour and land inputs. We assume that the production function is concave, homogeneous of degree one, differentiable, and that it has positive marginal products.

Goods produced by rural and urban sectors are called the *rural* and the *urban goods* respectively. Both goods are consumed by households. All households have the same differentiable, quasi-concave utility function,

$$u(a, z, h), \tag{1.6}$$

where a, z, and h are respectively the rural good, the urban good, and land for housing and where all goods are assumed to have positive marginal utilities.

The number of cities is denoted by n. For simplicity, we assume that all cities are identical, and we use the same notation for all cities. Since all cities have the same technology, and all households are the same, the assumption can usually be justified when the number of cities is large enough.

All cities are circular, and the amount of land available for housing between x and $x + dx$ from the center is $\theta(x)dx = 2\pi x dx$. As the production does not require land, the residential zone stretches from 0 to $\bar{x}$. The consumption of land for housing is constrained by

$$h(x)N(x) = \theta(x), \qquad 0 \le x \le \bar{x}, \tag{1.7}$$

where $N(x)dx$ is the number of households which live between x and $x + dx$, and $h(x)$ is the lot size of a house at x.

The total available land for the economy, H, is divided between cities and rural areas. The rural sector uses land both for production and for housing the rural workers. The land constraint for the entire economy is

$$H = n\int_0^{\bar{x}} \theta(x)dx + P_a h + H_a, \tag{1.8}$$

where h denotes the consumption of housing by rural residents. Note that, if h appears without the argument x, it denotes consumption by a rural resident. This distinction in notation will be used consistently. Implicit in constraint (1.8) is the assumption that the total available land is large enough to preclude overlapping of city areas.

Transportation requires many different inputs, but for the sake of simplicity, we assume that only the rural good is consumed in commuting. We continue to assume that goods, urban and rural, can be transported costlessly. The market clearing conditions for rural and urban goods are respectively

$$G(P_a, H_a) = n\int_0^{\bar{x}} [a(x) + t(x)]N(x)dx + P_a a, \tag{1.9}$$

$$nF(P_c) = n\int_0^{\bar{x}} z(x)N(x)dx + P_a z, \tag{1.10}$$

where $z(x)$ and $h(x)$ are the urban consumptions of urban and rural goods, and z and a are the rural consumptions. Commuting costs, $t(x)$, for a city resident at x satisfy

$$t(0) = 0\,. \tag{1.11}$$

The labour force in a city is assumed to equal the number of households living in the city:

$$P_c = \int_0^{\bar{x}} N(x)dx\,. \tag{1.12}$$

The population, *P,* of the whole economy, which is assumed to be given, is divided into urban and rural sectors:

$$P = nP_c + P_c\,. \tag{1.13}$$

2 A Fixed Number of Cities

We first derive the optimal solution under the assumption that the number of cities is exogenously given. As mentioned above, in this and the next two sections we assume that the marginal effect on production of increasing the population of a city is zero: $f_P(\ell, P_c) = 0$. Firms gain no advantage having other firms in the city, therefore, and each city has only one firm. The cities which we consider are based on economies of scale which are internal to the firm: $f_\ell > f/\ell$. Since $P_c = \ell$ for a single firm city, our notation can be simplified by using the aggregate production function, $F(P_c)$, and assuming

$$F'(P_c) > F(P_c)/P_c\,. \tag{2.1}$$

The utility level is maximized subject to the constraints (1.7)-(1.10), (1.12), (1.13), and the equal-utility constraints,

$$u(a(x), z(x), h(x)) = u\,,\ \ 0 \le x \le \bar{x}\,, \tag{2.2}$$

$$u(a, z, h) = u\,, \tag{2.3}$$

which require that all households receive the same utility level.

The Lagrangian for this problem is

$$
\begin{aligned}
\Lambda = u &+ n\int_0^{\bar{x}} \nu(x)[u(a(x), z(x), h(x)) - u]N(x)dx + \nu_a[u(a, z, h) - u]P_a \\
&+ \delta_c[nF(\int_0^{\bar{x}} N(x)dx) - n\int_0^{\bar{x}} z(x)N(x)dx - P_a z] \\
&+ \delta_a\left\{G(P_a, H_a) - n\int_0^{\bar{x}} [a(x) + t(x)]N(x)dx - P_a a\right\} \\
&+ n\int_0^{\bar{x}} \mu(x)[\theta(x) - h(x)N(x)]dx \\
&+ \mu[H - n\int_0^{\bar{x}} \theta(x)dx - P_a h - H_a] \\
&+ \gamma\left[P - n\int_0^{\bar{x}} N(x)dx - P_a\right]
\end{aligned}
\tag{2.4}
$$

where control variables are $a(x), z(x), h(x)$, and $N(x)$; control parameters are a, z, h, u, P_a, H_a, and $\bar{x}$; and $\nu(x)$, ν_a, δ_c, δ_a, $\mu(x)$, μ, and γ are respectively Lagrange multipliers for (2.2), (2.3), (1.10), (1.9), (1.7), (1.8), and (1.13). Note that P_c is eliminated by substituting (1.12) into (1.10) and (1.13). The Lagrange multipliers have basically the same interpretation as in Chapter I: $\nu(x)$ and ν_a are weights attached to the utilities of different households to obtain equal utility levels; δ_c and δ_a are respectively shadow prices of the urban and rural goods; $\mu(x)$ is the shadow rent of land at distance x from the center of a city and μ the shadow rent of the rural land; and γ is the shadow 'price' of a household.

The Lagrange multipliers express shadow prices in utility terms. It is convenient to transform shadow prices into pecuniary terms. Taking the rural good as a numeraire, we define $p = \delta_c/\delta_a$, $R(x) = \mu(x)/\delta_a$, $R_a = \mu/\delta_a$ and $s = -\gamma/\delta_a$. p is the shadow price of the urban good, $R(x)$ the shadow rent at x, R_a the shadow rent of the rural land, and a the marginal social cost—the negative of the shadow price—of a household.

First order conditions are immediately obtained by applying the result in the appendix on optimum control theory. They become, after simple rearrangements,

$$
\frac{u_z(a(x), z(x), h(x))}{u_a(a(x), z(x), h(x))} = \frac{u_z(a, z, h)}{u_a(a, z, h)} = p, \qquad 0 \le x \le \bar{x}, \tag{2.5a}
$$

$$
\frac{u_h(a(x), z(x), h(x))}{u_a(a(x), z(x), h(x))} = R(x), \qquad 0 \le x \le \bar{x}, \tag{2.5b}
$$

$$
\frac{u_h(a, z, h)}{u_a(a, z, h)} = R_a, \tag{2.5c}
$$

$$
G_H = R_a, \tag{2.5d}
$$

$$G_P + s = a + pz + R_a h\,, \tag{2.5e}$$

$$pF' + s = a(x) + pz(x) + R(x)h(x) + t(x)\,, \qquad 0 \le x \le \bar{x}\,, \tag{2.5f}$$

$$R(\bar{x}) = R_a\,.$$

(2.5a)-(2.5c) are the usual conditions equating marginal rates of substitution to (shadow) prices. Note that (2.5a) implies that all households have the same marginal rates of substitution between the urban good and the rural good. This is an immediate consequence of our assumption that it costs nothing to transport either good.

(2.5d) states that the value of marginal productivity of land is equal to the shadow rent of land. From (2.5c), (2.5d) and (2.5g), rural households, rural producers and urban residents at the edge of a city all face the same shadow rent. This condition implies that shadow rent varies continuously over space.

The social cost, *s*, of a household must be the value of resources it consumes minus the value of its marginal product:

$$s = a + pz + R_a h - G_p$$

$$s = a(x) + pz(x) + R(x)h(x) + t(x) - pF'\,.$$

These yield respectively (2.5e) and (2.5f).

If workers are paid the value of their marginal products, and if all prices equal shadow prices, a household must be given a subsidy which is equal to s in order to satisfy the budget constraint. Because of the resource constraints, (1.9) and (1.10), however, the sum of the subsidies must equal the total surplus in the economy, which is the sum of the total rent and the total profit:

$$s = \frac{l}{P}\left\{n\int_0^{\bar{x}} R(x)\theta(x)dx + R_a\left[P_a h + H_a\right] + np\left[F + P_c F'\right]\right\}. \tag{2.6}$$

If this optimal solution is decentralized using the usual price mechanism, an urban producer might incur a loss at the optimum, since we allowed increasing returns to scale. In such a case the government must give the producer a subsidy equal to the loss. If the subsidy does not weaken a firm's incentive to minimize costs, the price mechanism attains the optimal allocation. Unfortunately, administering such a subsidy requires a prohibitive amount of information.

These problems may not arise if the producer can act as land developer, collecting the residential land rent. We consider this institutional framework in section 4.

Of course, if the urban sector has decreasing or constant returns to scale, the urban sector earns a positive or zero profit, and hence a subsidy is not necessary. In such cases, however, there is no reason to have cities, since by reducing city size transportation costs can always be reduced without raising production costs. As will

be shown in the next section, the optimal solution with nonincreasing returns to scale requires that the urban sector spread uniformly over space.

3. A Variable Number of Cities

When we treat the number of cities as an endogenous variable, it is convenient to assume that the number is so large that it can be safely approximated by a continuous variable. Taking a derivative of the Lagrangian (2.4) with respect to n and substituting other first order conditions, we obtain

$$pF(P_c) - [pP_cF'(P_c) - \int_0^{\bar{x}} R(x)\theta(x)dx] - R_a \int_0^{\bar{x}} \theta(x)dx = 0. \tag{3.1}$$

The number of cities should be increased up to the point where an additional city has zero social net value. The net value of an additional city is the value of the gross product of the city minus the costs of producing it. The value of the gross product is pF. The cost of producing it is the cost of supporting workers. Workers consume the rural good, the urban good, and land: and they pay commuting costs. The total social value of their consumption is

$$\int_0^{\bar{x}} [a(x) + pz(x) + t(x)]N(x)dx + R_a \int_0^{\bar{x}} \theta(x)dx.,$$

where land must be evaluated at the opportunity cost, R_a, instead of the urban rent. This is not yet the social cost of supporting workers of an additional city. Since the society incurs the social cost of a household, *s*, regardless of whether a household lives in the city or not, sP_c must be subtracted from the costs. The net value of an additional city is then

$$pF(P_c) - \int_0^{\bar{x}} [a(x) + pz(x) + t(x)]N(x)dx - R_a \int_0^{\bar{x}} \theta(x)dx + sP_c = 0.$$

Substitution of (2.5f) into this equation yields (3.1).

If the number of cities is optimal, the population of a city is such that it minimizes the value of per capita consumption of resources, or the average cost of maintaining the utility level. Otherwise, there is some other population level which achieves the same utility level with a lower per capita consumption of resources, and the value of resources used by the entire urban sector can be reduced by changing the population size of all cities while changing the number of cities accordingly to keep the population of the entire urban sector unchanged. The resources saved could be used to raise the utility level, which proves that the allocation cannot be optimal. Note that consumption of resources in this argument does not need to be modified by subtracting the social cost of a household.

This observation facilitates another interesting interpretation of the optimality

condition.[2] The net cost of maintaining the utility level is equal to the total city consumption, plus the opportunity cost of land, minus total production. Expressed in terms of city population, the net cost is simply a total cost function,

$$TC(P_c) = \int_0^{\bar{x}} [a(x) + pz(x) + t(x)]N(x)dx + R_a \int_0^{\bar{x}} \theta(x)dx - pF(P_c) \quad .$$

The per capita, or average, cost is

$$AC(P_c) = TC(P_c)/P_c \; .$$

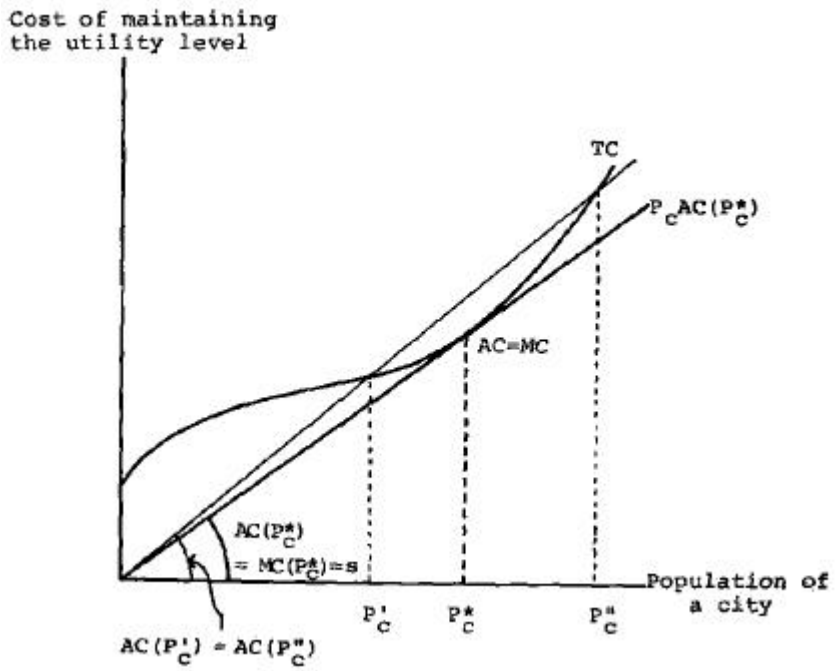

Figure 2. Optimum City Size

As illustrated in Figure 2, the average cost is minimized when it equals the marginal cost. The marginal cost is the cost of adding a household to the city. Since at the optimum the cost of adding a household must equal everywhere in the city, we may consider the addition at any radius. Addition at the edge of the city is easiest because there is no ambiguity about whether the rural rent or the urban rent expresses the value of land. The cost of adding a household at $\bar{x}$ is the value of consumption minus the marginal product:

$$MC(P_c) = a(\bar{x}) + pz(\bar{x}) + t(\bar{x}) + R_a h(\bar{x}) - pF'(P_c) \quad ,$$

which, from (2.5f), equals the social cost of a household, s, and also equals the value of

[2] This interpretation was suggested by Arnott. Similar interpretation is published in Arnott (1979).

consumption minus the marginal product at any other radius when consumption of land is evaluated at the urban shadow rent:

$$MC(P_c) = a(x) + pz(x) + t(x) + R(x)h(x) - pF'(P_c) \quad , \qquad 0 \le x \le \bar{x} \quad .$$

Multiplying this equation by $N(x)$ and integrating it from 0 to $\bar{x}$ yields

$$P_c MC(P_c) = \int_0^{\bar{x}} [a(x) + pz(x) + t(x)]N(x)dx + \int_0^{\bar{x}} R(x)\theta(x)dx - pP_c F'(P_c) \quad .$$

At the optimum the average cost equals the marginal cost, which implies equation (3.1) above.

We can rewrite (3.1) as

$$-p[F(P_c) - P_c F'(P_c)] = \int_0^{\bar{x}} [R(x) - R_a]\theta(x)dx \quad . \tag{3.1'}$$

If factor prices are equal to the values of marginal productivities, the left can be interpreted as the operating loss of an urban firm. Then (3.1') states that, *in a single firm city the firm's operating loss is equal to the aggregate differential rent* (the urban rent minus the rural rent) *if the number of cities is optimal.* Using this equation, we can rewrite (2.6) as

$$s = HR_a / P, \tag{3.2}$$

which says that the social cost of a household equals the per capita rural rent.

With constant or decreasing returns to scale, firms earn nonnegative profits. By (3.1') the aggregate differential rent would be nonpositive at the optimum, implying that cities have simply vanished. This is quite reasonable since smaller cities have the advantage of lower transportation costs with no disadvantage on the production side.

If the urban sector has increasing returns to scale, bigger cities have the advantage of lower average production costs. The optimum city size or the optimum number of cities is determined so as to balance the transportation costs and the benefit from increasing returns to scale. (3.1) shows that this balance is attained when the loss of the urban sector equals the aggregate differential rent.

By solving the problem of Section 2, the utility level can be obtained as a function, u(n), of the number of cities. The second order condition requires that

$$\frac{d^2u(n)}{dn^2} \le 0 .$$

By the Envelope Theorem,[3] $du(n)/dn$ is equal to the partial derivative of the Lagrangian (2.4) with respect to n.

$$\frac{du(n)}{dn} = \frac{\partial \Lambda}{\partial n} = \delta_a \{ p[F - P_c F'] + \int_0^{\bar{x}} [R(x) - R_a] \theta(x) dx \} \tag{3.3}$$

Hence, using the fact that the term with $d\delta_a/dn$ vanishes by (3.1), the second derivative is

$$\frac{d^2 u(n)}{dn^2} = \delta_a \frac{d}{dn} \{ p[F - P_c F'] + \int_0^{\bar{x}} [R(x) - R_a] \theta(x) dx \}, \tag{3.4}$$

which must be nonpositive at a maximum.

The net benefit from an additional city is the sum of differential rent and profit. The optimum is attained at the point where the net benefit is zero. In order to have maximum rather than minimum, however, the net benefit must be decreasing at the optimum, and the sum of the aggregate differential rent and the profit of an urban producer must be a decreasing function of the number of cities.

The second order condition is usually satisfied if the degree of increasing returns to scale declines as a city becomes larger, because the aggregate differential rent is usually larger in a larger city.

An important implication of this second order condition is that, if the aggregate differential rent exceeds the loss of the urban sector, there are too few cities. Since cities tend to be larger when there are fewer cities, city size is likely to be too big in this case. Notice, however, that this result is obtained under the condition that all variables other than the number of cities are optimally chosen. If there is some distortion like monopsony pricing in the labour market, this condition is not satisfied, and the difference between the differential rent and a firm's loss does not serve as a signal of whether or not the city size is too big. In the next section a paradoxical kind of monopsony pricing will be shown to exist at the market equilibrium of our model.

4. Market City Sizes

As noted in section 2, a firm must be given an appropriate subsidy to achieve the optimal solution in a decentralized market system. The result in section 3 shows that if the number of cities is optimal, the subsidy equals the total differential rent of a city. Thus the optimal solution can be decentralized by giving a firm a subsidy equal to the differential rent and distributing among all rural and urban households the rest of the rent (which equals the rural rent for the entire land, $R_a H$) as an equal lump-sum social dividend.

---

[3] See Appendix III for the explanation of the Envelope Theorem.

There is, however, a way to achieve the optimal solution that does not require as much knowledge and action on the part of the government. It turns out that *the optimal solution can be achieved by allowing firms to lease the urban land including the residential area.* The entire available land is owned collectively by all households in the economy. Rural producers and rural households rent the land and pay the rural rent. An urban firm, acting as a developer of a whole city, also rents urban land at the rural rent, but subleases it to city residents at the competitive market rent. Then the firm maximizes the sum of profit from production, which is usually negative, and the net rent. Firms, like land, are owned collectively by all households, and profits are distributed equally among all households. We show that such a system of urban-producers/ city-developers attains the optimal allocation, providing, of course, that the firms are perfectly competitive.

The number of cities (and hence the number of firms) is assumed to be so large that a firm acts as a price and utility taker: since there are no transportation costs for the urban good, firms directly compete with each other in the product market, and a single firm cannot significantly affect the price of the urban good. Under the assumption of perfect mobility, households move to the city where they can obtain the highest level of utility. Faced with freely mobile households, a firm must make sure that its employees obtain at least the same utility level as they would in any other city. This leads to utility taking behaviour as the number of firms becomes large. Notice, however, that a firm does not take the wage rate as given. Households decide to migrate on the basis of the utility level and not the wage rate. As long as the utility level is not lower than at any other place, the wage rate can be freely chosen.

A firm maximizes the profit on the entire development which is the revenue from the sales of its product, plus the land rent, minus the total wage bill, minus the total payment of the rural rent:

$$pF(P_c) + \int_0^{\bar{x}} R(x)\theta(x)dx - wP_c - R_a \int_0^{\bar{x}} \theta(x)dx, \qquad (4.1)$$

where w is the wage rate.

Four variables, w, P_c, $\bar{x}$, and $R(x)$, are involved in this maximization problem, but the firm faces the constraints imposed by competition with other firms. The maximum rent that households can pay, if they are to achieve the given utility level, is a function of their wage. We can, therefore, reduce the problem to that of maximizing (4.1) with respect to the wage. First, using the indirect utility function of households, we express $R(x)$ as a function of the wage.

Since all firms and the entire land are collectively owned by all households in the economy, households obtain equal shares of profits of firms and the revenue from the rural rent paid by both the rural and urban sectors. Then the budget constraint is

$$w + s = a(x) + pz(x) + R(x)h(x) + t(x) \quad , \qquad 0 \le x \le \bar{x}, \qquad (4.2)$$

where s is the share of the rent and profit, and satisfies (2.6). The following *indirect*

utility function is obtained as a result of utility maximization under the budget constraint:

$$v(I(x),\, p, R(x))\,, \tag{4.3}$$

where $I(x)$ is the net income:

$$I(x) \equiv w + s - t(x)\,. \tag{4.4}$$

Since the utility level, *u*, is taken as given, a firm maximizes profit under the constraint:

$$v(I(x),\, p, R(x)) = u\,. \tag{4.5}$$

This constraint enables us to express $R(x)$ as the *bid rent function,*

$$R(x) = R(I(x),\, p, u)\,. \tag{4.6}$$

As in (I.1.14), we have

$$R_I\,(I(x),\, p, u) = 1/h(x)\,, \tag{4.7}$$

where R_I is the partial derivative of the bid rent function with respect to the net income $I(x)$.

Substituting (4.7) and (4.4) into (4.1), we see that the firm's problem is to maximize

$$pF(P_c) - wP_c - R_a \int_0^{\bar{x}} \theta\,(x)dx + \int_0^{\bar{x}} R(w + s - t(x),\, p, u)\theta\,(x)dx \tag{4.8}$$

subject to the constraints

$$P_c = \int_0^{\bar{x}} R_I\,(w + s - t(x),\, p, u)\theta\,(x)dx\,, \tag{4.9}$$

$$R(w + s - t(\bar{x}),\, p, u) = R_a\,, \tag{4.10}$$

where *p*, *u*, and *s* are taken as given since the firm is small.

Now the population of the city can also be written as a function of the wage. Although the price of the product and the utility level are taken as fixed, the wage rate affects the supply of labour, because households will move to achieve the given utility level. From (4.10), $\bar{x}$ can be expressed as a function of *w*, and (4.9) becomes the following labour supply function,

$$P_c(w) = \int_0^{\bar{x}(w)} R_I\,(w + s - t(x),\, p, u)\theta\,(x)dx\,, \tag{4.11}$$

where s is given by (2.6).

Using this labour supply function we can demonstrate that firms have a kind of monopsony power despite the fact that they are competitive in the usual sense. The slope of the

supply curve is

$$P_c'(w) = \int_0^{\bar{x}} R_{II}\theta(x)dx + R_I(I(\bar{x}), p, u)\theta(\bar{x})\bar{x}'(w)$$

$$= \int_0^{\bar{x}} R_{II}\theta(x)dx + N(\bar{x})/t'(\bar{x}), \tag{4.12}$$

where the second equality is obtained from (4.7) and

$$\bar{x}'(w) = 1/t'(\bar{x}). \tag{4.13}$$

Since R_I satisfies

$$R_I(I(x), p, u) = 1/h[R(I(x), p, u)p, u], \tag{4.14}$$

where $h[\ ]$ is a compensated demand function for land, we have

$$R_{II}(I(x), p, u) = -h_R R_I / h^2 = -h_R / h^3 > 0. \tag{4.15}$$

Hence, (4.12) becomes

$$P_c'(w) = \int_0^{\bar{x}} -(h_R / h^2)N(x)dx + N(\bar{x})/t'(\bar{x}) > 0. \tag{4.16}$$

Thus *the supply curve of labour is upward sloping, and firms have apparent monopsony power in the labour market.*

Using the labour supply function (4.11), we can also reduce the problem (4.8)-(4.10) to maximization of

$$pF(P_c(w)) - wP_c(w) - R_a\int_0^{\bar{x}(w)}\theta(x)dx + \int_0^{\bar{x}(w)} R(w+s-t(x), p, u)\theta(x)dx \tag{4.17}$$

with respect to w. The first order condition is

$$(pF' - w)P_c'(w) - P_c(w) + \int_0^{\bar{x}} R_I\theta(x)dx + [R(w+s-t(\bar{x}), p, u) - R_a]\theta(\bar{x})\bar{x}'(w) = 0 \tag{4.18}$$

From (4.10) and (4.11), this becomes

$$pF'(P_c) = w. \tag{4.19}$$

Thus, even though a firm faces an upward-sloping supply curve in the labour market, it behaves like a price taker and sets the wage rate at the level where the value of marginal product equals the wage.

The reason is that when the utility level is fixed, the decrease in income of workers is fully reflected in a decrease in expenditures on land. Thus the increase in profit caused by lowering the wage is completely offset by the decrease in land rent, and the firm behaves as if there were no monopsony gain.

It now follows that all the first order conditions for the optimum are satisfied in market equilibrium: (2.5f) is obtained from (4.2) and (4.19); (2.5g) is equivalent to (4.10); (2.5a)-(2.5c) are the results of utility maximization of households; (2.5d) and (2.5e) result from profit maximization in the rural sector; and free entry insures equation (3.1), which states that the maximized profit (including differential rent) is zero in equilibrium. Therefore, if the first order conditions are sufficient to characterize the optimal solution, *the market equilibrium is optimal under the institutional framework in which a firm can act as the developer of an entire city.*

When firms cannot act as developers, however, a market equilibrium differs from the optimal allocation. A firm maximizes

$$pF(P_c) - wP_c, \tag{4.20}$$

taking the price of the product and the utility level of the workers as given. Households receive equal shares of profits of firms and the total land rent including the urban rent, and s is given by (2.6). In this case, too, a firm has monopsony power in the labour market in the sense that it faces an upward sloping supply curve. The labour supply function is given by (4.11). A firm takes s as given in this case, as in the last.

The first order condition for profit maximization is

$$pF'(P_c) = w + P_c / P_c'(w). \tag{4.21}$$

Free entry insures that the maximized profit is zero:

$$pF(P_c) - wP_c = 0. \tag{4.22}$$

Multiplying (4.21) by P_c and subtracting it from (4.22), we obtain

$$p\left[F(P_c) - P_c F'(P_c)\right] = -P_c^2 / P_c'(w) < 0, \tag{4.23}$$

which implies that market equilibrium occurs when the firm operates in the region of increasing returns to scale. The profit is, however, zero, since by (4.21) a firm exploits monopsony power, and pays a wage rate lower than the value of the marginal product of labour.

From (3.1), the optimal city size also occurs in the region of increasing returns to

scale. Since market equilibrium and optimal city sizes are in the same range of the production function, observation about returns to scale cannot be used to determine which is larger in general. It depends on the amount of monopsony power and the size of the aggregate differential rent. In principle there is no reason why the equilibrium city size should coincide with optimum city size, and a city in which firms maximize profits but do not act as land developers has zero probability of achieving an exactly optimum city size.

5. The Marshallian Externality Case

In the previous section scale economies were internal to the firm. We now assume scale economy internal to a city but external to a firm. Cities form because firms are more productive if they can draw on a larger population. To capture this effect, we repeat the analysis of the previous sections with one change. The production function of a firm is still

$$f(\ell, P_c) \tag{1.1}$$

but now we allow an increase in population to increase the firm's productivity:

$$f_P > 0\,. \tag{5.1}$$

This version of the *Marshallian externality* can result in multi-firm cities, since the presence of additional firms is now an advantage. If m is the number of firms, the total product of a city is

$$Y = mf(\ell, P_c) \tag{1.2}$$

where

$$m = \frac{P_c}{\ell}\,. \tag{1.3}$$

We assume increasing average returns to labour when the firm is small, with a gradual shift to decreasing returns as ℓ increases.

As in the previous case, we first consider the case of a *L*, fixed number of cities. The optimization problem can be solved in the same way as before, if (1.2) and (1.3) are substituted into the proper places. The first order condition (2.5f) is replaced by two conditions:

$$f = \ell f_\ell\,, \tag{5.2}$$

$$pf_\ell + s_c + s = a(x) + pz(x) + R(x)h(x) + t(x)\,, \quad 0 \le x \le \bar{x}\,, \tag{5.3}$$

where

$$s_c = pmf_P \tag{5.4}$$

(5.2) shows that the marginal product of labour equals the average product, and implies that firms operate under constant returns to scale at the optimum. The optimum is attained in long-run equilibrium at which all firms operate at the bottom of the average cost curve since constant returns to scale hold when the average cost curve is flat.

From (5.3), the total value of household consumption equals the sum of the three terms on the left: the value of marginal productivity of labour; the value of the marginal external economy, s_c, which an urban resident gives to the urban production sector; and the social dividend, s, equal for all households in both cities and the rural area. Therefore, a household must be given the Pigouvian subsidy, s_c, in addition to wages and the social dividend, *s*. A city resident gives external benefits to urban producers, and should be given a subsidy equal to the value of his marginal contribution to urban production.

Taking the number of cities as a variable now, the optimality condition becomes, after simple rearrangements,

$$s_c P_c = \int_0^{\bar{x}} [R(x) - R_a]\theta(x)dx . \tag{5.5}$$

Thus, at the optimum the total Pigouvian subsidy equals the total differential rent in a city. As before, the equal lump-sum social dividend s is given by (3.2).[4]

The optimal allocation is a market equilibrium in the following institutional setting. All land is equally and collectively owned by all households in the economy. Residents in a city form a cooperative, or a city government, which rents all the land for the city at the rural rent. Each household, in turn, rents land for housing from the city government, and pays the market-determined rent. Since the urban residential rent is higher than the rural rent, the city government has a surplus revenue. The surplus is returned to city residents as an equal subsidy. It will be shown later that the optimal

---

[4] It is easy to show that (5.5) is equivalent to (3.1). From (5.2), we can express labour input to a firm as a function, $\ell(P_c)$, of the population of a city. The aggregate production function can then be written as

$$F(P_c) = \frac{P_c}{\ell(P_c)} f(\ell(P_c), P_c) .$$

Differentiation of the aggregate production function yields

$$F'(P_c) = \frac{P_c}{\ell} f_P(\ell, P_c) + \frac{1}{\ell} f(\ell, P_c) + \left[\frac{P_c}{\ell} f_\ell(\ell, P_c) - \frac{P_c}{\ell^2} f(\ell, P_c) \right] \ell'(P_c)$$
$$= \frac{P_c}{\ell} f_P + \frac{1}{\ell} f$$

where the second equality is obtained from(5.2). Noting

(1.3) and (5.4), we finally obtain

$$s_c P_c = -p[F(P_c) - P_c F'(P_c)]$$

which shows that (5.5) is equivalent to (3.1).

solution is a market equilibrium under this institutional arrangement. Unfortunately, however, the optimal allocation is not a *unique* market equilibrium. A wide range of city sizes greater than the optimum can also be equilibria, and there is no reason to believe that the optimum is likely to be attained.

This point can be illustrated in the following way. If we specify the number of cities, a market equilibrium is obtained by substituting

$$s_c = \frac{1}{P_c}\int_0^{\bar{x}} [R(x) - R_a]\theta(x)dx$$

and (3.2) into the first order conditions (2.5a)-(2.5e), (2.5g), (5.2) and (5.3). Since the resulting population size is normally the same for all cities, we can consider the equilibrium utility level as a function, $u^*(P_c)$, of the population of a city. For simplicity, the function is assumed to be single-peaked as in Figure 3.

Clearly, city sizes less than P_c^*, where the equilibrium utility level attains its maximum, cannot be equilibria. If a household moves to another city, the utility level will rise in the receiving city, and fall in the city which has lost population. Therefore, a household has an incentive to move to another city. The receiving city would continue to grow at least until P_c^* was reached. The losing city would eventually disappear.

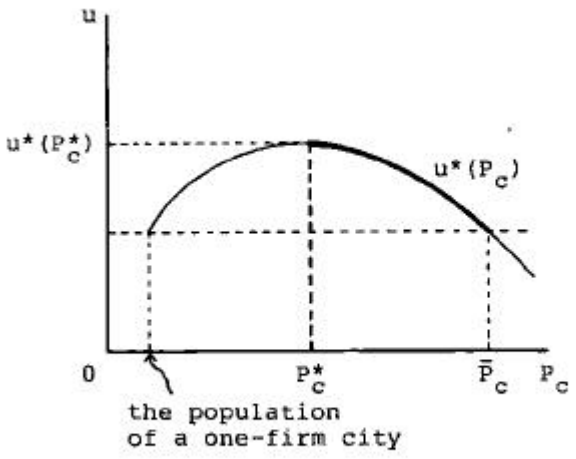

Figure 3. Market City Sizes

City sizes greater than P_c^*, however, can be an equilibrium. Households do not have an incentive to move to another city at a city size greater than P_c^* since an increase in the population of the receiving city would lower the utility there. For the same reason they do not have an incentive to move to the rural area, either. The only

way, therefore, to reduce the city size is to create a new city. If we do not allow for coalition or entrepreneurship to form a large new city, all new cities must start from one firm. In this case, a new city will not be formed unless the size of existing cities exceeds P_c^* by so much that the general utility level falls to that of a one-firm city. Therefore, the city sizes between P_c^* and $\overline{P}_c$ tend to remain the same.

It is easy to see that the city size, P_c^*, which maximizes the utility level among market equilibria coincides with that of the direct optimum. Though the direct optimum does not have, among its constraints, the conditions for a market equilibrium, the first order conditions for the direct optimum include all the conditions required for a market equilibrium. The two problems, therefore, must have the same optimum.

Thus *the minimum market city size coincides with the optimal city size when there is a Marshallian externality, and there is a strong tendency for market city sizes to become too large.* This result suggests that government intervention is necessary to achieve the optimum city size

Whether intervention is required to not, the actual situation may be less serious than the model suggests. Historical development has provided us with a hierarchy of cities rather than a single type. Cities produce different sets of commodities, and bigger cities produce more commodities than smaller ones. A new city at a certain level of the hierarchy can be created by adding firms producing new commodities to an existing city at a lower level of the hierarchy. This does not require a very large population shift.

The above special institutional arrangement which allows cities to collect land rent and distribute the revenue among city residents is not usually possible in a private ownership economy. In a private ownership economy, migrational decisions are not affected by the land rent households can *earn*, though they are certainly affected by the level of the rent they must pay. One reason is that households may be able to invest in houses in cities where they do not live. Another is that even if all houses are owner-occupied, households must pay the discounted value of future rent when they move to a city. The benefit of future high rent, which might attract households to a city, is thereby neutralized by the purchase cost of the house. Thus the usual private ownership economy is closer to the case of $s_c = 0$ and the city size which maximizes the utility level among market equilibria is different from the direct optimum.

It is not clear in a general case whether this city size is bigger or smaller than the optimum city size. If there is no rural sector, it is obvious that this city size coincides with the optimum. Divergence from the optimum is caused by the fact that the absence of the Pigouvian subsidy distorts the allocation of households between the urban and the rural sectors. In the real world it seems likely that the population in the urban sector is too small, because the incentive to live in cities is weaker due to the lack of the Pigouvian subsidy. However, the problem is more subtle than it appears, since it involves determining the number of cities. It is not quite clear how the number of cities is affected by the distortion.

We have assumed that it is the population of a city that generates an external economy. Obviously, this is not the only formulation. For example, we can assume that the total product of a city induces the externality, as in Henderson (1974). In that case, the Pigouvian subsidy should be given to firms as excise subsidy on their products. With this change, the above analysis can be applied and the same conclusions are obtained.

6. Differences in City Sizes

So far we have considered only cities which have the same allocation, both at the optimum and in market equilibrium. This is clearly unrealistic. Relaxing the simplifying assumptions of previous sections, we can obtain differences among cities.

First, production functions may differ among cities because of differences in climate, factor endowment, and so on, or simply because technology does not diffuse instantaneously. Since cities with technological advantage tend to attract more households than others, city sizes vary.

This extension turns out to be fairly simple. In the case of increasing returns to scale, internal to a firm, the only change is that we must distinguish cities notationally since they in general have different allocations. The first order conditions (2.5a), (2.5b), (2.5f) and (2.5g) hold in all cities. In particular, (2.5f) and (2.5g) for the i-th city must read

$$pF_P^i(P_c^i, H_c^i) + s = a^i(x) + pz^i(x) + R^i(x)h^i(x) + t(x) \tag{6.1}$$

$$R^i(\bar{x}^i) = R_a\,. \tag{6.2}$$

Combining these equations, we have

$$pF_P^i + s - t(\bar{x}^i) = a^i(\bar{x}^i) + pz^i(\bar{x}^i) + R_a h^i(\bar{x}^i)\,. \tag{6.3}$$

Since all households must receive the same utility, the right side is equal for all cities. Hence, the value of the marginal product of labour minus the commuting costs at the edge of a city is the same for all cities:

$$pF_P^i - t(\bar{x}^i) = pF_P^j - t(\bar{x}^j) \qquad \text{for any } i, j. \tag{6.4}$$

When the number of firms is optimized, the *marginal* firm will obtain zero profit (including the aggregate differential rent) and other firms will earn positive profits.[5] In exactly the same way as in the case of identical cities, it can be shown that, if firms

[5] Here, it is implicitly assumed that there is no competitive bidding for the right to build a plant in a specific city. This is the reason why a firm located in an advantageous city earns excess profit. The profit is caused by the Presence of some unpriced factors such as good climate, clean water, etc. If these factors are competitively priced, all firms earn zero profit. Even if there is no market for these factors, competitive bidding for the site of a plant drives down the profit to zero and the rent is captured by the owner of the site.

act as land developers, the market equilibrium attains the optimal allocation.

We can analyze the Marshallian externality case in a similar way. It is easy to see that the value of the marginal product of labour, plus the Pigouvian tax, minus the commuting costs at the edge of a city is equal for all cities;

$$pf_{\ell}^{\,i} + s_c^i - t(\bar{x}^i) = pf_{\ell}^{\,j} + s_c^j - t(\bar{x}^j) \qquad \text{for any } i, j \,.$$

The condition for the optimum number of cities is that the aggregate Pigouvian subsidy equals the aggregate differential rent in the *marginal* city. However, the equality does not hold in inframarginal cities. This causes a difference from the case of identical cities. If all cities are identical, the aggregate Pigouvian subsidy must equal the aggregate differential rent in all cities. This is the reason why we obtained the result that, if the differential rent is returned to city residents as an equal subsidy, the optimal allocation is one of market equilibria. If cities are not identical, the result does not hold, since the optimum Pigouvian subsidy is not equal to the average differential rent in inframarginal cities. Therefore, even the best allocation among market equilibria does not coincide with the optimal allocation.

A second class of differences which can give rise to differing cities includes all the ways that household tastes and skills may vary. An extended analysis, unfortunately, is so messy that we have reluctantly decided to spare our readers.

Although it is certainly more realistic to include these factors, they alone cannot explain the differences we observe in modern economies. The fact that cities produce differing bundles of commodities probably explains more of the variation in their sizes than, for example, consumer tastes.

Consider the effect of introducing more than one urban good into the model with increasing returns to scale. If the goods have different production functions, the cities will have different sizes.

If we ask whether a city can produce more than one good in our model, we discover an important implication of the assumption that transporting goods costs nothing. Commuting costs can be saved by separating firms producing different goods, without incurring any additional costs, so two-product cities will not occur.

If transporting urban goods is costly, however, cities producing more than one good might well arise. The saving on transporting wet concrete or bottled coke to demanders, for example, might justify the extra commuter costs that result from having a concrete plant and a bottling plant in each city.

The cost of goods transport has a strong influence on city form as well as size, although the subject is outside the range of this chapter. Even if two or more commodities are produced in a city, the firms will not necessarily all locate at the center. Retail stores, for example, disperse throughout a city to reduce the transportation costs of shopping for consumers. Moreover, there is no a priori reason to expect that a concrete plant and a bottling plant locate at the center. They might locate at the edge of a city to take advantage of lower land costs, and form a

multi-centered city.

There is another problem in multi-product cities caused specifically by the cost of goods transport. There can be only one firm with the greatest returns to scale in a city. If there were two, and if we could ignore the fact that the numbers of firms in other industries must be integers,[6] then we could split the city into two. Production costs would not increase in any industry and commuting costs would decrease, and society be better off with two cities instead of one.

Therefore, we have to introduce externalities in order to attain a more realistic system of cities. The simplest way is to add another urban good to the framework in section 5. If there are two urban goods, we obtain three types of cities:

two producing only one good, and one producing both. It is easy to see that the same results as those in section 5 can be obtained for each type of city.

However, there is no guarantee that cities producing both goods are bigger than cities producing only one good. For example, if externality works only through the total population of a city, cities producing two goods have no more benefit from becoming bigger than cities producing one good. Therefore, we might want to assume that there is a special benefit which arises from having two industries together.

Although introducing cross product externalities is attractive, and would give rise to more realistic system of cities in our model, the analysis is simply too difficult for the present work. We do not, therefore attempt to build a model of a system of cities of this type here.

Notes

Until Alonso's work (1971), the analysis of city sizes had been limited to the cost side, and the city size which gave the minimum cost had been considered optimal. Alonso introduced the output side, regarding a city as an aggregate production unit. There are two types of optimum city size in this model. For the residents the optimum size is that which maximizes the difference between the average product (AP) and the average cost (AC). For a national government interested in maximizing total product under conditions of labour surplus, the optimum size is where the marginal product (MP) is equal to the marginal cost (MC). If the supply of labour is limited, this condition should be modified. MP may not equal MC although the difference between MP and MC must be the same for all cities. Alonso pointed out that if individuals maximize the difference between AP and AC, per capita tax of MP-AP-(MC-AC) can result in the optimum city size.

Although Alonso's work was a big step forward in constructing the economic theory of city sizes, his approach has the following shortcomings. First, the analysis is partial in nature, since only one city is considered: if the city is placed in a general

---

[6] For example, if there are two firms of the greatest degree of increasing returns and three firms of the second greatest degree of increasing returns, splitting this city into two may involve an extra social cost since a city cannot have one and a half firms.

equilibrium framework, we may face different problems. Second, the welfare aspect of the analysis is not very clear, since utility functions for households are not introduced. Third, the cause of increasing average product is not explicitly formulated. It is not clear, therefore, how individual firms and households behave in a market economy: increasing returns to scale for a firm, and external economies among firms have very different implications on individual behaviour. Fourth, the spatial aspect of cities is ignored.

There have been several attempts to overcome these shortcomings. Borukhov (1975) built a very simple model of an economy consisting of many cities. He showed that Alonso's second condition for the optimum city size is correct if the number of cities is given: at the optimum the difference between MP and MC is equal for all cities, but MP exceeds MC by an amount which has been interpreted as the opportunity costs of siting the population in alternative cities. If the number of cities is a variable, however, this condition is not sufficient to characterize the optimal solution. Since Borukhov was worried about integerness of the number of cities, he could not obtain a transparent condition for the optimum number of cities. However, if one is willing to approximate the number of cities by a continuous variable, and to assume that all cities are the same (as done in this chapter), it is easy to see that at the optimum the difference between MP and MC is equal to the difference between AP and AC. This means that the difference between AP and AC is maximized at the optimum number of cities. Therefore, the optimum for the residents coincides with that for a national government.

If the difference between MP (MC) and AP (AC) is caused by externalities, the Pigouvian tax/subsidy discussed by Alonso is necessary. However, our result suggests that the net Pigouvian tax/subsidy is zero at the optimum number of cities. Unfortunately, this does not imply that the optimal allocation is automatically attained by market mechanism. As seen in section 5, city sizes tend to be too big because of the difficulty in forming a coalition to create a new sufficiently large city.

Henderson (1974) formulated a more sophisticated model with three industries. The first is the export industry, which faces a fixed export price. The export industry is assumed to have increasing returns to scale. The second is the housing industry, which is assumed to have constant returns to scale. Finally, the third industry produces an intermediate good which is used as an input (called sites) to the above two industries. This industry represents the spatial aspect of cities (for example, commuting costs) which works to discourage formation of big cities. Instead of explicitly introducing spatial dimension, Henderson assumed that sites are produced with labour under decreasing returns to scale. The optimum city size balances increasing returns to scale in the export industry, and decreasing returns to scale in the site industry.

One of the most important findings by Henderson is that a market economy tends to overshoot the optimum city size because of difficulty of forming a coalition to create a big city. Our argument in section 5 is based on his observation.

Henderson (1977) extended this analysis to a spatial model and obtained (independently of our work) results similar to ours in the Marshallian externality case.

One of the major differences is that he worked with special functional forms of production functions and utility functions, whereas we assume general functional forms.

Henderson (1974) and Tolley (1974) analyzed the size of a city, considering the rest of the economy as given. Both focused on the effect of pollution taxation on the city size. Henderson showed that, since pollution taxation increases the welfare of city residents, the city size rises with immigration from the rest of the economy. In Tolley's model pollution taxation increases the city size if the externalities originate in production of nontraded goods, but might reduce the city size if the externalities originate in export production.

Serck-Hanssen, in a pioneering but little known work (1969), first obtained the condition for the optimal number of cities discussed in section 3. Adopting a framework due to Losch, he considered firms supplying their products in a space over which consumers are homogeneously distributed. Instead of assuming commuting costs, he assumed positive transportation costs for products. His optimality condition is essentially the same as ours, although in his model there is a complication arising from the fact that the optimal market areas are not circular but hexagonal in a two-dimensional space.

Mirrlees (1974), Dixit (1973), and Starret (1974) derived conditions for optimal city size in models of closed economies similar to ours. All of them assumed increasing returns to scale in the urban industry, and obtained results equivalent to that in section 3: the excess of marginal over average productivity equals the average differential rent (minus a correction if environmental externalities are present as in Mirrlees' model). Concentrating on optimal allocations, they did not analyze how the market city size is determined.

Vickrey (1977), in a very simple model, derived the result that the aggregate land rent equals the loss of a firm at the optimum, and argued that competition among cities leads to an efficient allocation. Although his analysis is not rigorous, it has the same spirit as our analysis of a system of cities formed by land-developer firms.

Arnott and Stiglitz (1975) introduced a public good which is local to a city while assuming constant returns in the production sector. In this case the optimum city size is characterized by the condition, à la Henry George, that the cost of the public good is equal to the total differential rent of a city. They also derived the following interesting formula: if the commuting costs are given by a linear function of distance (in our notation $t(x) = tx$), the aggregate differential rent equals the aggregate transportation costs in a linear city ($\theta(x) = \theta$), or one half of the aggregate transportation costs in a circular city ($\theta(x) = 2\pi x$). Arnott (1979) generalized these results to include congestion in transportation, economies of scale in production, and other matters.

The central place theory originating from the seminal work of Christaller (1966) and Lösch (1954) has a close relationship with our discussion of a system of cities in section 6. Assuming that demand is uniformly distributed over space, the theory considers the spatial pattern of suppliers of goods. A hierarchical structure of central places is derived by superimposing the geographical networks of market areas for goods with different market sizes. As pointed out by Eaton and Lipsey (1979) among others,

the economic foundations of the theory are incomplete in an important respect. The crucial assumption to obtain a hierarchical structure is that the location of a firm producing a good with a large market area attracts producers of other goods with smaller market areas. Under this assumption, there is a hierarchy of central places: the biggest having producers of all goods, the second biggest having producers of goods with smaller market areas, and so on. However, there is no explicit analysis of the force that causes producers to group together in this way. Eaton and Lipsey built a model in which multipurpose shopping offers an incentive for the formation of central places. Our discussion in section 6 attempts to indicate how a theory of central places might be based on the economic forces causing the agglomeration of different industries.

REFERENCES

Alonso, W., (1971), "The Economics of Urban Size", *Papers of Regional Science Association* 26, 67-83.

Aoki, M., (1971), "Marshallian External Economies and Optimal Tax Subsidy Structure", *Econometrica* 39, 35-54.

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

Arnott, R. and J.E. Stiglitz, (1975), "Aggregate Land Rents, Aggregate Transport Costs and Expenditure on Public Goods", Discussion Paper #192, Institute for Economic Research, Queen's University.

Borukhov, E., (1975), "Optimality in City Size and System of Cities: A Comment", *Urban Studies* 12, 325-328.

Chipman, J.S., (1970), "External Economies of Scale and Competitive Equilibrium", *Quarterly Journal of Economics* 86, 347-385.

Christaller, W., (1966), *Central Places in Southern Germany,* (Prentice Hall, New Jersey).

Dixit, A., (1973), "The Optimum Factory Town", *The Bell Journal of Economics and Management Science* 4, 637-651.

Eaton, B.C. and R.G. Lipsey, (1979), "Microeconomic Foundations of Central Place Theory", Discussion Paper No. 327, Institute for Economic Research, Queen's University.

Henderson, J.V., (1974), "The Sizes and Types of Cities", *American Economic Review* 64, 640-657.

Henderson, J.V., (1974), "Optimum City Size: The External Diseconomy Question", *Journal of Political Economy* 82, 373-388.

Henderson, JV., (1977), *Economic Theory and the Cities,* (Academic Press, New York).

Livesey, D.A., (1973), "Optimum City Size: A Minimum Congestion Cost Approach", *Journal of Economic Theory* 6, 144-161.

Lösch, A., (1954), *The Economics of Location,* (Yale University Press, New Haven and London).

Mirrlees, J.A., (1972), "The Optimum Town", *Swedish Journal of Economics* 74, 114-135.

Richardson, H.W., (1972), "Optimality in City Size, System of Cities and Urban Policy: A Skeptic's View", *Urban Studies* 10, 29-48.

Serck-Hanssen, J., (1969), "The Optimal Number of Factories in a Spatial Market", in: H. Bos (ed.). *Towards Balanced International Growth,* (North-Holland, Amsterdam).

Starrett, D.A., (1974), "Principles of Optimal Location in a Large Homogeneous Area", *Journal of Economic Theory* 9, 418-448.

Tolley, G.S., (1974), "The Welfare Economics of City Bigness", *Journal of Urban Economics* 1, 324-345.

Vickrey, W., (1977), "The City as a Firm", in: M.S. Feldstein and P.P. Inman (eds.). *The Economics of Public Services,* (Macmillan, London).

CHAPTER III
LOCAL PUBLIC GOODS

The spatial equilibrium model in Chapter I can be used to analyze problems associated with optimal provision of local public services. In the case of *pure* public goods it is extremely difficult to achieve the optimal allocation by a decentralized mechanism. Local public goods which, while still public, are not perfectly public, however, allow the introduction of competition among suppliers and it is possible to devise a competitive mechanism which achieves the optimal allocation.

A pure public good is consumed collectively: its consumption by any individual does not reduce the amount available for others. The classic example of a virtually pure public good is national defense. It is claimed that the amount of "security" one person "consumes" from her nation's "defense expenditure" has no effect on the amount available for others: the entire population is able to consume a pure public good.

Conventional public good theory assumes that the number of consumers is fixed since the size of the community - usually a nation - is known. For local public services the assumption breaks down because the population of local communities is endogenous, determined in the system's search for equilibrium.[1] It is possible to take advantage of this problem, however.

We know from the Fundamental Theorem of Welfare Economics that, if there are only pure private goods, a competitive equilibrium is Pareto optimal, that is, no one can be made better off without making somebody else worse off. When there are public goods, however, a competitive equilibrium fails to attain Pareto optimality, and furthermore it is difficult to devise any other workable decentralized mechanism. The problem arises because households have an incentive to "misreveal" their preferences. By understating the marginal benefit it gains from the public good, a household can avoid being assessed its full share of the cost of providing the good, without suffering a reduction in supply. Supply is unaffected , because the contribution of a single household is negligible. This difficulty is often called the "free rider" problem.

Since a pure public good is consumed by all households concurrently, a marginal increase in supply benefits all households simultaneously. The marginal social benefit is therefore the sum of the benefits received by each household, which may be expressed as the sum of marginal rates of substitution between the public good and the numeraire.

If all households were to pay the full value of the benefit they received, profit maximization would yield an efficient allocation. Because of the free rider problem, however, it is extremely difficult to devise a pricing scheme in which every household has an incentive to reveal its marginal evaluation of the public good.[2]

---

[1] Stiglitz (1977) emphasized this aspect of local public goods.

[2] Although dark (1971), Groves (1973) and Groves and Ledyard (1977) invented a mechanism in which a household has an incentive to reveal its preferences correctly, this mechanism is rather artificial. Green and Laffont (1977) proved that this mechanism is the only one that does not have the preference revelation problem.

In the case of local public goods, competition between different communities can work in a manner similar to competition between suppliers of private goods. The preference revelation problem still remains *within* a community since a local public good has the same characteristics as a pure public good within a community. It is, however, possible to exploit the special property of local public goods, the fact that the population of beneficiaries is endogenous to the system. If a community increases the supply of local public goods, the community becomes more attractive, which induces immigration of households. This increases demand for housing, causing land rent to rise. The marginal social benefits of the public goods are therefore reflected, at least partially, in the marginal increase in land rent. If the community is infinitely small relative to the rest of the world, the marginal benefits equal the increase in the total land rent in the community. Then the behaviour which is characteristic of a land developer, maximizing land rent net of the cost of providing the public goods, leads to the efficient supply of the public goods.

In order to illustrate the basic principle, we start in section 1 with a simple case. Public goods are assumed to be extremely local in the sense that they are jointly consumed only by residents at a location. To simplify the analysis we assume that public goods supplied at a certain distance from the city center can be consumed only by residents living at that distance from the center. In effect we pretend that neighbourhoods form a series of concentric rings, each of unit width, around the city center. It may seem a bit peculiar, but the assumption is nothing more than a mathematical convenience which yields perfectly sensible and general results. This type of public good represents, in an extreme form, goods consumed only by households living very close to the location of supply; street lighting, for example, or neighbourhood beautification, or snow removal. The extremely local public goods are embedded in the closed city of Chapter I.

Not surprisingly, the optimum solution must meet the Samuelsonian condition that the sum of marginal rates of substitution be equal to the marginal cost of the public good. Another interesting property of the optimal solution is that the differential rent (the difference between the urban rent and the rural rent) at the edge of the city equals the cost of the public good there.

The optimal solution can be achieved either by centralized control, which requires impractical amounts of information, or through a decentralized mechanism such as a system of neighbourhood development corporations which rent land at the rural rent and maximize their profits. In the second half of section 1 a system of competitive land developers with a developer in each neighbourhood is described and its optimality demonstrated.

In section 2 we examine a crowding phenomenon by assuming that the cost of producing the same amount of the public good rises as the number of residents increases. The major difference in this case is that the optimal solution requires a congestion tax on households. The congestion tax at any location equals the marginal increase in the cost of the public good caused by adding a household there. The system of competitive land developers achieves the optimal allocation if a land developer charges the congestion tax and maximizes rent plus tax minus the costs of providing the public good.

In section 3 we consider a local public good which is jointly consumed by all residents in an entire city, rather than by residents at a certain radius. Museums, theaters, sewage systems, and large parks may fit this category. Competition between cities is introduced by assuming that there are many identical cities. The results are parallel to those in the increasing-returns-to-scale case of the previous chapter, as well as those in the extremely local public good case of the present chapter. If a competitive land developer develops an entire city, the local public good is optimally supplied when the number of cities is very large. Moreover, the zero profit condition from free entry insures the optimum number of

cities.

Crucial in deriving our results is that, in the eyes of a developer, the utility level of the residents is fixed. This suggests that we can extend the result to more general models as long as this condition is guaranteed. In Appendix II we consider one example of such an extension, in which two inputs, land and labour, are used in production.

It is worth emphasizing that our results depend on the assumption that all households in the economy are identical in terms of both skills and preferences. Although we may relax this assumption to include different types of households, we must assume that there are many households in each type in the whole economy and that one region contains a very small fraction of the households in each type. Since identical households receive the same utility level in equilibrium, regardless of where they live, a change in the supply of local public goods in one small region has a negligible effect on the general utility level. If all households are different, however, the utility levels of residents cannot be taken as constant even in the case where the population of the region is very small compared with the rest of the world. Therefore, at best we can only say that the system of competitive land developers approximates the optimal allocation of local public goods. How good an approximation it achieves is an empirical question. Considering the fact that there is no perfect mechanism to supply public goods, however, our scheme of letting competitive land developers supply local public goods is worth a serious consideration. Our result would suggest, for example, that when a land developer develops a new community, the developer rather than a local government should pay for the public good supplied in the community.

1. An Extremely Local Public Good

Consider an extremely local public good in the public-ownership, closed-city case of Chapter I. The amount of public good supplied between x and $x+dx$ is denoted by $X(x)dx$. Though we consider only one public good for notational simplicity, the conclusions obtained in this section are valid for any number. The public good is extremely local in the sense that the public good supplied at x is jointly consumed only by residents of a ring of unit width between $x-\frac{1}{2}$ and $x+\frac{1}{2}$. If we assume that public goods supplied at different radii are perfect substitutes, then a household at x had available.

$$\int_{x-1/2}^{x+1/2} X(x')dx'.$$

or approximately $X(x)$, of the public good and its utility function can be written

$$u(z(x),h(x),X(x)). \tag{1.1}$$

It is assumed that the consumer good is the only input in the production of the public good. The public good is assumed to be produced separately at each location at a cost $C(X(x))$. Then the resource constraint (I.1.30) is rewritten as follows.

$$\int_0^{\bar{x}} \{[z(x)+t(x)]N(x)+c(X(x))+R_a\theta(x)\}dx = F(P) \tag{1.2}$$

The land constraint is the same as (I.2.2), and the population constraint as (1.1.24):

$$\theta(x) = h(x)N(x) \tag{1.3}$$

$$P = \int_0^{\bar{x}} N(x)dx\,. \tag{1.4}$$

The sum of the equal utilities,

$$\int_0^{\bar{x}} N(x)dx\,, \tag{1.5}$$

is maximized under the constraints (1.2), (1.3), (1.4) and the equal utility constraint,

$$u(z(x), h(x), X(x)) = u\,. \tag{1.6}$$

The first order conditions for this problem become, after simple rearrangements:

$$\frac{u_h}{u_z} = R(x)\,, \tag{1.7a}$$

$$\frac{u_X}{u_z} N(x) = c'(X(x))\,, \tag{1.7b}$$

$$y = z(x) + t(x) + R(x)h(x)\,, \tag{1.7c}$$

$$R(\bar{x}) - R_a = \frac{c(X(\bar{x}))}{\theta(\bar{x})}\,. \tag{1.7d}$$

(1.7a) and (1.7c) are the same as in Chapter I. (1.7a) equates the marginal rate of substitution between housing and the consumer good to the shadow rent. (1.7c) states that the household expenditure on private goods, evaluated at the shadow prices, must be the same everywhere in the city.

Conditions (1.7b) and (1.7d) are new. (1.7b) is the Samuelsonian condition for efficient supply of the public good described in the introduction: the marginal cost of the public good at x must equal the sum over all residents at x of the residents' marginal rates of substitution between the public good and the consumer good. A unit increase in the supply of the public good between x and $x+dx$ raises the utility level of a household there by u_X. Since $N(x)dx$ households receive the benefits of the public good, the marginal social benefit in utility terms is $u_X N(x)dx$, and in pecuniary terms $(u_X / u_Z)N(x)dx$. The social optimum is achieved when the marginal benefit equals the social marginal cost, $c'(X(x))dx$.[3]

(1.7d) shows that the shadow rent at the boundary of the city is not equal to the rural

---

[3] If we go back to the original formulation, a household at x has available $\int_{x-\frac{1}{2}}^{x+\frac{1}{2}} X(x')dx'$ of the public good. Consider the costs and benefits of a unit increase of $X(x)$ between x and $x+dx$. The costs are $c'(X)dx$. On the other hand, the utility level of a household between $x-1/2$ and $x+1/2$ rises by $u_X dx$, and the marginal benefit a household receives is $(u_X / u_z)dx$ in pecuniary terms. The social benefit is obtained by summing this over all households between $x-1/2$ and $x+1/2$ so that the optimality condition is

$$\left[\int_{x-\frac{1}{2}}^{x+\frac{1}{2}} \frac{u_X}{u_z} N(x')dx'\right]dx = c'(X(x))dx\,.$$

Equation (1.7b) is obtained if we can approximate $(u_X / u_z)N(x')$ for all x' between $x-1/2$ and $x+1/2$ by $(u_X / u_z)N(x)$.

rent as in Chapter I, but rather greater than the rural rent by the cost-per-unit-area of producing the public good there.

This optimal solution can be achieved in the following ways. First, local governments might supply the local public good so as to equate the sum of marginal rates of substitution to marginal cost of the public good at each location. The city would lease the land to those who pay the highest rent, which would be $R(x) = u_h / u_Z$ in market equilibrium. Part of the revenue would then be used to produce the public good and the rest returned to residents as an equal subsidy. The public good would be supplied out to the radius where the market rent minus the rural rent equals the cost of the public good per unit acre. Under this arrangement utility maximization by households ensures conditions (1.6) and (1.7a) and the market equilibrium attains the optimal allocation.[4] Unfortunately, this method is not practical since local governments must know the marginal rates of substitution, and these are very hard to discover.

The second way to implement the optimal solution can be seen as a system of land developers. Imagine a large number of developers in a city, each developing an extremely small area, and each supplying the local public good in their area. The developers rent land from the rural landlords and sublet it to city residents at the market rent. In our circular city, it is convenient to allow each developer to develop a band around the city center at a given radius. The developer's profit, which is the differential rent minus the cost of providing the public good, becomes

$$\left[R(x) - R_{\bar{a}}\right]\theta(x) - c(X(x)) .$$

In order to ensure that all households obtain the same utility level, we assume that the profit is distributed equally among all city residents.

Since each developer is very small, its action does not significantly affect the utility or the income levels. Therefore, when he changes the supply of the public good, land rent moves in such a way that utility and income both remain unchanged. The change in land rent can be obtained as follows. A household maximizes the utility function (1.1) under the budget constraint (1.7c), which can be summarized as the indirect utility function,

$$v(y - t(x), R(x), X(x)) \qquad (1.8)$$

as in (I.1.7). Equating the indirect utility function to the fixed utility level, u, we obtain the

---

[4] The reader may wonder whether a household would not prefer to rent land directly from the rural owners or the central government and live outside the boundary of the city, where the public good is not supplied. If the optimal solution requires a positive supply of the public good at the boundary of the city, then households do not have an incentive to live in the places where the public good is not supplied. It suffices to show that households obtain higher utility at the boundary if the public good is supplied than not, since locations farther than the boundary are even less desirable.

From (1.7c) and (1.7d), the following resource constraint is satisfied at $\bar{x}$.

$$y = z(\bar{x}) + t(\bar{x}) + R_a h(\bar{x}) + \frac{c(X(\bar{x}))}{N(\bar{x})} \qquad (*)$$

A household which lives on the other side of $\bar{x}$ has the budget constraint;

$$y = z + t(\bar{x}) + R_a h \quad . \qquad (**)$$

Since the same amount of resource is used up in both cases, under (*) be higher than or equal to the maximum attainable utility level under the budget constraint (**). Otherwise, the utility level of $\bar{x}$ can be increased by making the supply of the public good zero without lowering the utility level of other locations.

bid rent function,

$$R(y - t(x), u, X(x)) \tag{1.9}$$

as in (I.1.12).

A profit maximizing developer at x maximizes

$$[R(y - t(x), u, X(x)) - R_a]\theta(x) - c(X(x)) \tag{1.10}$$

with respect to $X(x)$, which yields

$$R_X \theta(x) = c'(X). \tag{1.11}$$

This implies that the optimality condition (1.7b) is satisfied. By Roy's Identity (1.1.10) the bid rent function satisfies

$$R_X = \frac{l}{h}\frac{v_X}{v_I}.$$

Noting that $v_X = u_X$ and $v_I = u_Z$ by the Envelope Theorem[5], we can rewrite this equation as

$$R_X = \frac{l}{h}\frac{u_X}{u_Z}. \tag{1.12}$$

Equation (1.5b) follows, since from the land constraint (1.3), $\theta(x)/h(x) = N(x)$.

The land developer operates only when profit can be made:

$$[R(x) - R_a]\theta(x) - c(X(x)) \geq 0. \tag{1.13}$$

This condition insures that (1.7d) is satisfied at the edge of the city.

Thus the system of land developers achieves the optimality conditions (1.7b) and (1.7d). Since other conditions are also satisfied in market equilibrium, the optimal allocation can be reproduced if the local public good is supplied by extremely small land developers.

Note that developers need to know only the land rent, and not the utility function. Therefore, the informational requirement is the same as the usual price mechanism. There still remains, however, a difference from the market system for private goods. Since firms and households maximize their objective functions taking prices as given, maximization processes are not affected by situations outside them, whereas the maximization problem for land developers involves an important *endogenous price,* namely, land rent, which is determined through reactions of households to the supply of the public good. Therefore, the profit-maximizing level of the local public good can only be found after observing

[5] See Appendix III on the envelope property.

levels of land rent corresponding to many different supply levels.

The system of land developers may be interpreted as the mechanism proposed by Negishi (1972) and combining Margolis' principle of fiscal profitability with Tiebout's voting with one's feet. According to the principle of fiscal profitability, a local government pays for the local public good from a tax on land, and determines the supply of the public good which maximizes the rent net of the tax. This behaviour is identical to the profit-maximizing behaviour of a developer. Voting with one's feet allows households to choose the local government that offers the preferred bundle of local public goods. In our model the free choice of location represents voting with one's feet. This, coupled with the assumption of extremely small local governments, will insure that local governments take as given the utility level of residents.

The above result relies on the fact that the marginal benefits of the public good are capitalized in land rent. Multiplying (1.12) by $\theta(x)$, we obtain

$$\theta(x)R_X = \frac{u_X}{u_z}N(x): \tag{1.14}$$

the marginal increase in land rent at x, caused by a unit increase of the public good, equals the sum of the marginal rates of substitution between the public good and the consumer good, which in turn equals the marginal benefits of the public good. This result is characteristic of a small economy in which the utility level can be taken as given, and is independent of the public good being optimally supplied. The benefit of the public good must accrue to somebody or become a deadweight loss. Since there is no deadweight loss in the first best world, all the benefits must be received by somebody. In our model, the only place the benefits can go is land rent.

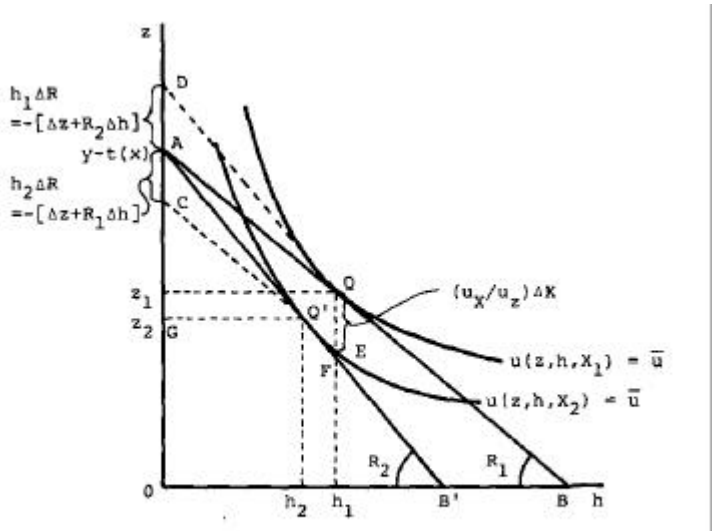

Figure 1. Capitalization of the Benefits of a Public Good

Figure 1 illustrates the capitalization of the benefits of public goods. Consider an increase in the supply of the public good from x_1 to x_2. Then a smaller bundle of (z,h) is necessary to achieve the same utility level, u, and the indifference curve shifts toward the origin. The equilibrium consumption moves from Q to Q'. The benefits of the increase in the supply of the public good can be represented by the amount of resources freed by this move. Since both z and h change, we must evaluate the change by using some relative

price. There are at least two possibilities. If we use the before-the-change rent, R_1, the benefits of this change are AC in Figure 1, or $-[\Delta z + R_1 \Delta h]$; and if we use the after-the-change rent, R_2, the benefits are AD, or $-[\Delta z + R_2 \Delta h]$.

From Figure 1 (or from simple algebraic manipulations) it is clear that AC, which is $AG - CG$, also equals the change in rent, $\Delta R = R_1 - R_2$, multiplied by the after-the-change consumption level of housing, h_2, i.e., $h_2 \Delta R$; and that AD equals the change in rent multiplied by the before-the-change consumption, $h_1 \Delta R$.

Although it is not clear in this partial analysis which measure of benefit is a better approximation[6], if the change in X is infinitesimal, the two measures coincide, and the problem disappears. For a marginal change in X, therefore, the benefits a household receives equal $h\frac{dR}{dX}$, which is equivalent to (1.10). The social benefit is the sum of the benefits of all households who consume the public good and is given by $\theta(x)\frac{dR}{dX}$ in our model. Thus the rise in land rent completely capitalizes the marginal benefits of the public good.

The diagram also shows that the marginal rate of substitution between the public good and the consumer good is the correct measure of the marginal benefit of the public good which a household receives. When the consumption of land is held constant, a reduction in the consumption of the consumer good made possible by the increase in the public good equals QE. If the change in the supply of the public good is small, QE is approximately $\Delta z = (u_X / u_z)\Delta X$, since by total differentiation

$$u_z dz + u_X dx = du = 0,$$

where $\Delta X \equiv X_2 - X_1$. Moreover, as ΔX approaches zero, QF approaches $(u_X / u_z)\Delta X$. QF equals AD, and hence gives the benefit of the marginal increase evaluated at the after-the-change price. Thus u_X / u_z is the correct measure of the marginal benefit of the public good.

2. An Extremely Local Congestible Public Good

In the previous section we assumed that the local public good was a pure public good at each radius. In particular, we assumed that the costs of providing the same level of the public good did not depend on the number of consumers. This assumption does not hold for most public services. For example, the same park gives different levels of services depending on the number of people using it. The cost of providing the same level of park services usually increases as the number of users increases.

In this section we assume that the cost of producing the same level of the public good increases as population density increases. The cost function for the local public good is modified as

---

6 Following the approach due to Negishi (1972), Harris (1978) showed, in the context of public inputs rather than public consumption goods, that the value of the change evaluated at the after-the-change prices is the lower bound of the benefits and that the value at the before-the-change prices is the upper bound. Since in our case the value of the change evaluated at the after-the-change price is greater (in the absolute value) than the value at the before-the-change prices, Harris' result must clearly be modified. It is still an open question whether a similar relationship can be established in our model.

$$c(X(x), N(x)), \tag{2.1}$$

where

$$C_N > 0 \quad .$$

As in the previous section, the optimal solution can be easily obtained. The first order conditions are (1.7a) and

$$\frac{u_X}{u_Z} N(x) = C_{X'} \tag{2.2a}$$

$$y = z(x) + t(x) + R(x)h(x) + C_N \tag{2.2b}$$

$$c(X(\bar{x}), N(\bar{x})) = [R(\bar{x}) - R_a]\theta(\bar{x}) + N(\bar{x})c_N . \tag{2.2c}$$

(2.2a) is the same as before: the marginal cost of the public good must equal the sum of marginal rates of substitution between the public good and the consumer good for all households at each radius. Terms in (2.2b) and (2.2c) containing C_N are new. In order to achieve this solution in a market system, a household must pay a congestion tax equal to the marginal cost of adding a household, $c_{N'}$, and varying with distance from the center. Then (2.2c) states that the government budget is balanced at the edge of the city. The sum of the revenues from the congestion tax and the land rent is exactly equal to the sum of the rural rent and the cost of the public good at $\bar{x}$.

Consider again a system of competitive neighbourhood developers supplying the public good. As before we assume that no developer is large enough to affect the utility and income levels. We now assume that each developer charges a congestion tax (or the membership fee to join the location) and maximizes profit including the tax. If the congestion tax at x is denoted by $s(x)$, the developer at x maximizes

$$R(x)\theta(x) + s(x)N(x) - c(X(x), N(x)) . \tag{2.3}$$

The policy variables for the developer are $s(x)$ and $X(x)$. $R(x)$ and $N(x)$ are determined through the market's adjustment.

As in the previous section (c.f., Equation (1.9)), we can derive the bid rent function;

$$R(y - t(x) - s(x), u, X(x)) . \tag{2.4}$$

As in (I.1.14), the function satisfies

$$R_I(I(x), u, X(x)) = \frac{1}{h(x)}, \tag{2.5}$$

where $I(x) \equiv y - t(x) - s(x)$. The number of households at x, therefore, satisfies

$$N(x) = \theta(x) R_I(y - t(x) - s(x), u, X(x)) . \tag{2.6}$$

Thus a developer maximizes

$$\begin{aligned} &R(y - t(x) - s(x), u, X(x))\theta(x) \\ &+ s(x)\theta(x) R_I(y - t(x) - s(x), u, X(x)) \\ &- c[X(x), \theta(x) R_I(y - t(x) - s(x), u, X(x))] \end{aligned} \tag{2.7}$$

with respect to $s(x)$ and $X(x)$. It is easy to see that optimization with respect to $s(x)$ yields

$$s(x) = c_N . \tag{2.8}$$

As in the previous section, optimization with respect to $X(x)$ yields (2.2a), and the nonnegative-profit condition guarantees (2.2c). Thus the optimal supply of the public good and the optimal level of the congestion tax are obtained.

In the previous section we showed that the marginal benefit of the public good is fully reflected in the increase in land rent. It may seem plausible that, when there is a congestion tax, some of the benefit of an increase in the supply of the public good will show up as an increase in tax revenue, so that the marginal benefit would equal the change in the sum of land rent and the congestion tax. Differentiating the sum, however, yields

$$\begin{aligned}
&\frac{d}{dX}[\theta(x)R(y - t(x) - s(x), u, X(x)) + N(x)s(x)] \\
&= \theta(x)[-R_I \frac{ds}{dX} + R_X] + N(x)\frac{ds}{dX} + s(x)\frac{dN}{dX} \\
&= \theta(x)R_X + s(x)\frac{dN}{dX} \\
&= \frac{u_X}{u_z} N(x) + s(x)\frac{dN}{dX}
\end{aligned} \tag{2.9}$$

where the second and the last steps use (2.5) and (2.8) respectively. The change in the sum, therefore, exceeds the marginal benefit, and the difference is the increase in tax revenue caused by an induced change in population, $s(x)(dN/dX)$. The increase in population raises the tax revenue, but at the same time increases the cost of producing the public good. From (2.8), the two increases are equal at the optimum, and the in-crease in tax revenue, being completely absorbed by the increased costs, does not constitute net social gain.

(2.9) also shows that, if the congestion tax, $s(x)$, is held constant, the earlier result follows:

$$\begin{aligned}
&\frac{d}{dx}\left[\theta(x)R(y - t(x) - s(x), u, X(x))\right] \\
&= \frac{u_X}{u_z} \qquad \text{if} \quad s(x) = \text{constant}
\end{aligned}$$

Thus, if, for example, the marginal cost of a population increase, c_N, is constant, the marginal benefit of the public good exactly equals the increase in land rent.

3. A Public Good Local to a City

In this section a local public good is assumed to be jointly consumed by all residents of a city. Consider n identical cities which produce the consumer good under constant

returns to scale. A city's production function is wP_c, where P_c is the population of the city and the marginal product of labour, w, is constant. Note that the existence of a public good provides a reason for having a city: an increase in population lowers the *per capita* cost of supplying the same amount of the public good. Cities, therefore, may exist even if production technology has constant returns to scale.

The utility function of a household is

$$u(z(x), h(x), X), \tag{3.1}$$

where X is the consumption of the local public good and is equal for all residents in a city. The cost in terms of the consumer good of the public good is

$$C(X), \tag{3.2}$$

where there is no congestion effect.[7] We do not explicitly introduce a rural sector but the rural rent, R_a, is assumed to be paid by cities. Then the resource constraint is

$$\int_0^{\bar{x}} \{[z(x)+t(x)]N(x)+R_a\theta(x)\}dx + C(X) = wP_c. \tag{3.3}$$

The total population, P, of city residents is assumed to be given. The population constraints are

$$P = nP_c \tag{3.4}$$

and

$$P_c = \int_0^{\bar{x}} N(x)dx. \tag{3.5}$$

Our optimization problem is one of maximizing the sum of equal utilities,

$$n\int_0^{\bar{x}} N(x)dx, \tag{3.6}$$

under the above constraints (3.3)-(3.5) and the constraint that all households have the same utility level,

$$u(z(x), h(x), X) = u. \tag{3.7}$$

If the number of cities is fixed, we obtain the first order conditions (1.5a), (1.5b), and

$$\int_0^{\bar{x}} \frac{u_X}{u_Z} N(x)dx = C'(x), \tag{3.8a}$$

$$R(\bar{x}) = R_a. \tag{3.8b}$$

---

[7] This formulation implicitly assumes no transportation costs for the local public good. In this sense the public good is like a telephone system, a cable television network or a sewage system but not like a theater or a central park. Transportation costs of a local public good can, however, be easily introduced and do not change our results. If the public good is supplied at the center of the city, we may even interpret $t(x)$ as including transportation costs of the public good.

(3.8a) is the Samuelsonian condition that the sum of marginal rates of substitution over all residents in a city must equal the marginal cost of the public good. (3.8b) is a familiar equality between the urban rent at the edge of a city and the rural rent.

If the number of cities is a policy variable, we must add the following condition:

$$\int_0^{\bar{x}} [R(x) - R_a]\theta(x)dx = C(X). \tag{3.9}$$

The total differential rent is equal to the total cost of the public good. Therefore, if a city government collects land rent, pays the rural rent, and supplies the local public good, its budget is balanced at the optimal number of cities.

Now, consider the benefit of the public good in a market economy. We first derive a formula which holds for any type of city, and then consider the special cases of a closed city and a small open city in an economy with many cities. In our market cities, city governments are assumed to collect the land rent, and to return the surplus, after the payment of the rural rent and the cost of the public good, to residents as an equal subsidy. Since everybody has the same marginal productivity, the wage rate is also the same, and therefore income is the same for all households. Then the budget constraint for a household is given by

$$y = z(x) + t(x) + R(x)h(x), \tag{3.10}$$

for an appropriate income y.

The bid rent function can be derived as in the previous sections:

$$R(x) = R[y - t(x), u, X]. \tag{3.11}$$

The effect on land rent of a change in the supply of the public good is

$$\begin{aligned}\frac{dR(x)}{dX(x)} &= R_I \frac{dy}{dx} + R_u \frac{du}{dx} + R_X \\ &= \frac{l}{h(x)} \frac{dy}{dx} + \frac{l}{v_I h(x)} \frac{du}{dX} + \frac{l}{h(x)} \frac{u_X}{u_Z},\end{aligned} \tag{3.12}$$

where the second equality is obtained from (2.5), (I.1.15) and (1.10). Multiplying both sides by $\theta(x)$, integrating from 0 to $\bar{x}$, and rearranging terms, we obtain

$$\int_0^{\bar{x}} \frac{u_X}{u_Z} N(x)dx = \int_0^{\bar{x}} \frac{dR(x)}{dx}\theta(x)dx + \left[\int_0^{\bar{x}} \frac{l}{v_I} N(x)dx\right]\frac{du}{dx} - P_c \frac{dy}{dx}. \tag{3.13}$$

Thus the marginal benefit of the public good is reflected in the changes of land rent, the utility level and the income level. Notice that this equation holds for any degree of openness of a city.

First, consider the public-ownership case of a single closed city, where the population of the city is fixed. The argument applies as well to an economy with many cities when the number of cities is given. The income of a household satisfies

$$P_c y = P_c w + \int_0^{\bar{x}} [R(x) - R_a]\theta(x)dx - C(X). \tag{3.14}$$

Differentiating this equation, and noting that (3.8b) holds in equilibrium, we obtain

$$P_c \frac{dy}{dx} = \int_0^{\bar{x}} \frac{dR(x)}{dX}\theta(x)dx + [R(\bar{x}) - R_a]\theta(\bar{x})\frac{d\bar{x}}{dX} - C'(x)$$

$$= \int_0^{\bar{x}} \frac{dR(x)}{dX}\theta(x)dx - C'(x). \tag{3.15}$$

Substituting (3.15) into (3.13) yields

$$\left[\int_0^{\bar{x}} \frac{1}{v_I} N(x)dx\right]\frac{du}{dx} = \int_0^{\bar{x}} \frac{u_X}{u_z} N(x)dx - C'(x), \tag{3.16}$$

which states that, if the marginal benefit exceeds the marginal cost, the utility level of residents rises as the supply of the public good is increased. At the optimum, where the utility level is maximized, we have

$$\frac{du}{dX} = 0$$

which, coupled with (3.16), yields the Samuelsonian condition (3.8a) for the optimum supply of the public good. Notice that in a closed city the land rent does not necessarily reflect the benefit of the public good.

Next, consider a small, open city. When the number of cities is very large, a city may be considered to be very small. In such a case the utility level of households can be considered as given for a city and (3.13) becomes

$$\int_0^{\bar{x}} \frac{dR(x)}{dX}\theta(x)dx = \int_0^{\bar{x}} \frac{u_X}{u_z} N(x)dx + P_c \frac{dy}{dX}.$$

Therefore, if the income level is given, an increase in land rent fully reflects the benefits of the public good:

$$\int_0^{\bar{x}} \frac{dR(x)}{dX}\theta(x)dx = \int_0^{\bar{x}} \frac{u_X}{u_z} N(x)dx. \tag{3.17}$$

There are at least two such cases. First, if land is owned by absentee landlords, the income of city residents is not affected by the supply of the public good. More important in our context is the case where a central government collects all the fiscal surpluses of city governments and distributes them as an equal subsidy. If a city is small compared to the whole economy, the policy in that city affects the subsidy received by its residents only negligibly, and the income level can be considered as fixed.

The latter case completely parallels the treatment of the extremely local public good case: if a profit-maximizing city developer, owned equally by all households in the economy, supplies the public good, the optimal supply of the public good is achieved. A city developer maximizes the profit

$$\int_0^{\bar{x}} [R(x) - R_a]\theta(x)dx - C(X)$$

with respect to X among market equilibria. Then at the maximum we have

$$\frac{d}{dx}\left\{\int_0^{\bar{x}}[R(x)-R_a]\theta(x)dx-C(X)\right\}$$
$$=\int_0^{\bar{x}}\frac{dR(x)}{dx}\theta(x)dx+\frac{d\bar{x}}{dX}[R(\bar{x})-R_a]-C'(x)$$
$$=\int_0^{\bar{x}}\frac{dR(x)}{dx}\theta(x)dx-C'(X)=0.$$

where derivates are taken across equilibria. Combining this equation with (3.17), we obtain the condition (3.8a) for the optimum supply of the public good. Therefore, a system of land developers does not require that the region be either homogeneous or physically small to achieve an optimum. We do need smallness in the sense that the utility and the income levels of residents are not affected by policies within a city.

If the number of cities is optimal, the profit of a city developer is zero from (3.9). Therefore, the zero profit condition from free entry insures the optimal number of cities. This result parallels those in the cases of increasing returns and Marshallian externality in Chapter II. The main difference is that in the public good case the supply of the public good must be determined, as well as the population of a city, while there is no such variable in previous cases.

In the case of Marshallian externality the market city tended to be too large. This problem does not appear when city formation results from the existence of public goods. Consider the utility level attainable in a city given the allocation in the rest of the world. In the Marshallian externality case the utility level first rose as the population of the city increases, reached a maximum at P_c^*, and then fell as illustrated in Figure 2a.

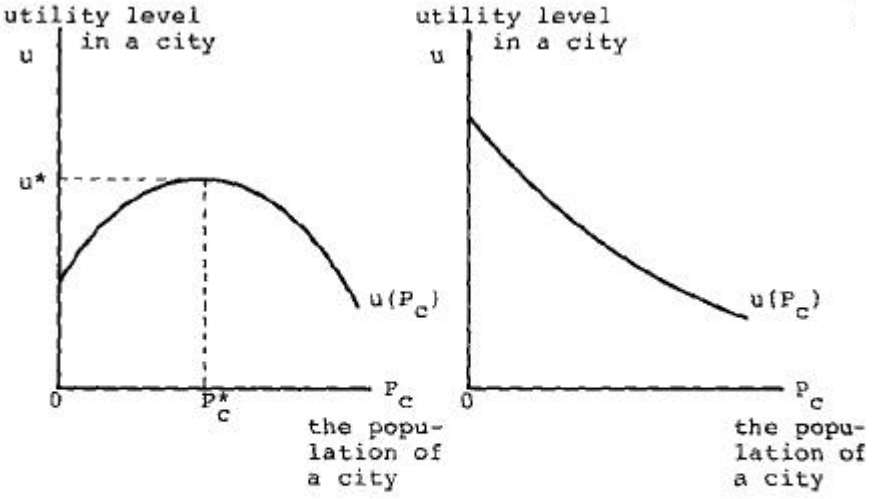

Figure 2a
The Marshallian Externality Case

Figure 2b
The Public Good Case

Figure 2. Comparison with the Marshallian Externality Case

Since the utility level was low when the population was small, it was difficult for a new small city to attract residents. In the public good case, however, the situation is different. For the same supply level of the public good the utility level achievable in a city is higher when the population of the city is smaller as illustrated in Figure 2b. Since the public good is financed by the land rent, residents do not pay any tax for the public good. The residents are therefore better off in a smaller city, since they can enjoy the same amount of the public

good with smaller average commuting costs. In such a case a new small city has no trouble attracting residents.

Notes

The analyses in this chapter derive from two separate bodies of literature. The first is concerned with attaining an efficient supply of a local public good. The second with the relationship between land rents and the benefits of public goods.

Samuelson (1954) has shown that it is extremely doubtful that any decentralized market system can determine the optimal level of a pure public good. His main argument is that there is always an incentive to misreveal one's preferences. For local public goods, however, Tiebout (1956) has argued that a decentralized market mechanism can indeed work. Freedom of personal migration among jurisdictions works as voting with one's feet which insures efficiency.

As shown elsewhere (Kanemoto (1976)), this hypothesis is not correct if local governments are passive in supplying local public goods. An argument similar to the discussion of optimum and market city sizes in the Marshallian externality case in Chapter II can be applied to show that, though the optimal; solution is one of market equilibria, there are many other equilibria, and there is no reason to believe that the optimal solution is likely to be attained.

The multiplicity of equilibria occurs since a sudden formation of a new community which is sufficiently large to be viable is usually impossible in a decentralized economy. Therefore, one way to avoid the difficulty is to allow free coalition. As shown by Pauly (1970), however, an efficient allocation is a core only if the total population is divisible by the best community size. Otherwise, a core does not exist. Furthermore, informational requirement to attain a core would be formidable.

Another way of avoiding the difficulty is to introduce an active role of local governments. McGuire (1974) and Berglas (1976) assumed a profit maximizing behaviour of the suppliers of a local public good. They showed that if there are sufficiently many suppliers, an efficient allocation of the public good is attained. For this to be true, however, a firm should be able to determine the number of the members of the club as well as the supply of the public good and the tax (or the membership charge, in their club theory terminology). Though this may be plausible in a club theory, it is usually difficult for a local government to control the population of its jurisdiction. If the population of a community is determined by free migration, the difficulty of forming a sufficiently large new community will remain to be an obstacle to achieving the efficient community size.

If there is a factor whose supply is fixed, notably land, this difficulty disappears. As a local government's policy, Margolis (1968) suggested the principle of fiscal profitability: local governments seek to minimize the burden to the local tax payers. However, he remained doubtful on the optimality of the supply of public goods in a model with the principle of fiscal profitability and voting with one's feet.

Negishi (1972) developed a formal model to analyze this problem and showed that Pareto optimality can be attained under the following three assumptions. First, the marginal rate of substitution between land and local public goods is equal to the reciprocal of the ratio of land inputs to local public goods. Second, local public goods are financed by proportional taxes on land. Third, local governments believe that marginal and average land value productivities of a public good are equal. Unfortunately, these assumptions

(especially, the first one) are quite restrictive.

We have shown that Negishi's first and second assumptions are not necessary to establish efficiency of the principle of fiscal profitability coupled with voting with one's feet, if a jurisdiction is very small relative to the whole economy.

The second source of our analysis is the literature on the relationship between land rents and the benefits of public projects. Polinsky and Shavell (1975) and Pines and Weiss (1976) showed that the marginal increase of the land rent in an open and small region correctly reflects the marginal benefit of a public project. Pines and Weiss. added a qualification: if relative prices of goods are affected by the public project (for example, in the case of leisure), this may not be true. We show in Appendix II, however, that, even if the wage rate is affected by the supply of the public good, the marginal benefit is correctly reflected in land rent. We have shown elsewhere (Kanemoto (1978)) that the conclusion holds for models which are still more general than the one used in the appendix, even when leisure is introduced.

The model of an extremely local public good is similar to models in Schuler (1974) and Helpman, Pines and Borukhov (1976).) Their main concern is, unlike ours, the spatial pattern of the supply of the local public good.

The model of a public good local to a city is similar to that of Arnott and Stiglitz (1975) who considered only the optimal allocation. They obtained the result that, in a city with the optimum population, the aggregate land rent equals the total expenditure on public goods. This result was first obtained by Flatters, Henderson, and Mieszkowski (1974) and sometimes called the Henry George Theorem or the Golden Rule. We found that this property follows from the conditions for the , optimal number of cities. It is apparent that the problem of the optimal number of cities is equivalent to the problem of the optimum population of a city in a model with identical cities.

Arnott (1979) discussed market city sizes. His approach, in contrast to ours, was to assume away entrepreneurship of city developers. He therefore repeated the argument which Henderson (1974) gave in the case of Marshallian externality and concluded that the market city size tends to be greater than the optimum.

REFERENCES

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

Arnott, R. and J.E. Stiglitz, (1975), "Aggregate Land Rents, Aggregate Transport Costs and Expenditure on Public Goods, " Discussion Paper #192, Institute for Economic Research, Queen's University.

Berglas, E., (1976), "On the Theory of Clubs, " *American Economic Review* 66, 116-121.

Clarke, E.H., (1971), "Multipart Pricing of Public Goods, " *Public Choice* 11, 17-33.

Flatters, F., V. Henderson, and P. Mieszkowski, (1974), "Public Goods, Efficiency, and Regional Fiscal Equalization, " *Journal of Public Economics* 3, 99-112.

Green, J. and J.J. Laffont, (1977), "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods, " *Econometrica* 45, 427-438.

Groves, T., (1973), "Incentives in Teams, " *Econometrica* 41, 617-631.

Groves, T. and J. Ledyard, (1977), "Optimal Allocation of Public Goods: A Solution to the 'Free-Rider' Problem, " *Econometrica* 45, 783-809.

Harris, R., (1978), "On the Choice of Large Projects, " *Canadian Journal of Economics* 11, 404-423.

Helpman, E., D. Pines and E. Borukhov, (1976), "The Interaction between Local Government and Urban Residential Location: Comment, " *American Economic Review* 66, 961-967.

Henderson, J.V., (1974), "The Sizes and Types of Cities, " *American Economic Review* 64, 637-651.

Kanemoto, Y., (1976), "A Reexamination of the Tiebout Hypothesis, " unpublished manuscript.

Kanemoto, Y., (1978), "Optimal Provision of Public Goods in a Spatial Economy, " Discussion Paper #78-45, Department of Economics, University of British Columbia.

Margolis, J., (1968), "The Demand for Urban Public Services, " in: Perloff, H.S. and L. Wingo (eds.). *Issues in Urban Economics,* (Johns Hopkins Press, Baltimore).

McGuire, M., (1974), "Group Segregation and Optimal Jurisdiction, " *Journal of Political Economy* 82, 112-132.

Negishi, T., (1972), "Public Expenditure Determined by Voting with One's Feet and Fiscal Profitability, " *Swedish Journal of Economics* 74, 452-458.

Negishi, T., (1972), *General Equilibrium Theory and International Trade,* (Amsterdam, North Holland).

Pauly, M.V., (1970), "Cores and Clubs, " *Public Choice* 9, 53-65.

Pines, D. and Y. Weiss, (1976), "Land Improvement Projects and Land Values, " *Journal of Urban Economics* 3, 1-13.

Polinsky, A.M. and S. Shavell, (1975), "The Air Pollution and Property Value Debates, " *Review of Economics and Statistics* 57, 100-104.

Samuelson, P.A., (1954), "The Pure Theory of Public Expenditures, " *Review of Economics and Statistics* 36, 387-389.

Schuler, R.E., (1974), "The Interaction between Local Government and Urban Residential Location, " *American Economic Review* 64, 682-696.

Stiglitz, J.E., (1977), "The Theory of Local Public Goods, " in: Feldstein, M.S. and P.P. Inman (eds.). *The Economics of Publi Services*, (MacMillan, London).

Tiebout, C.M., (1956), "A Pure Theory of Local Expenditures, " *Journal of Political Economy* 64, 416-424.

CHAPTER IV

TRAFFIC CONGESTION AND LAND USE FOR TRANSPORTATION: OPTIMUM AND MARKET CITIES

Traffic congestion probably induces the most important kind of externality in cities; the waste of resources and time in inefficient transportation may be enormous. In this chapter we extend the basic model to examine commuter congestion when city land is used for transportation. A new problem we face is how to allocate land between transportation and residential uses, or how much road to build.

Traffic produces a variety of externalities, noise, pollution, and risk of injury among them, which affect people whether they are traveling or not. Traffic congestion, however, tends to affect travelers most. Each additional vehicle on the road adds to the congestion and increases the travel time for others. Since the additional traveler decides to travel on the basis of her own costs, and does not have to compensate other travelers for the increased costs she imposes on them, her decision may be socially inefficient.

Decisions may be inefficient in various ways: it may be better to use less congested roads, or other modes of transportation, to travel at a less congested time, or less frequently, and so on. In this chapter, we concentrate on the distortion of residential decisions, assuming that no other decisions can be changed. Households are not charged for costs they impose on others, so they pay less than the social cost for their transportation. Since land rents reflect the differential commuting costs, they are also distorted; and the lot sizes chosen by households are therefore socially inefficient.

The obvious way of achieving an efficient allocation is to levy congestion tolls equal to the costs that a commuter imposes on others. In practice, however, it is technically and politically difficult to introduce congestion tolls.

Allocation of land between transportation and residential uses introduces another complication. Policy makers generally presume that market prices correctly reflect the social marginal value of goods. This presumption, however, is erroneous if congestion tolls are not levied. The private transportation costs are different from the social transportation costs, and market rents are not equal to social rents. The usual benefit-cost criterions based on market prices are, therefore, misleading.

In the next chapter we analyze the "second best" solution under which congestion tolls are not levied but social benefits and costs rather than market benefits and costs are used as a criterion for building roads. In this chapter, however, we compare the market allocation with the optimum allocation based on true social costs. At the optimum, congestion tolls are levied and the amount of land allocated to roads is optimal, while at the market allocation congestion tolls are not levied and roads are built according to the erroneous benefit-cost criterion based on market prices.

For the sake of simplicity we make the following (drastic) simplifying assumptions about the transportation sector.

(a) Automobiles are the only mode of transportation. Although extending the model to include alternate modes would introduce many interesting problems, the analysis would require at least another chapter.

(b) Circumferential travels are costless. This assumption allows us to maintain the one-dimensional framework. We may imagine that there are so many radial roads that any household can reach one of them with negligible costs. Miyao (1978) relaxed this assumption and considered a two-dimensional rectangular city. In order to carry out this extension, he had to simplify other aspects of the model. Although the two-dimensional case introduces the problem of route choice, we do not expect qualitatively different results.

(c) All commuters arrive at and leave the CBD at the same time, and travel at the same speed. This assumption simplifies the analysis greatly since the traffic volume at each radius can be represented by the number of workers living outside the radius. In reality, people are probably brighter than this, and try to avoid the peak time. Our assumption essentially describes the upper limit of urban congestion.

(d) There are no road construction costs. The only costs of building roads are the opportunity cost of land. This assumption can be easily relaxed.

(e) Allocation within the CBD (central business district) can be ignored. In effect, we assume that commuting costs inside the CBD are zero. This assumption was relaxed by Livesey (1973) and Sheshinski (1973) in a model simpler than ours.

(f) Time costs can be ignored. This is consistent with previous chapters, and does not affect the results.

1. The Model

Two new elements are required to extend the transportation sector of the basic model in Chapter I. First, transportation is assumed to require land. Land allocated to transportation use at radius x is denoted by $L_T(x)$. The land constraint becomes

$$L_H(x) + L_T(x) \leq \theta(x), \tag{1.1}$$

where, as in Chapter I, $L_H(x)$ denotes the amount of residential land.

We continue to ignore the allocation within the CBD and assume that the residential zone stretches from $x = 0$ to $x = \bar{x}$. We do not, however, assume that the CBD is a point. This change is made because, if $\theta(0)$ is zero or close to zero, all the available land is devoted to roads near the CBD. In such a case, the nonnegativity constraint, $L_H(x) \geq 0$, for $0 \leq x \leq \bar{x}$, is binding, and we obtain a corner solution. For simplicity, we assume that enough land is available near the center to preclude the corner solution.

The second new element is traffic congestion: commuting costs for each individual depend on the number of others using the same road at the same time. Specifically, it is assumed that the commuting cost per mile per household at radius x is a function of the volume of traffic $T(x)$ and the amount of land allocated for transportation $L_T(x)$ at that radius:

$$g(T(x), L_T(x)), \tag{1.2}$$

where the cost increases as the volume increases,

$$g_T(T, L_T) > 0\,, \tag{1.3}$$

and decreases as more land is used as roads,

$$g_L(T, L_T) < 0\,. \tag{1.4}$$

We concentrate on the *total* amount of land used for roads at each radius, and do not analyze how the width of an individual road is determined. In this chapter the width of the road refers to the total amount of land used for transportation at a radius, instead of the width of an individual road.

Commuting costs incurred by a household living at x are

$$t(x) = \int_0^x g(T(x'), L_T(x'))dx'\,. \tag{1.5}$$

Differentiation of this equation with respect to x yields the following differential equation:

$$t'(x) = g(T(x), L_T(x))\,. \tag{1.6}$$

This differential equation, with the boundary condition at $x = 0$,

$$t(0) = 0\ \ , \tag{1.7}$$

is equivalent to (1.5).

Since all commuters arrive at and leave the CBD at the same time, and that they travel at the same speed, the traffic volume at a radius x is equal to the number of households living outside *x*:

$$T(x) = \int_x^{\bar{x}} N(x')dx'\,, \tag{1.8}$$

where *N(x)dx* is the number of households living between *x* and *x*+*dx*, and is given by (I.1.25): $N(x) = L_H(x)/h(x)$. This is equivalent to the differential equation,

$$T'(x) = -L_H(x)/h(x)\,, \tag{1.9}$$

with the boundary condition,

$$T(\bar{x}) = 0\,. \tag{1.10}$$

t(*x*) is usually called the *private transportation cost* and is different from the *social transportation cost*, since it does not include the external costs imposed on other commuters. The social transportation cost, $G(T, L_T)$, at radius x is an increase in the total transportation cost there, $Tg(T, L_T)$, caused by a marginal increase in traffic:

$$G(T, L_T) \equiv \partial\left[Tg(T, L_T)\right]/\partial T$$

$$= g(T, L_T) + Tg_T(T, L_T)\,. \tag{1.11}$$

An additional car on the road causes more congestion and increases the transportation costs of other commuters by g_T. Since there are *T* cars on the road, the total increase in the

costs for other travelers is Tg_T. This external cost must be added to the private transportation cost, g. In the transportation economics literature, the private transportation cost is sometimes called the *average transportation cost*, and the social transportation cost is called the *marginal transportation cost* for the obvious reason.

2. A Closed City

In a closed city the population of the city, P, is given. The population constraint (I.1.24) gives the boundary condition for (1.9) at $x = 0$:

$$T(0) = P. \tag{2.1}$$

To save space we analyze only the public-ownership case, which is slightly simpler than the absentee-landlord case. In our version of public ownership the resource constraint is

$$\int_0^{\bar{x}} \left[zL_H / h + Tg(T, L_T) + R_a\theta \right] dx \le Pw . \tag{2.2}$$

The available resource, Pw, is spent on the consumer good, zL_H / h, commuting costs, Tg, and the rural rent, $R_a\theta$. This constraint is different from the constraint (I.1.30) in Chapter I in the following two respects. First, equality is replaced by inequality. This does not change the conclusions because the constraint holds with equality both at the optimum and in market equilibrium. For technical reasons, the inequality constraint is more convenient in this and the next chapters, since the associated Lagrange multiplier can be signed. Second, the transportation cost, tN, is replaced by Tg. Equivalence of these two formulations can be easily seen by integration by parts.

2.1. The Optimum City

In the optimum city the sum of utilities,

$$\int_0^{\bar{x}} \left[uL_H(x) / h(x) \right] dx , \tag{2.3}$$

is maximized under the constraints (1.1), (1.9), (1.10), (2.1), (2.2), and the equal-utility constraint,

$$u(z(x), h(x)) = u , \qquad 0 \le x \le \bar{x} . \tag{2.4}$$

The control variables are the consumptions of the consumer good and housing, $z(x)$ and $h(x)$, and the total widths of the road and the residential area, $L_T(x)$ and $L_H(x)$; and the control parameters are the utility level, u, and the physical city size, $\bar{x}$.

The Theorem of Hestenes, which is stated in the appendix on optimal control theory, can be applied to this problem. The Hamiltonian is

$$\phi = \left[u - \lambda(x) \right] L_H(x) / h(x) - \delta \left[z(x) L_H(x) / h(x) + T(x) g(T(x), L_T(x)) + R_a\theta(x) \right], \tag{2.5}$$

where $\lambda(x)$ and δ are adjoint variables associated with the differential equation (1.9)

and the isoperimetric constraint (2.2) respectively. δ is a constant, and satisfies

$$\delta\left\{Pw-\int_0^{\bar{x}}\left[zL_H/h+Tg+R_a\theta\right]dx\right\}=0,\qquad \delta\geq 0. \tag{2.6}$$

As in previous chapters, δ can be interpreted as the shadow price of the consumer good in utility terms. The ad joint variable, $\lambda(x)$, satisfies the adjoint equation

$$\frac{\partial\phi}{\partial T}=-\lambda'(x)=-\delta\left[g(T,L_T)+Tg_T(T,L_T)\right] \tag{2.7}$$

The second equality shows that $\lambda'(x)$ equals the shadow price of the consumer good times the social transportation cost at *x* defined in section 1. Thus $\lambda'(x)$ is the social cost of transportation in utility terms. The first equality confirms this interpretation, since it says that a marginal increase in traffic at radius *x* decreases the sum of utilities by $\lambda'(x)$.

According to the maximum principle, the Hamiltonian must be maximized under the constraints (1.1) and (2.4). The Lagrangian for this problem is

$$\psi=\phi+\nu(x)\left[u(z(x),h(x))-u\right]+\propto(x)\left[\theta(x)-L_H(x)-L_T(x)\right]\ , \tag{2.8}$$

where $\nu(x)$ and $\propto(x)$ are respectively the Lagrange multipliers for the constraints (2.4) and (1.1). The first order conditions

are

$$\frac{\partial\psi}{\partial L_H}=\frac{u-\lambda(x)-\delta z(x)}{h(x)}-\propto(x)=0 \tag{2.9a}$$

$$\frac{\partial\psi}{\partial L_T}=-\delta Tg_L(T,L_T)-\propto(x)=0 \tag{2.9b}$$

$$\frac{\partial\psi}{\partial h}=-\frac{u-\lambda(x)-\delta z(x)}{h(x)}N(x)-\nu(x)u_h=0 \tag{2.9c}$$

$$\frac{\partial\psi}{\partial z}=-\nu(x)u_z-\delta N(x)=0\ . \tag{2.9d}$$

$\propto(x)$ must satisfy the condition that

$$\propto(x)\left[\theta(x)-L_H(x)-L_T(x)\right]=0,\qquad \propto(x)\geq 0, \tag{2.10}$$

and can be interpreted as the shadow rent of land in utility terms.

The transversality condition for $\bar{x}$ is

$$\psi(\bar{x})=\left[u-\lambda(\bar{x})-\delta z(\bar{x})\right]L_H(\bar{x})/h(\bar{x})-\delta\left[T(\bar{x})g(T(\bar{x}),L_T(\bar{x}))+R_a\theta(\bar{x})\right]=0, \tag{2.11}$$

which simply says that the city should not extend beyond the point where the marginal social contribution of developing an additional unit of land is zero. The transversality condition for *u*,

$$\int_0^{\bar{x}} N(x)dx = \int_0^{\bar{x}} \nu(x)dx \tag{2.12}$$

is the same as (I.2.22e) and has the same interpretation.

To simplify the interpretations of the optimality conditions, it is convenient to recast shadow prices in terms of the consumer good. Define

$$\tau(x) \equiv \frac{1}{\delta}\left[\lambda(x) - \lambda(0)\right]. \tag{2.13}$$

Then $\tau(x)$ satisfies both

$$\tau(0) = 0 \tag{2.14}$$

and

$$\tau'(x) = g(T, L_T) + Tg_T(T, L_T), \tag{2.15}$$

and can be interpreted as the social transportation cost of commuting from radius x to the center. Similarly, the social rent at x is

$$R(x) \equiv \propto(x)/\delta. \tag{2.16}$$

Equation (2.9a) may be interpreted as the optimal household budget. We can rewrite (2.9a) as

$$u = \delta z(x) + \propto(x)h(x) + \lambda(x).$$

Dividing through by δ, defining

$$y \equiv \left[u - \lambda(0)\right]/\delta, \tag{2.17}$$

and using (2.13) and (2.16), we obtain

$$y = z(x) + R(x)h(x) + \tau(x). \tag{2.18}$$

This equation expresses the socially optimal allocation of household income at x if y is the income, $R(x)$ the market rent, and $\tau(x)$ the commuting costs. Then, by (2.15), households must pay the social transportation costs, or the private transportation costs plus the costs of externalities imposed on other travelers. In other words, some way must be found to collect a congestion toll if the price system is to achieve the optimum city.

Notice that in this simple model congestion tolls can be levied according to the location of residence. A household living at x should pay the amount

$$\int_0^x T(x')g_T(T(x'), L_T(x'))dx',$$

of congestion tolls. However, this kind of distance tax is not optimal in a more general model in which households can choose when to travel or the best mode among several modes of transportation.

Rewriting Equation (2.9b) as we did (2.9a), we obtain

$$-Tg_L(T, L_T) = R(x). \tag{2.19}$$

Now since $-Tg_L(T, L_T)$ is simply the marginal reduction in total transportation costs from widening the road at x with the traffic volume fixed, (2.19) reveals that at the optimum the marginal reduction in transportation costs from widening the road equals the shadow rent. For later use, we define $-Tg_L(T, L_T)$ as the market benefit, $B(x)$, of widening the road:

$$B(x) \equiv -T(x)g_L(T(x), L_T(x)) . \tag{2.20}$$

Combining (2.9c) and (2.9d), and solving for u_h / u_z yields

$$\frac{u_h}{u_z} = R(x) , \tag{2.21}$$

which says that the marginal rate of substitution between land and the consumer good equals the shadow rent at the optimum. This condition is obtained if a household maximizes utility and pays the congestion tax, and therefore allocates its budget according to (2.18).

The transversality condition (2.11) becomes[1]

$$R(\bar{x}) = R_{a'} \tag{2.22}$$

that is, the urban rent at the edge of the city equals the rural rent.

Thus the optimum solution can be attained by the decentralized market mechanism if three conditions are met: all households are given the equal income y; congestion tolls equal to the external costs, Tg_T, are levied at each x; and roads are built according to the benefit-cost criterion, equating the marginal reduction of transportation costs from widening the road to the market rent.

Note that the marginal benefits of the road are given by the marginal reduction of transportation costs with the volume of traffic fixed. This is true even though the construction of a new road changes the allocation of the entire economy: the change in commuting costs induces a change in land rent, and hence in the consumption decisions of households, which changes the residential structure of the city. Due to the envelope property, all the indirect effects cancel out each other and the benefits are simply given by the direct saving in transportation cost.[2] This is a general property of the first best optimum. As will be shown in the next chapter, however, the effects of induced changes do not wash out in the second best world where congestion tolls are not allowed.

When we consider the relationship between the total congestion tolls and the total land rent of the road, one of the standard results from production theory is obtained: profit is negative under marginal cost pricing when there are increasing returns to scale, zero in the constant returns case, and positive in the decreasing returns case. First, consider the case where transportation technology exhibits constant returns to scale: the average transportation cost, g, remains the same if the volume of traffic, T, and the width of the road, L_T, are increased with the same proportion. In the constant-returns-to-scale case, $g(T, L_T)$ can

---

[1] We have been able to prove this only in the case where $g(T, L_T)$ is finite at $\bar{x}$ and $g_L(0, L_T) < \infty$ for all $L_T > 0$. In that case $T(\bar{x})g(T(\bar{x}), L_T(\bar{x})) = 0$, and $L_H(\bar{x}) = \theta(\bar{x})$, and hence (2.11) implies $R(\bar{x}) = R_a$. The first equality is obvious from $T(\bar{x}) = 0$. The second equality is obtained since otherwise $R(\bar{x})$ is zero from (2.19) and $T(\bar{x}) = 0$. From (2.11) and (2.9a), this implies that $R_a\theta(\bar{x}) = 0$, which cannot happen if $R_a > 0$ and $\theta(\bar{x}) = 0$.

[2] Wheaton (1977) and Arnott (1976) observed this well-known result in the context of urban land use.

be written as

$$g(T, L_T) = \tilde{g}(T / L_T).$$

Then, the total congestion tolls at x are

$$T^2 g_T = (T(x)^2 / L_T(x))\tilde{g}, (T(x) / L_T(x)),$$

and the total land rent at x is

$$-TL_T g_L = (T(x)^2 / L_T(x))\tilde{g}'(T(x) / L_T(x)).$$

Thus the congestion tolls exactly cover the land rent of the road at each radius.

A proportionate increase of T and L_T decreases the average transportation cost, g, in the case of increasing returns to scale, and increases in the case of decreasing returns to scale. Therefore, for a change in T and L_T satisfying

$$\frac{dT}{T} = \frac{dL_T}{L_T},$$

the corresponding change in g,

$$\begin{aligned} dg &= g_T dT + g_L dL_T \\ &= (Tg_T + L_T g_L) L_T dL_T, \end{aligned}$$

is negative in the case of increasing returns and positive in the case of decreasing returns. Thus the total congestion tolls are less than the total land rent in the increasing returns case,

$$T^2 g_T < -TL_T g_L,$$

and greater in the decreasing returns case,

$$T^2 g_T > -TL_T g_L.$$

If the transportation authority pays for land rent of the road and collects congestion tolls, its budget is balanced in the constant returns case, it makes a profit in the decreasing returns case, and suffers a loss in the increasing returns case, which is analogous to the results in the usual production theory.

2.2. The Market City

Let us consider an allocation where the congestion tolls cannot be levied. The width of the road is not determined by the market, but by the benefit-cost criterion based on market prices (to be explained below).

When congestion tolls are not levied, households pay only the private transportation costs, $t(x)$, given by (1.6) and (1.7). Assuming that all households receive the same income y, we obtain the budget constraint (I.1.3),

$$y = z(x) + R(x)h(x) + t(x) \qquad 0 \le x \le \bar{x},$$

where $R(x)$ is land rent. The first order condition for utility maximization is given by (I.1.4):

$$\frac{u_h}{u_Z} = R(x).$$

The utility level must be the same everywhere in the city because of spatial arbitrage. This is equivalent to the condition

$$h(x)R'(x) + t'(x) = 0, \qquad 0 \le x \le \bar{x}, \tag{2.23}$$

which is obtained from (I.1.16).

Since the residential rent is equal to the rural rent at the edge of the city, we have (I.1.13):

$$R(\bar{x}) = R_a.$$

In the public-ownership case on which we shall concentrate, the differential rent is returned to residents. Thus the income level is given by

$$y = w + \frac{1}{P}\{\int_0^{\bar{x}} R(x)L_H(x)dx - \int_0^{\bar{x}} R_a\theta(x)dx\}. \tag{2.24}$$

It is easy to see that this is equivalent to (2.2) with equality.

It is assumed that the (erroneous) benefit-cost criterion based on market prices is adopted to determine the allocation of land between housing and transportation uses. Roads are widened until the market benefit equals the market rent. The market benefit, $B(x)$, is the reduction of transportation costs from a marginal increase in land used for roads, which is given by (2.20). Then we have

$$-T(x)g_L(T(x), L_T(x)) = R(x) \tag{2.25}$$

in equilibrium. Note that this is the same as the benefit-cost criterion (2.19) adopted in the optimum city. Although this naive benefit-cost criterion leads to the optimum allocation of land when congestion tolls are levied, it is no longer optimal in the absence of congestion tolls.

Since no available land is left vacant unless the rent is zero, (1.1) holds with equality:

$$L_H(x) + L_T(x) = \theta(x), \qquad 0 \le x \le \bar{x}. \tag{2.26}$$

Comparing these equations with those obtained in the optimum city, we can see that the only difference lies in transportation costs. In the market city residents pay the private (or average) transportation cost, while in the optimum city they also pay congestion tolls, which make up the difference between the private and social (or marginal) transportation cost.

2.3. Comparison Between the Optimum and Market Cities

In this section the optimum and market cities characterized in the previous sections are

compared. Unfortunately, the complexity of the model prevents us from carrying out the comparison in the general case. We, therefore, calculate numerical examples using the Cobb-Douglas type utility function.

$$u = h^{\alpha} z^{1-\alpha}, \tag{2.27}$$

and the Vickrey type transportation cost function without a constant term,

$$t'(x) = g(T/L_T)^K, \tag{2.28}$$

where g and k are positive constants.

These functions are chosen for the convenience of computation and are not quite realistic. The properties of the functions are as follows. The Cobb-Douglas utility function (2.27) implies that the proportion of income net of commuting costs spent on land is always α. In other words, the income elasticity of demand for land is one and the price elasticity is minus one. The transportation cost function (2.28) represents constant returns to scale in transportation technology. Since there is no constant term, transportation costs are zero when there is no other car on the road. Transportation costs rise when the traffic density, T/L_T, or the number of travelers per unit width of the road, rises. The elasticity of transportation costs with respect to traffic density is *k* and constant.

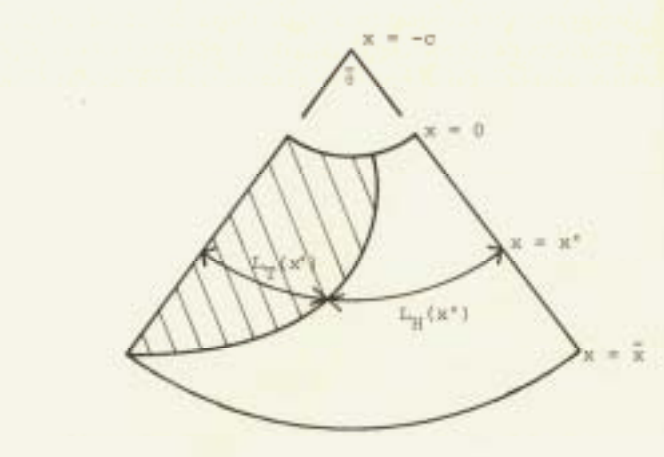

Figure 1. The spatial configuration of the city

The city is assumed to be circular, although not necessarily a complete circle. Since commuting costs in the CBD are zero by assumption (e), we need only consider the residential zone, where the supply of land is

$$\theta(x) = \bar{\theta}(x + c), \tag{2.29}$$

with positive constants $\bar{\theta}$ and *c*. The constant *c* is chosen so that roads do not cover all the land at $x = 0$. In the numerical calculations, $\bar{\theta} = 2$ and $c = 50$.

The results of calculations are shown on Tables 1 and 2.[3] In Table 1, *k* is assumed to be 1 and *g* to be 10^{-5}. α is assumed to be 0.2, which means that a fifth of the income net of

[3] For the details of calculations, see the Appendix to Chapter V, Part I of Kanemoto (1977).

transportation costs is spent on land. It should be remembered that actual housing is included in the consumer good. The number of households in the city is 100,000, and 1 unit of resources expressed in terms of the consumer good is available for each household. The rural rent is 1 per unit of land.

Table 1

Comparison between Optimum and Market Cities: $k = 1$

	Optimum	Market
Rent at radius *0* $(R(0))$	31.3	14.9
Income per household (y)	1.30	1.03
City size $(\bar{x})$	94.29	120.3
Utility level (u)	0.3955	0.3640
Total area $(\cdot\ 10^3)$	6.39	12.0
Total area of roads $(\cdot\ 10^3)$	2.14	5.88
Total rent $(\cdot\ 10^4)$	1.91	1.50
Total transport costs $(\cdot\ 10^4)$	1.72	2.79
$g = 10^{-5}, k = 1.0, w = 1, P = 100{,}000, R_a = 1.0$ $\alpha = 0.2, \theta(x) = 2(x + 50)$		

Table 2

Comparison between Optimum and Market Cities: $k = 2$

	Optimum	Market
Rent at radius *0* $(R(0))$	18.9	5.22
Income per household (y)	1.33	0.83
City size $(\bar{x})$	123.0	177.7
Utility level (u)	0.4450	0.3619
Total area $(\cdot\ 10^4)$	1.26	2.91
Total area of roads $(\cdot\ 10^3)$	5.10	18.5
Total rent $(\cdot\ 10^4)$	1.85	1.22
Total transport costs $(\cdot\ 10^4)$	1.34	2.23
$g = 0.5 \cdot\ 10^{-8}, k = 2.0, w = 1, P = 100{,}000, R_a = 1.0$ $\alpha = 0.2, \theta(x) = 2(x + 50)$		

There is a striking difference in physical city size between the optimum and market cities: the length of the residential zone of the optimum city is just over three quarters of that of the market city, and the total area of the residential zone (including the road) is just over a half. Because congestion tolls are levied in the optimum city, the land rent tends to be higher and consequently the optimum city is denser than the market city.

The rent at $x = 0$ in the optimum city is more than twice as high as that in the market city, and the total land rent of the residential land in the optimum city is greater than that of the market city even though the market city is considerably bigger. The total rent is 19.1% of the total available resources in the optimum city and 15% in the market city.

In the optimum city the total transportation cost not including congestion tolls is about 62% of those in the market city. Transportation costs constitute 17.2% of the total available resources in the optimum city and 27.9% in the market city. Thus the absence of congestion tolls results in the excessive use of resources in transportation. Since $k = 1$, congestion tolls in the optimum city equal the private transportation cost. This means that when congestion tolls are included, the total commuting costs paid by households are twice as much as the total transportation costs calculated in Table 1. Therefore, although less resources are devoted to transportation in the optimum city than in the market city, households pay more commuting costs in the optimum city if we include congestion tolls. Of course, the revenue from congestion tolls is returned to the city residents in our model, and congestion tolls do not represent any consumption of resources.

The total land allocated to housing is greater in the market city. On the average, therefore, residents in the optimum city consume less land. Notice, however, that housing consumption need not decrease because it is a part of the composite consumer good. Since the total transportation costs (excluding congestion tolls) are smaller in the optimum city, the total consumption of the consumer good is greater. This overwhelms the decrease of the consumption of land and the utility level is higher in the optimum city. Thus the main advantage of the optimum city lies in the fact that the total transportation costs are reduced through dense habitation.

Notice that household income y is 1.3 although we assumed that only one unit of the consumer good was available to each household. The difference is the average expenditure on rent and congestion tolls which is returned to city residents in the public-ownership case.

The road width functions are plotted in Figure 2. The superscripts ^{0} and ^{m} denote respectively the optimum and market solutions. The road in the market city is wider than that

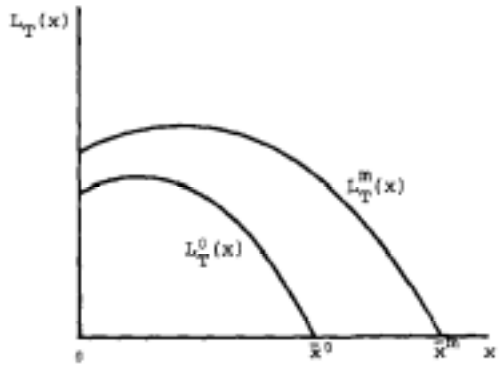

Figure 2
Optimum and Market Road Width Functions: A Closed City

in the optimum city everywhere in the city. In this sense, the benefit-cost criterion based on market prices has a tendency to overinvest in roads. The ratio between the width of the road and the available land is plotted in Figure 3. In both optimum and market cities the ratio decreases monotonically with distance from the center.

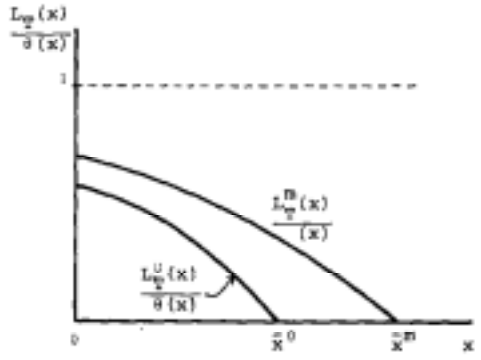

Figure 3
The Proportion of Land Devoted to Roads: A Closed City

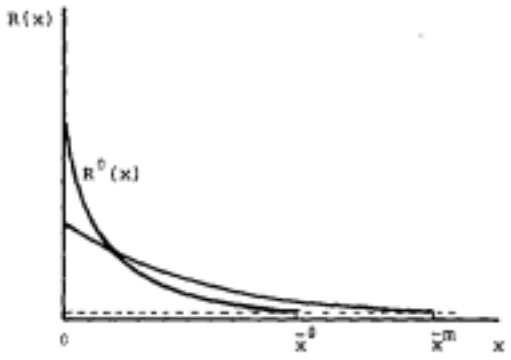

Figure 4
Optimum and Market Rent Profiles: A Closed City

The rent function is plotted in Figure 4. The rent is higher in the optimum city than in the market city near the center but lower near the edge.

As shown in Figure 5, near the CBD the traffic density is higher in the optimum city, which reflects the fact that the road is narrower in the optimum city. Near the edge of the city, however, the traffic density is higher in the market city even though the market city has the wider road, because the optimum city has fewer commuters near the edge simply because the optimum city is smaller.

Table 2 shows the results of the case of $g = 0.5 \cdot 10^{-8}$, $k = 2$ and $\alpha = 0.2$. The assumption of $k = 2$ implies more acute congestion than in the previous case. This is the reason why the difference in the utility level is greater here. All the qualitative results are the same, however.

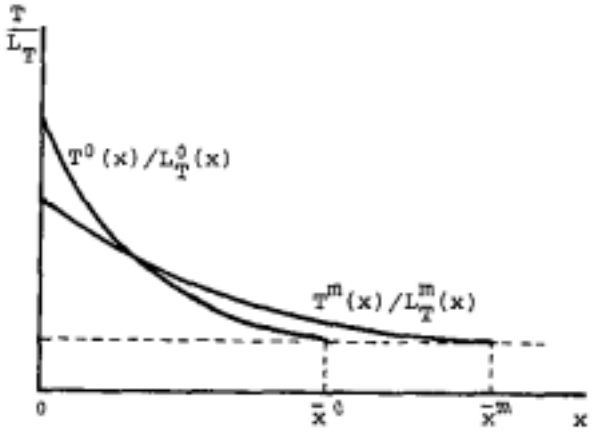

Figure 5
Optimum and Market Traffic Density Functions: A Closed City

3. An Open City

In this section we consider a small, open city where the utility level in the city must equal the level outside the city. This case would be relevant when a planner of a small city is contemplating a long-run policy. In order to isolate problems pertaining to traffic congestion from others, we assume constant returns to scale in production: the aggregate production function of the city is

$$wP,$$

with a constant w. The analysis can be easily extended to the case where the aggregate production function of a city exhibits increasing or decreasing returns to scale.

In a small, open city, the utility level of residents is given and the population of the city becomes an endogenous variable. Therefore, (2.1) is replaced by

$$u(z(x), h(x)) = \bar{u}\,, \tag{3.1}$$

where $\bar{u}$ is the utility level given for the city.

Only the absentee-landlord case is considered in this section. Absentee landlords receive congestion tolls as well as land rent. The income of a household, then, is given by w, and the resource constraint (2.2) no longer holds.

3.1 The Optimum City

As in Chapter I, we maximize the net product of the city after the cost of maintaining the given utility level of residents. Thus our problem is one of maximizing

$$\int_0^{\bar{x}} \{[w - z(x) - t(x)]L_H(x)/h(x) - R_a\theta(x)\}dx\,, \tag{3.2}$$

subject to the constraints (1.1), (1.6), (1.7), (1.9), (1.10), and (3.1). The Hamiltonian is

$$\Phi = [w - z(x) - t(x)]\frac{L_H(x)}{h(x)} - R_a\theta(x) - \lambda(x)\frac{L_H(x)}{h(x)} + \eta(x)g(T(x), L_T(x))\,, \tag{3.3}$$

and the Lagrangian for the problem of maximizing the Hamiltonian under the constraints (1.1) and (3.1) is

$$\Psi = \Phi + \nu(x)[u(z(x), h(x)) - \bar{u}] + \propto(x)[\theta(x) - L_H(x) - L_T(x)], \tag{3.4}$$

where $\lambda(x)$ and $\eta(x)$ are respectively adjoint variables associated with differential equations (1.9) and (1.6). $\nu(x)$ and $\propto(x)$ are Lagrange multipliers for (3.1) and (1.1).

If we define $R(x) = \propto(x)$ and $\tau(x) = t(x) + \lambda(x)$, the first order conditions become, after simple manipulations:

$$\tau'(x) = g + Tg_T \tag{3.5}$$

$$w = z + Rh + \tau \tag{3.6}$$

$$-Tg_{L_T}(T, L_T) = R(x) \tag{3.7}$$

$$\frac{u_h}{u_z} = R(x) \tag{3.8}$$

$$\tau(0) = 0 \tag{3.9}$$

$$R(\bar{x}) = R_a\,. \tag{3.10}$$

It can be seen immediately that these equations coincide with (2.15), (2.18), (2.19), (2.21), (2.14) and (2.22) obtained in a closed city if w is replaced by y. Therefore, the difference between open and closed cities lies in the determination of income and utility levels. In a closed city these two variables are determined so as to satisfy the population constraint (2.1) and the resource constraint (2.2), whereas in an open city the utility level is given from outside, and the income level is equal to the marginal productivity of labor, which is also assumed to be given. Note that neither land rent nor congestion tolls are returned to residents in this section.

If the production sector is competitive, the wage will be equal to the marginal productivity of labor. Therefore, if the land is owned by absentee landlords, the optimal

solution can be obtained by levying congestion tolls and by constructing roads so as to equate the market rent to the marginal benefit from widening the road.

3.2 The Market City

In the market city, it is assumed that production is carried out competitively. Since we consider only the absentee landlord case, the income of residents is the competitive wage w. In the absence of congestion tolls, the commuting costs are given by $t(x)$. When the market rent is given by $R(x)$, a household faces the budget constraint,

$$w = z(x) + R(x)h(x) + t(x).$$

A household maximizes the utility level under this budget constraint, which yields the first order condition:

$$\frac{u_h}{u_z} = R(x).$$

The maximized utility level must be equal to the given utility level, $\bar{u}$, everywhere in the city.

Roads are built according to the benefit-cost criterion based on market prices:

$$-Tg_L(T, L_T) = R(x) \quad .$$

At the edge of the city, the market rent equals the rural rent:

$$R(\bar{x}) = R_a .$$

Again, we can observe that the only differences between closed and open cities are boundary conditions which determine the income level and the utility level: in a closed city the income level is given by (2.23) and the population constraint (2.1) must hold, but in an open city the utility level is fixed at $\bar{u}$ and the income level is also a constant w.

3.3. Comparison Between Optimum and Market Cities

The optimum and market cities obtained in the previous sections are compared.

First, since $\tau(0)$ and $t(0)$ are both zero, households at $x = 0$ face the same budget constraint,

$$w = z + R(0)h ,$$

in both optimum and market cities. In order for the utility levels to be the same, the rents at $x = 0$ must be the same in the two cities:

$$R^0(0) = R^m(0) ,$$

where superscripts ^{0} and ^{m} respectively denote optimum and market solutions.

Since congestion tolls are levied in the optimum city, it is expected that households pay more transportation costs in the optimum city. If this is true, the rent function has a steeper slope in the optimum city than in the market city and the land rent in the optimum city is lower than that in the market city everywhere in the city except at $x = 0$ where they

are equal. Though we have not been able to show this in a general case, it is true if we assume the Cobb-Douglas type utility function (2.27) and the Vickrey type transportation cost function (2.28).[4] The rent profiles in this case are depicted in Figure 6.

The traffic density has the same pattern as the rent function. At $x = 0$, optimum and market cities have the same traffic densities and in the rest of the city the market city has a higher traffic density.

The market city size is bigger than the optimum city size. It can be shown that, in the case of the Cobb-Douglas utility function and the Vickrey transportation cost function, the residential zone is exactly $k+1$ times longer in the market city than in the optimum city. In a closed city, the difference between market and optimum city sizes is not as large as in an open city. This is because the population is fixed in a closed city and the rent at the CBD and the income level are both higher in the optimum city.

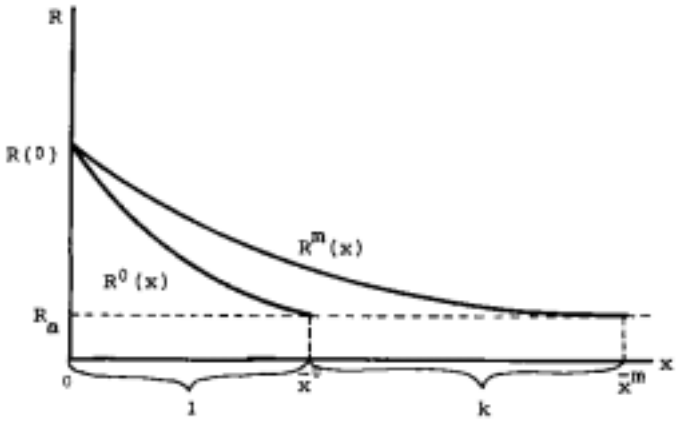

Figure 6
Rent Profiles in Market and Optimum Cities: An Open City

The widths of roads have been calculated for a variety of parameters. In most cases the market city has a wider road than the optimum city though we have found some exceptional cases where the optimum city has a wider road near the center. However, even in such cases the ratio between the total area of roads and the total area of residential land is greater in the market city. One example in which $w = 1, k = 1, h = 10^{-5}$ and $u = -0.364$ is plotted in Figure 7. In this case roads are wider everywhere in the market city.

[4] See Chapter V, Part I of Kanemoto (1977).

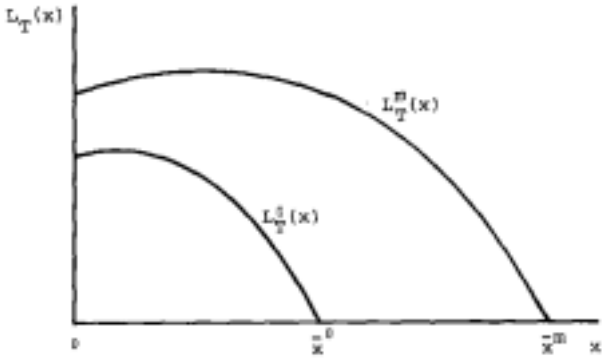

Figure 7
Road Width Functions in Market and Optimum Cities:
An Open City

4. An Economy Consisting of Many Cities

In this section we briefly consider an economy consisting of many cities, under the assumption that the number of cities is a variable. Only the optimum allocation is analysed since the market city size may be indeterminate as shown in Chapter II.

For simplicity, we assume that no one lives in the rural area and that all cities are identical. Then the population of the economy, P, the population of a city, P, and the number of cities, n, must satisfy the relationship

$$P = nP_c . \tag{4.1}$$

The boundary condition for $T(x)$ at $x = 0$ is now

$$T(0) = P_c . \tag{4.2}$$

The aggregate production function of a city is

$$F(P_c), \tag{4.3}$$

and we assume increasing returns to scale. The resource constraint (2.2) must be rewritten as

$$\int_0^{\bar{x}} [zL_H / h + Tg(T, L_T) + R_a\theta]dx \le F(P_c) . \tag{4.4}$$

Now, our problem is one of maximizing the common utility level, u, subject to the resource constraint (4.4), the traffic flow constraint (1.9), the equal utility constraint (2.4), the land constraint (1.1), the population constraint (4.1), and the boundary conditions for $T(x)$, (4.2) and (1.10). The Lagrangian for this problem is

$$\Lambda = u + \delta n\left[F(P_c) - \int_0^{\bar{x}} (zL_H/h + Tg + R_a\theta)dx \right]$$
$$+ n\int_0^{\bar{x}} \lambda(x)[-L_H/h - T'(x)]dx$$
$$+ n\int_0^{\bar{x}} \nu(x)[u(z,h) - u]dx + n\int_0^{\bar{x}} \propto(x)(\theta - L_H - L_T)dx$$
$$+ \gamma(nP_c - P) + \varepsilon n[P_c - T(0)]. \qquad (4.5)$$

If we define

$$\tau(x) \equiv \frac{1}{\delta}(\lambda(x) - \lambda(0)) \qquad (4.6)$$

and

$$R(x) \equiv \propto(x)/\delta\,, \qquad (4.7)$$

the first order conditions become, after some rearrangements,

$$F'(Pc) = z(x) + R(x)h(x) + \tau(x) \qquad 0 \le x \le \bar{x}\,, \qquad (4.8)$$

and (2.14), (2.15), (2.19), (2.21), and (2.22). The only new condition is (4.8), which means that a worker is paid the value of marginal productivity of labor.

The condition can be related to the results in Chapter II. Multiplying (4.8) by $N(x)$ and integrating from 0 to $\bar{x}$ yields

$$P_c F'(P_c) = \int_0^{\bar{x}} [z(x)N(x) + \tau(x)N(x) + R(x)L_H(x)]dx\,. \qquad (4.9)$$

The resource constraint (4.4) holds with equality, and the total transportation costs are the same regardless of how costs at different radii are added,

$$\int_0^{\bar{x}} Tgdx = \int_0^{\bar{x}} tNdx\,.$$

We therefore have

$$F(P_c) = \int_0^{\bar{x}} \left[z(x)N(x) + t(x)N(x) + R_a\theta(x)\right]dx\,. \qquad (4.10)$$

Subtracting (4.8) from (4.9) yields

$$\left[F(P_c) - P_c F'(P_c)\right] + \int_0^{\bar{x}} \left[R(x)L_H(x) - R_a\theta(x)\right]dx$$
$$+ \int_0^{\bar{x}} \left[\tau(x) - t(x)\right]dx = 0. \qquad (4.11)$$

The first square bracket is the profit from production, which is negative because of increasing returns to scale; the second term is the net rent revenue after the payment of the rural rent; and the third term the total congestion toll. Thus the loss incurred by a producer equals the sum of the net rent revenue and the total congestion toll. This is more general

than the result in Chapter II, which states that the operating loss of a producer equals the total differential rent, or the market rent minus the rural rent. Notice also the similarity with the results of Section 2 of Chapter III which considers congestion in the consumption of local public goods.

If transportation technology has constant returns to scale, the total congestion toll equals the land rent on the road. In this case (4.11) is equivalent to the result obtained in Chapter II,

$$\left[F(P_c) - P_c F'(P_c)\right] + \int_0^{\bar{x}} \left[R(x) - R_a\right]\theta(x)dx = 0: \qquad (4.12)$$

the loss of a producer equals the total differential rent.

Notes

Traffic congestion has usually been analyzed in nonspatial frameworks. Strotz (1965) extended the usual analysis to a spatial model in which a city is divided into a finite number of homogeneous rings. He characterized the optimal solution and showed that the optimal solution requires congestion tolls. He also showed that congestion tolls exceed or less than expenditure on roads if transportation technology has decreasing or increasing returns to scale respectively.

Solow and Vickrey (1971) formulated a model of a long narrow city in which distance is a continuous variable. They solved the problem of minimizing transportation costs using calculus of variation. Mills and de Ferranti (1971) consider a similar transportation-cost-minimization problem in a circular city Livesey (1973) and Sheshinski (1973) extended their model to analyze land use within the CBD. Legey, Ripper and Varaiya (1973) extended the model to include capital. They also introduced the market city where roads are built according to the benefit-cost criterion based on market prices and compared optimum and market cities. It was shown that the market city is more dispersed than the optimum city.

All these papers considered closed cities where the tot product (or the total population) of the city was fixed. Kanemoto (1975) introduced an open city where the city faces fixed export price.

None of the above models allow for substitution between land and other factors. Dixit (1973), Oron, Pines and Sheshinski (1973) and Riley (1974) considered traffic congest in a model which allows for the choice of housing lot size and therefore substitution between land and other goods. Indepedent of our work, Robson (1976) compared optimum and market cities in the same model as ours. He considered the case of $\alpha = 1/2$ in the utility function (2.27). Though calculations are easiest in this case, the assumption implies that all households spend half of their incomes-after-commuting-costs on land which is quite unrealistic.

Kanemoto (1976) considered a production city with substitutability between labor and land in an open city framework. The results are parallel to those in section 3 on the open city.

REFERENCES

Dixit, A., (1973), "The Optimum Factory Town, " *The Bell Journal of Economics and Management Science* 4, 637-651.

Kanemoto, Y., (1975), "Congestion and Cost-Benefit Analysis in Cities, " *Journal of Urban Economics* 2, 246-264.

Kanemoto, Y., (1976), "Optimum, Market and Second-Best Land Use Patterns in a von Thünen City with Congestion, " *Regional Science and Urban Economics* 6, 23-32.

Kanemoto, Y., (1977), *Theories of Urban Externalities*, Ph.D. thesis, Cornell University.

Legey, L., M. Ripper and P. Varaiya, (1973), "Effect of Congestion on the Shape of a City, " *Journal of Economic Theory* 6, 162-179.

Livesey, D.A., (1973), "Optimum City Size: A Minimum Congestion Cost Approach, " *Journal of Economic Theory* 6, 144-161.

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

Miyao, T., (1978), "A Note on Land-Use in a Square City, " *Regional Science and Urban Economics* 8, 371-379.

Oron, Y., D. Pines and E. Sheshinski, (1973), "Optimum vs. Equilibrium Land Use Patterns and Congestion Toll, " *The Bell Journal of Economics and Management Science* 4, 619-636.

Riley, J., (1974), "Optimal Residential Density and Road Transportation, " *Journal of Urban Economics* 1, 230-249.

Robson, A., (1976), "Cost-Benefit Analysis and the Use of Urban Land for Transportation, " *Journal of Urban Economics* 3, 180-191.

Sheshinski, E., (1973), "Congestion and the Optimum City Size, ” *American Economic Review* 63, 61-66.

Solow, P.M. and W.S. Vickrey, (1971), "Land Use in a Long Narrow City, " *Journal of Economic Theory* 3, 430-447.

CHAPTER V

TRAFFIC CONGESTION AND LAND USE FOR TRANSPORTATION: THE SECOND BEST CITY[1]

In the previous chapter we introduced traffic congestion and analyzed the optimum and market allocations. With transportation congestion, an additional traveler imposes external costs on other travelers by slowing them down. The optimal solution requires congestion tolls to "internalize" this externality. It is, however, difficult to charge congestion tolls because of very high administrative costs. In fact, there are very few roads where congestion tolls are levied and there is no city where congestion tolls are adopted in the whole city. It is, therefore, very important to consider what can be done given the constraint that congestion tolls are not allowed.

In the market city of the preceding chapter, we assumed that roads are built according to a naive benefit-cost criterion: the direct saving in transportation costs from widening the road is equated to the market land rent. This benefit-cost criterion leads to a misallocation of land between transportation and residential uses since, given the absence of congestion tolls, the market rent does not correctly reflect the true social scarcity of land.

In this chapter we consider the second best problem, which is to optimize the allocation of land between roads and residence when congestion tolls are not levied. The benefit-cost criterion that must be adopted to achieve the second best allocation is more complicated than the one in the optimum city or the market city. The cost side must be the *shadow rent*, or the *social rent*, which is no longer equal to the market rent. The benefit side also differs from the marginal direct saving in transportation cost (unless compensated demand for land is completely price or rent inelastic). The reason is as follows. A reduction of transportation costs from widening the road induces a change in the market rent. If demand for land is responsive to a price change, this has a side effect of changing the consumption decisions of households. As shown in the previous chapter, the social value of the change is zero due to the envelope property if the market rent equals the social rent. In the second best city, however, the market rent is not equal to the social rent, and a change in the consumption decision results in a net social gain or loss. The loss or gain is the difference between the social benefit and the marginal reduction in transportation costs.

Since the naive benefit-cost criterion usually adopted by policy makers leads to a suboptimal allocation of land, it is of interest to know the *direction* of the misallocation, that is, whether there is overinvestment or underinvestment in roads. The direction of

---

[1] This chapter is based on my 1977 paper in the Journal of Urban Economics. I would like to thank Academic Press, Inc. for permitting me to include an extended version of the paper in this book.

misallocation may be determined by comparing the market city with the second best city, but the second best city is, unfortunately, so complicated that we have not been able to carry out the comparison directly. We therefore examine the direction of a change that the naive benefit-cost criterion suggests at the second best optimum. More specifically, we compare the marginal saving in transportation costs and the market rent when roads are built in the second best way.

This comparison yields unambiguous results only if the benefit-cost criterion is adopted in a small region while the allocation in the rest of the city is held constant. The criterion leads to overinvestment in roads if the marginal saving in transportation costs is greater than the market rent at the second best optimum. If, however, the criterion is adopted in the entire city, interrelationships among different locations introduce complicated reactions, and we cannot obtain a definite answer.

In the second best city, the market rent at the edge of the city does not equal the rural rent although the shadow rent does. This result is in sharp contrast to those obtained in the optimum and market cities. The city must be expanded out to the radius where the contribution of an additional unit of land equals the rural rent. This requires the shadow rent to be equal to the rural rent. Since the market rent equals the shadow rent in the optimum city, the market rent also equals the rural rent at the edge of the city. In the second best city, however, the market rent is no longer equal to the shadow rent and hence is not equal to the rural rent at the edge.

Imposing another constraint that the market rent equals the rural rent at the edge of the city does not essentially change the situation. It is always possible, for instance, to make the width of the road zero and transportation costs per mile infinite at the edge of the city. This can cause a sudden drop in the market rent profile at the city's edge so that the market rent equals the rural rent after the drop and the constraint can be satisfied without changing the allocation inside the city. The only way to make the constraint significant is to restrict the shape of the road width functions, for example, to the class of linear functions as in Solow (1973).

The case where compensated demand for land is completely price inelastic is peculiar in the following two respects. First, the social marginal benefit of the road equals the direct marginal saving in transportation costs, since a change in rent caused by widening the road does not induce any change in consumption decision. Second, the absolute level of the market rent is indeterminate as long as difference in rents at different locations is such that the utility levels are equal. The second property mislead Solow and Vickrey (1971) and Kanemoto (1975) to conclude that the market rent is lower than the shadow rent everywhere in the city. In this case, there is no need for a jump in the market rent to make the market rent equal to the rural rent at the edge of the city, since the level of the market rent is indeterminate. This, coupled with the result that the slope of the shadow rent is steeper than that of the market rent, implies that the market rent is lower than the shadow rent everywhere in the city. This result, however, is misleading since it does not carry over to the case where the elasticity is not zero even when the elasticity is extremely small.

This chapter is organized as follows. The model is set up in section 1. Section 2 is the largest section in this chapter and devoted to the case of a closed city. The section is divided into three subsections: in subsection 2.1 the first order conditions for the second best optimum are derived and interpreted, in subsection 2.2 the benefit (the

direct marginal saving in transportation costs) and cost (the market rent) of the naive benefit-cost criterion based on market prices are compared at the second best optimum, and in subsection 2.3 the case of completely inelastic demand for land is considered. An open city is analyzed in section 3, and an economy consisting of many cities in section 4.

1. The Model

In this chapter we make the same technological assumptions as in Chapter IV. The only difference lies in the nature of the optimization problem: in this chapter congestion tolls are not allowed but the width of the road is optimized, whereas in the optimum city both congestion tolls and the width of the road could be chosen, and in the market city congestion tolls were not allowed and the road was built according to the erroneous benefit-cost criterion based on market prices.

Since congestion tolls are not allowed, households pay the private (or average) transportation cost, $t(x)$, defined by (IV.1.6) and (IV.1.7):

$$t'(x) = g(T(x), L_T(x)), \tag{1.1}$$

$$t(0) = 0. \tag{1.2}$$

If we denote the income of a household by y and the rent at x by $R(x)$, a household at x maximizes the utility function, $u(z(x), h(x))$, under the budget constraint

$$y = z(x) + R(x)h(x) + t(x). \tag{1.3}$$

Because of spatial arbitrage, the rent function, $R(x)$, must be such that the utility levels are equal everywhere in the city. As in section I.1.1, all this information can be summarized in the bid rent function,

$$R(x) = R(y - t(x), u), \tag{1.4}$$

which satisfies (I.1.14) and (I.1.15):

$$R_I(y - t(x), u) = 1/h(x), \tag{1.5}$$

$$R_U(y - t(x), u) = -1/v_I h(x), \tag{1.6}$$

where u is the equal utility level. Consumptions of the consumer good and housing are given by the compensated demand functions,

$$z(x) = z(R(x), u), \tag{1.7}$$

$$h(x) = h(R(x), u), \tag{1.8}$$

which satisfy (I.1.19) and (I.1.20):

$$z_R(R(x), u) \geq 0, \tag{1.9}$$

$$h_R(R(x), u) \leq 0. \tag{1.10}$$

The volume of traffic at x, $T(x)$, satisfies (IV.1.9) and (IV.1.10):

$$T'(x) = -L_H(x)/h[R(y-t(x),u),u], \tag{1.11}$$

$$T(\bar{x}) = 0. \tag{1.12}$$

The widths of the residential area and the road must satisfy the land constraint (IV.1.1):

$$L_H(x) + L_T(x) \le \theta(x). \tag{1.13}$$

2. A Closed City

In a closed city the population of the city is fixed, which yields the boundary condition (IV.2.1) for $T(x)$ at $x = 0$:

$$T(0) = P. \tag{2.1}$$

Using (1.4), (1.7), (1.8) and a different representation of transportation costs (tN instead of Tg), we can rewrite the resource constraint (IV.2.2) as

$$\int_0^{\bar{x}} \{\frac{z[R(y-t(x),u),u]+t(x)}{h[R(y-t(x),u),u]} L_H(x) + R_a\theta(x)\}dx \le Pw. \tag{2.2}$$

2.1. Derivation and Interpretation of First Order Conditions

In the second best problem the distortion of relative prices caused by the absence of congestion tolls is taken as given. The bid rent function (1.4) and demand functions, (1.7) and (1.8), of the consumer good and land capture the response of households to this distortion. Thus the second best problem maximizes the sum of utilities, (IV.2.3),

$$\int_0^{\bar{x}} \frac{uL_H(x)}{h[R(y-t(x),u),u]}dx, \tag{2.3}$$

under the constraints (1.1), (1.2), (1.11), (1.12), (1.13), (2.1), and (2.2). There are two state variables in this problem: $t(x)$ and $T(x)$. The control variables are $L_H(x)$ and $L_T(x)$. The control parameters are y, u, $\bar{x}$, and $t(\bar{x})$.

We assume that the market rent at the edge of the city, $R(\bar{x})$, is not restricted to equal the rural rent. In this case, there is no constraint on $t(\bar{x})$. The constraint on the market rent at the edge of the city does not cause any essential difference in the optimum allocation if we assume that transportation costs per mile become infinite, as the width of the road tends to zero. Under this assumption it is possible to have the same allocation inside the city and at the same time to satisfy the constraint by causing a jump in the market rent. Since the difference in allocation occurs only in an infinitesimal interval, this is possible without violating the resource constraint. Thus the constraint on the market rent is superfluous.

The Hamiltonian for the second best problem is

$$\Phi = [u - \lambda(x)]\frac{L_H(x)}{h[R(y - t(x), u)u]} + \eta(x)g(T(x), L_T(x))$$
$$- \delta\{\frac{z[R(y - t(x), u), u] + t(x)}{h[R(y - t(x), u), u]} L_H(x) + R_a\theta(x)\}, \tag{2.4}$$

where $\lambda(x)$, $\eta(x)$, and δ are respectively adjoint variables associated with (1.11), (1.1), and (2.2). Forming the Lagrangian,

$$\Psi = \Phi + \mu(x)[\theta(x) - L_H(x) - L_T(x)], \tag{2.5}$$

where $\mu(x)$ is the Lagrange multiplier for the constraint (1.13), we obtain the following necessary conditions for the optimum:

$$\frac{\partial \Phi}{\partial T} = -\lambda'(x) = \eta g_T(T, L_T), \tag{2.6}$$

$$\begin{aligned}\frac{\partial \Phi}{\partial t} &= -\eta'(x) \\ &= -\delta N + [u - \lambda - \delta(z + t)]\frac{N}{h} h_R R_I + \delta N z_R R_I\end{aligned} \tag{2.7}$$

$$\frac{\partial \Psi}{\partial L_H} = [u - \lambda - \delta(z + t)]\frac{1}{h} - \mu = 0, \tag{2.8}$$

$$\frac{\partial \Psi}{\partial L_T} = \eta g_L(T, L_T) - \mu = 0, \tag{2.9}$$

where μ and δ satisfy

$$\mu(x)[\theta(x) - L_H(x) - L_T(x)] = 0, \qquad \mu(x) \geq 0 \tag{2.10}$$

$$\delta\{Pw - \int_0^{\bar{x}} [zL_H/h + Tg + R_a\theta]dx\} = 0, \qquad \delta \geq 0. \tag{2.11}$$

The transversality conditions for $\bar{x}$, $t(\bar{x})$, u, and y are

$$\begin{aligned}\Phi(\bar{x}) = &[u - \lambda(\bar{x}) - \delta(z(\bar{x}) + t(\bar{x}))]N(\bar{x}) - \delta R_a\theta(\bar{x}) \\ &+ \eta(\bar{x})g(T(\bar{x}), L_T(\bar{x})) = 0\end{aligned}, \tag{2.12}$$

$$\eta(\bar{x}) = 0 \tag{2.13}$$

$$\int_0^{\bar{x}} \{N - [u - \lambda - \delta(z + t)]\frac{N}{h}[h_R R_u + h_u] - \delta N[z_R R_u + z_u]\}dx = 0, \tag{2.14}$$

$$\int_0^{\bar{x}} \{[u - \lambda - \delta(z + t)]\frac{N}{h} h_R R_I + \delta N z_R R_I\}dx = 0. \tag{2.15}$$

For convenience, we divide the shadow prices, λ, η, μ, and the utility, u, by δ, and substitute the original notations for the variables obtained. This operation converts the shadow prices from utility terms into pecuniary terms. Substituting (2.8) into (2.7), and noting that the rent function and the compensated demand functions satisfy both (1.5) and

$$Rh_R + z_R = 0,^{2} \tag{2.16}$$

we can rewrite (2.7) as

$$-\eta'(x) = -N - \frac{\mu - R}{R} eN, \tag{2.17}$$

where e is the price (rent) elasticity of compensated demand for land defined by

$$e = -\frac{Rh_R}{h} \geq 0. \tag{2.18}$$

The inequality is obtained because the substitution effect, h_R, is always nonpositive as in (1.10). Notice that e is a function of R and u and hence in general varies over space.

From (2.7), $-\eta'(x)$ can be interpreted as the social benefit of a unit increase of the commuting costs, $t(x)$, of residents living at x. When $t(x)$ increases by one unit, the total commuting costs are paid by $N(x)$ households who are living at x. This is represented by the first term on the RHS of (2.17). In addition to this direct effect, the increase of $t(x)$ has a side effect on the consumption decisions of households. The market rent, $R(x)$, must fall to compensate the increase of the commuting costs, $t(x)$, which induces a change in consumptions of housing and the consumer good. The second term on the RHS of (2.17) captures this indirect effect.

By the envelope property the second term vanishes when the social rent is equal to the market rent. The envelope property, (2.16), insures that, in the neighborhood of the equilibrium (or optimal) point, the changes in consumptions of the two goods evaluated at market prices counteract each other. In the first best world, therefore, where market prices reflect social values, the social cost of a unit increase of $t(x)$ is $N(x)$.

There is another case where the second term vanishes. When housing demand is completely price inelastic, $e = 0$, the change of the rent does not affect the consumption decision. Therefore, there is no side effect even when the social rent is different from the market rent. This is also a first best situation because the decisions of households are not affected by the existence of congestion and the first best solution can be attained without congestion tolls.

(2.17) shows that the adjustment of consumption has a socially desirable effect if R is greater than μ, which makes sense intuitively. An increase in commuting costs, $t(x)$, lowers the market rent, $R(x)$. When the market rent is higher than the social rent, a fall in the market rent brings it closer to the social rent, and the adjustment of

---

2 This can be shown as follows. By the definition of compensated demand functions, $h(R,u)$ and $z(R,u)$ must satisfy

$$u = u\left[h(R,u), z(R,u)\right],$$

for any R. Differentiating both sides with respect to R, we obtain

$$u_h h_R + u_z z_R = 0.$$

Since $u_h / u_z = R$, this implies

$$Rh_R + z_R = 0 \quad .$$

consumption works in the socially desirable direction.

Integrating (2.17) from x to $\bar{x}$ and using the transversality condition (2.13), we obtain

$$\eta(x) = -T - \int_x^{\bar{x}} \frac{\mu - R}{R} eN dx' . \tag{2.19}$$

Thus, $-\eta(x)$ is the social cost of increasing commuting costs of all households living between x and $\bar{x}$ by one unit.

Using this interpretation of $\eta(x)$, we can interpret $\lambda'(x)$ in (2.6) as the social congestion cost due to a unit increase in traffic. A unit increase in traffic between x and $x+dx$ causes more congestion there and raises transportation costs to pass through the ring by $gT(T, L_T)dx$. Since all households living beyond the ring must pass through the ring, the social cost of this increase in transportation costs is approximately $-\eta(x) g_T dx = \lambda'(x)dx$.

From (2.8) and the budget constraint (1.3), we have

$$\mu(x) = R(x) + \frac{u - y - \lambda(x)}{h(x)}, \tag{2.20}$$

where $\mu(x)$ is the shadow rent of land at x and the right hand side is the marginal value of land in residential use. The shadow rent differs from the market rent, and hence from the marginal rate of substitution between housing and the consumer good. The difference is caused by the second term on the right side, which reflects the congestion costs.

From (2.9) the shadow rent $\mu(x)$ also satisfies

$$\mu(x) = \eta(x) g_L(T(x), L_T(x)) . \tag{2.21}$$

The right side can be interpreted as the social marginal value of land in transportation use. A marginal increase of land allocated to roads lowers transportation costs at the radius. The social value of this decrease is given by the right side of (2.21).

From (2.6) and (2.21), we obtain

$$\lambda'(x)T - \mu L_T = -\eta[Tg_T + L_T g_L] ,$$

where, as shown in subsection 2.1 of Chapter IV, the square bracket on the right side is negative if transportation technology exhibits increasing returns to scale and positive in the case of decreasing returns. Since g_L is negative and $\mu(x)$ is nonnegative, (2.21) implies that $\eta(x)$ is nonpositive. Thus the following relationships hold between the total social congestion costs and the total shadow rent of roads at any radius:

$$\begin{array}{rl} < & \text{in the increasing returns case} \\ \lambda'(x)T(x) = \mu L_T & \text{in the constant returns case} \\ > & \text{in the decreasing returns case.} \end{array} \tag{2.22}$$

This result is more general than the condition obtained for the first best solution, where the relationship was expressed in terms of the actual congestion tolls and the road rent.

Using (2.19), we can rewrite (2.21):

$$\mu(x) = B(x) - g_L(T, L_T)\int_x^{\bar{x}} \frac{\mu - R}{R} eNdx', \tag{2.23}$$

where

$$B(x) \equiv -Tg_L(T, L_T) \tag{2.24}$$

is the marginal direct saving in transportation costs from widening the road as defined by (IV.2.20), and is sometimes called the *market benefit*. The second term on the right of (2.23) represents the social cost of the adjustment in the consumption of land for housing, which is characteristic of the second best world.

The naive benefit-cost criterion based on market prices cannot achieve the second best allocation of land. Although the social marginal values of land in residential and transportation uses are equal at the second best optimum, the market rent of the residential land is not in general equal to the market benefit of the road, since the market values differ from the social values as shown in (2.20) and (2.23).

It is easy to see that the transversality conditions, (2.12) and (2.13), imply that the shadow rent equals the rural rent at the edge of the city:[3]

$$\mu(\bar{x}) = R_a . \tag{2.25}$$

The transversality condition, (2.15), can be written more simply:

$$\int_0^{\bar{x}} \frac{\mu - R}{R} eNdx = 0 . \tag{2.26}$$

Though this equation is very important in deriving qualitative results (it is used in both Theorem 1 and Theorem 2 below), it is difficult to provide an interesting interpretation.

(2.14) can be simplified by using uncompensated demand functions for land and for the consumer good, $\hat{h}(I, R)$ and $\hat{z}(I, R)$, defined in (I.1.5) and (I.1.6) respectively. Compensated and uncompensated demand functions satisfy the following relationships derived in (3.14) and (3.16) of Appendix III.

$$h_u v_I = \hat{h}_I$$

$$z_u v_I = \hat{z}_I$$

$$\hat{h}_R = h_R - h\hat{h}_I .$$

From these equations and (I.1.9), (1.6), and (2.16), (2.14) can be written

$$\int_0^{\bar{x}} \left[\frac{1}{\delta} - \frac{1}{v_I}\right] N dx = -\int_0^{\bar{x}} [\mu - R] \frac{\hat{h}_R}{h} \frac{N}{v_I} dx . \tag{2.27}$$

---

[3] In deriving this condition, we assumed that g is finite at $\bar{x}$. It seems very unlikely that g becomes infinite at $\bar{x}$ because traffic is very light and available land is very large there.

This equation describes the relationship between the social value of the numeraire good (δ) in utility terms and the marginal utility of income (v_I). When there is no congestion, the right side vanishes and we obtain (I.2.23d) which says that the averages of reciprocals of these two are equal. When the shadow rent is not equal to the market rent, the reciprocal averages differ by the term on the right.

From (2.20) and (2.23), the benefit-cost criterion that must be used to achieve the second best allocation differs from the naive one adopted in the optimum and market cities. Unfortunately, it is not easy to calculate the correct benefit and cost. We can express the difference, $r(x)$, between the shadow rent, which represents the correct social cost, and the market rent by

$$r(x) \equiv \mu(x) - R(x) = \left[u - y - \lambda(x)\right] / h(x). \tag{2.28}$$

The difficulty is that the values of u and $\lambda(x)$ are not directly observable. The policy maker can, however, observe $h(x), N(x), T(x), L_T(x)$, and $R(x)$ without too much difficulty, and can estimate, with some more difficulty, the compensated price elasticity, $e(x)$. We therefore express $r(x)$ in terms of these variables. From (2.6) and (2.19), $r(x)$ satisfies

$$\begin{aligned} & r'(x)h(x) + r(x)h'(x) \\ & = \left[T(x) + \int_x^{\bar{x}} \frac{r(x')}{R(x')} e(x')N(x')dx' \right] g_T(T(x), L_T(x)), \end{aligned} \tag{2.29}$$

and from (2.26),

$$\int_0^{\bar{x}} \frac{r(x)}{R(x)} e(x)N(x)dx = 0. \tag{2.30}$$

The difference between the shadow rent and the market rent can be calculated by solving the differential equation (2.29) with the boundary condition (2.30). Then the social marginal cost of widening the road is simply the sum of the difference, $r(x)$, and the market rent, $R(x)$. Although it is not extremely difficult to solve the differential equation numerically in simple models like ours, the calculation is likely to be formidable in a more realistic model.

Once the difference between the shadow rent and the market rent is obtained, the social benefit can be easily calculated from (2.23):

$$B(x) - g_L(T, L_T) \int_x^{\bar{x}} \frac{r(x')}{R(x')} e(x')N(x')dx'.$$

2.2. Comparison of the Market Benefit and the Market Rent

Having simplified and interpreted first order conditions, we can now proceed to examine the consequence of the benefit-cost analysis based on market prices. Our ultimate goal is to compare the market benefit, $B(x)$, and the market rent, $R(x)$, at the second best optimum. It is convenient to compare the social rent, $\mu(x)$, with each of these first.

In this subsection we consider the case where compensated demand for land is not completely price inelastic: $e > 0$.

The social rent is equal to the market rent in the optimum city with optimal congestion tolls. If, however, congestion t tolls are not levied, the market rent diverges from the social rent. Since transportation costs are lower than they should be,

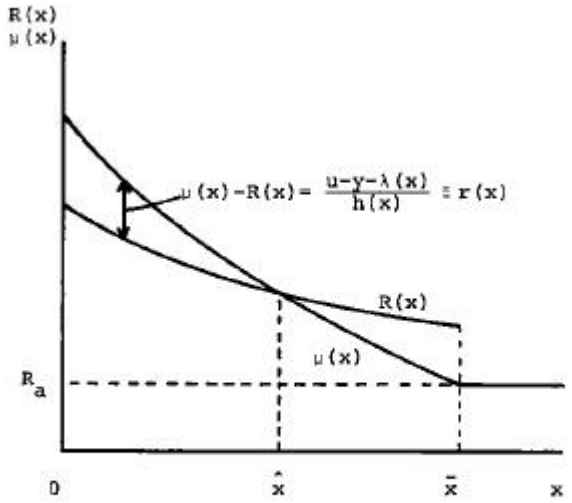

Figure 1. The Relationship between the Market Rent and the Social Rent

Households tend to locate too far from the CBD. People seeking land farther from the center bid up the rent at larger radii, and the market rent tends to be *flatter* than the social rent. The following Theorem shows that the market rent crosses the social rent at some intermediate radius, and that the social rent must be higher than the market rent inside the radius and lower outside the radius. This is illustrated in Figure 1.

Theorem 1: If $e > 0$ *for any radius, then there exists an* $\hat{x}$ *strictly between 0 and* $\bar{x}$ $(0 < \hat{x} < \bar{x})$ *such that* $\mu(\hat{x}) = R(\hat{x})$, *and*

$$\mu(x) > R(x) \qquad \text{for} \quad 0 \le x < \hat{x},$$

$$\mu(x) < R(x) \qquad \text{for} \quad \hat{x} < x \le \bar{x}.$$

Proof:

From (2.26) and $e > 0$, it is impossible to have $\mu(x) > R(x)$ for all x or $\mu(x) < R(x)$ for all x. Since both $\mu(x)$ and $R(x)$ are continuous, they must cross somewhere: there exists an $\hat{x}$, $0 \le \hat{x} \le \bar{x}$, where $\mu(\hat{x}) = R(\hat{x})$. From (2.20), at this point $\lambda(x)$ satisfies

$$\lambda(\hat{x}) = u - y .$$

From (2.6), (2.9), (2.10), (IV.1.3), and (IV.1.4), we obtain

$$\lambda'(x) = -\mu g_T / g_L > 0 .$$

This inequality is strict at $\hat{x}$ since $\mu(\hat{x}) = R(\hat{x}) > R_a > 0$. Hence we obtain the following inequalities:

$$\lambda(x) < u - y \qquad x < \hat{x}$$

$$\lambda(x) > u - y \qquad \hat{x} < x .$$

From (2.20), these inequalities imply

$$\mu(x) > R(x) \qquad x < \hat{x}$$

$$\mu(x) < R(x) \qquad \hat{x} < x ,$$

which in turn implies that $\hat{x}$ must be strictly between *0* and $\hat{x}$ to satisfy (2.26).

Q.E.D.

We next compare the market benefit and the social rent. The next Theorem shows that they are equal at $x = 0$ and that the market benefit is greater than the social rent in the rest of the city. Thus the market benefit overestimates the true social benefit. This is illustrated in Figure 2.

The result can be understood intuitively as follows. Recall that the difference between the market benefit and the social rent is the social value of the adjustment of consumptions in response to a decrease in transportation costs. First, consider the social value of the adjustment caused by a transportation improvement at $x = 0$. The improvement reduces commuting costs for all households by the same amount, which is equivalent to

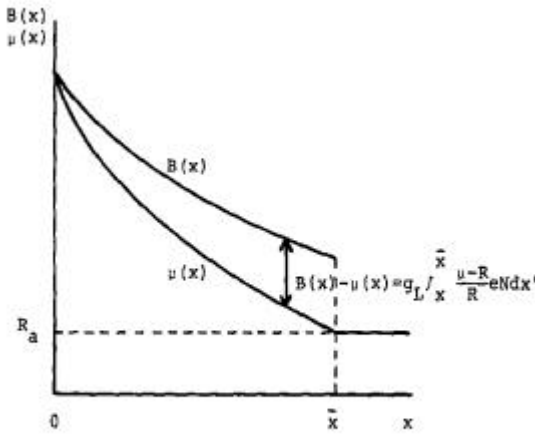

Figure 2. The Relationship between the Market Benefit and the Social Rent

an increase in the income, *y*, of every household in the city. Since *y* is optimally chosen, the change in the utility level caused by an infinitesimal increase in y is zero. The social value of the consumption adjustment is, therefore, zero for an improvement at $x = 0$.

Next, consider an improvement at any radius x beyond $\hat{x}$ in Theorem 1. This decreases commuting costs of households living farther than *x* and raises the market rent there. Since the social rent is lower than the market rent beyond $\hat{x}$, this works in a

socially undesirable direction and causes a social loss. Thus the social benefit (and hence the social rent) is less than the market benefit at any radius beyond $\hat{x}$.

An improvement inside $\hat{x}$ benefits both households living outside $\hat{x}$ and inside $\hat{x}$. The consumption adjustments of households outside $\hat{x}$ cause social losses for the same reason as above, but those of households inside $\hat{x}$ are socially beneficial since the social rent is higher than the market rent there. The next Theorem shows, however, that the former is always greater than the latter except for an improvement at $x=0$ in which case the two are equal.

Theorem 2: If $e>0$ *for any x, then we obtain*

$$\mu(0)=B(0)$$

and

$$\mu(x)<B(x), \qquad \text{for} \qquad 0<x\leq\bar{x}.$$

Proof:

We first show that for any x strictly between 0 and $\bar{x}$,

$$\int_x^{\bar{x}}\frac{\mu-R}{R}eNdx'<0.$$

For x greater than or equal to $\hat{x}$, this can be immediately obtained since $\mu(x)<R(x)$ from Theorem 1. For x less than $\hat{x}$ this is obtained from

$$\int_x^{\bar{x}}\frac{\mu-R}{R}eNdx'=-\int_0^x\frac{\mu-R}{R}eNdx'<0.$$

Hence (2.23) yields

$$\mu(x)<B(x) \qquad 0<x<\bar{x}.$$

At $x=0$, the following equality is obtained:

$$\mu(0)=B(0)-g_L(T(0),L_T(0))\int_0^{\bar{x}}\frac{\mu-R}{R}eNdx$$

$$=B(0),$$

where the second equality is obtained from (2.26), since g_L can be seen to be finite at $x=0$.

At $x=\bar{x}$, however g_L becomes infinite and we must use *L'Hôpital's Rule* to obtain

$$\begin{aligned}
&\lim_{x\to\bar{x}-} g_L \int_x^{\bar{x}} \frac{\mu - R}{R} eN dx' \\
&= \lim_{x\to\bar{x}-} \frac{\mu(x)}{\eta(x)} \int_x^{\bar{x}} \frac{\mu - R}{R} eN dx' \\
&= R_a \lim_{x\to\bar{x}} \frac{1}{\eta'(x)} \frac{\mu - R}{R} eN \\
&= \frac{R_a[R_a - R(\bar{x})]e}{R(\bar{x}) + [R_a - R(\bar{x})]e} \\
&< 0 \ ,
\end{aligned}$$

where the first equality is obtained from (2.9), the second equality by *L'Hôpital's Rule*, the third equality from (2.17), and the inequality from $R(\bar{x}) > \mu(\bar{x}) = R_a$ and the elementary result that the limit must be nonpositive when it is approached through nonpositive values. From (2.23) this implies

$$\mu(\bar{x}) < B(\bar{x})$$

Q.E.D.

Combining Theorems 1 and 2, we can immediately see that the market benefit is greater than the market rent near the center. However, it is not clear whether or not this remains to be true when we move farther from the center. The next proposition throws a light on this question.

Proposition 1: If the compensated demand for land is not completely price inelastic $(e > 0)$*, then the market benefit is always greater than the market rent near the CBD. Near the edge of the city, however, the market benefit is smaller than the market rent if the price elasticity is less than one, and is greater than the market rent if the elasticity is greater than one.*

Proof:

The first half is immediately obtained from Theorem 1 and 2.

From the proof of Theorem 2, we obtain

$$\begin{aligned}
R(\bar{x}) - B(\bar{x}) &= R(\bar{x}) - R_a + R_a - B(\bar{x}) \\
&= \frac{[R(\bar{x}) - R_a]R(\bar{x})(l - e)}{R(\bar{x}) + [R_a - R(\bar{x})]e} .
\end{aligned}$$

Noting that the denominator and the square bracket of the numerator are both positive, we get

$$R(\bar{x}) \underset{>}{\overset{<}{=}} B(\bar{x}) \quad \text{where} \quad e \underset{>}{\overset{<}{=}} 1 .$$

Q.E.D.

Figure 3 illustrates the relationship between the market benefit and the market rent in the case of price inelastic demand for land: the market benefit is greater than the market rent near the center of the city, but drops below it near the edge. As a result, the naive benefit-cost criterion has a tendency to overinvest in roads near the center and to underinvest near the edge. When demand for land is price elastic as in Figure 4, the benefit-cost criterion tends to overinvest in roads both near the center and near the edge of the city[4].

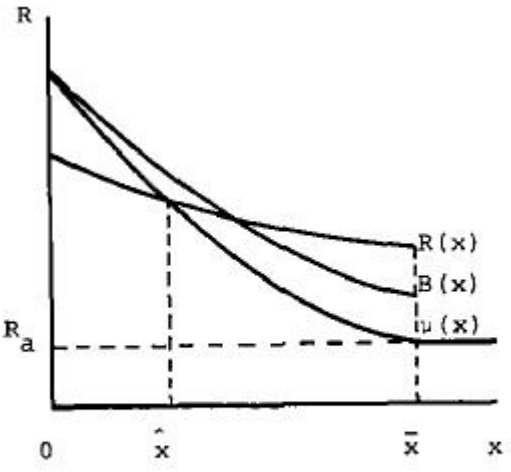

Figure 3
Price Inelastic Case: e < 1

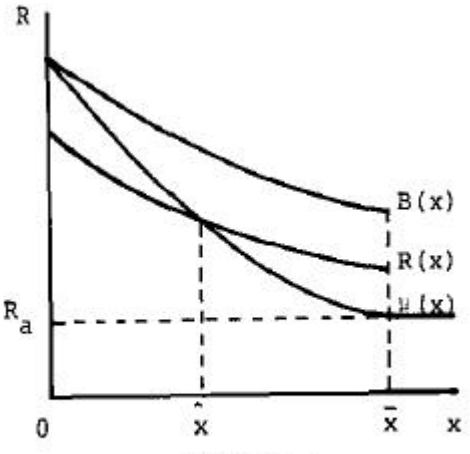

Figure 4
Price Elastic Case: e > 1

Since the Cobb-Douglas type utility function (IV.2.27) has the elasticity $1-\alpha$, which is always less than *1*, there is a tendency in that case to overinvest in roads near the center and to underinvest near the edge.

The conclusion depends on the elasticity of demand for land since difference between the market benefit and the social rent reflects the side effect due to the change of housing consumption, and the change of housing consumption is greater when the elasticity is bigger.

Notice that since these results are valid only in the neighbourhood of the second best solution, we do not have a definite answer as to whether the second best solution has a wider road than the market solution.

When the naive benefit-cost analysis based on market prices is adopted only in a small ring at x, and roads are built in other parts of the city to achieve the second best allocation, the above comparison between two equilibria is valid. If, for example, the market benefit is greater than the market rent in the ring between x and $x+dx$, the naive criterion calls for the road to be widened until the marginal market benefit of further widening falls to the market rent. When the ring is very narrow the market rent is not significantly affected by a change in road width there, and the preceding conclusions hold.

If, however, the naive benefit-cost criterion is adopted in the entire city, this

---

[4] Note that the case where $B(x)$ is lower than $R(x)$ somewhere in the middle of the city is not excluded.

argument cannot be applied because the market rent curve changes. Widening of the road in the rest of the city might cause such a rise in market rent at some locations that, even though the market rent at the second-best allocation was below the market benefit, the road might become narrower as a result of changes elsewhere.

Furthermore, since the market rent is higher than the rural rent at the edge of the city, the city tends to expand. This causes another tendency toward overinvestment in roads. The reader may think that this effect would not appear if the second best problem were solved with the additional constraint that the market rent equal the rural rent at the edge of the city. In our model, however, under the reasonable assumption that transportation costs per mile, $g(T, L_T)$, are infinite when the width of the road is zero, the constraint is superfluous and the effect does not disappear.

The constraint on the market rent at the boundary,

$$R(y - t(\bar{x}), u) = R_a, \tag{2.28}$$

would restrict y, $t(\bar{x})$, and u to a hypersurface. The optimum allocation for the problem with this additional constraint is essentially the same as that for the problem without the constraint: the allocation is exactly the same within the boundary $\bar{x}$, and $g(T, L_T)$ is made infinite at $\bar{x}$ causing a jump in $t(x)$ of an appropriate size to satisfy the constraint (2.28). Since the jump which occurs in an infinitesimally small interval does not involve a finite social cost, the same maximum without the constraint is attained.[5]

Now, we briefly consider the possibility that $t(x)$ has jumps even without the constraint (2.28). In such a case the usual maximum principle like the Theorem of Hestenes in Appendix IV cannot be applied since it assumes that state variables are continuous. Kanemoto (1977b) analyzed the case by considering the problem with an upper bound on $g(T, L_T)$ and letting the upper bound tend to infinity.

The following argument shows that a jump in $t(x)$ is indeed possible. Equation (2.21) suggests that $\eta(x)$ must be nonpositive, since $\mu(x)$ is nonnegative. There is no guarantee, however, that $\eta(x)$ is nonpositive since $\eta(x)$ must also satisfy (2.19). If compensated demand for housing is sufficiently price elastic, the indirect benefit from increasing transportation costs (the second term on the right side of (2.19)) may overwhelm the direct cost (-T), in which case $\eta(x)$ becomes positive. Then the necessary conditions for the optimum involve contradiction, which suggests that the maximum does not exist within the range of functions assumed by the maximum principle.

In order to show that such a case can occur, we rewrite (2.19) as

$$\eta(x) = -\int_x^{\bar{x}} \left[1 + \frac{\mu - R}{R} e \right] N dx' .$$

---

[5] It can be shown that, if g is infinite when L_T is zero, then a jump in $t(x)$ may occurs at $\bar{x}$. See Kanemoto (1977b). Although the proof there has a minor error, the conclusion can be easily seen to be correct.

This equation shows that, if $e(\bar{x}) > R(\bar{x})/[R(\bar{x}) - R_a]$, $\eta(x)$ is positive near $\bar{x}$. In particular, if $R_a = 0$ and $e(\bar{x}) > 1$, then $\eta(x)$ is positive. There certainly exists a well-behaved utility function whose compensated demand function is price elastic.

In Kanemoto (1977b) it was shown that, if $g(T, L_T)$ tends to infinity as traffic density, T/L_T, approaches infinity, a jump in $t(x)$ occurs at a point where $\eta(x)$ is positive. Theorem 1 remains valid even when a jump occurs. Theorem 2 and Proposition 1 are also valid if R_a and $R(\bar{x})$ are replaced by the left side limits,

$$\bar{\mu}(\bar{x}) = \lim_{x \uparrow \bar{x}} \mu(x),$$

$$\bar{R}(\bar{x}) = \lim_{x \uparrow \bar{x}} R(x).$$

If g remains finite even when T/L_T approaches infinity, L_T becomes zero for a finite length. It can be easily seen that if the upper bound for g is sufficiently large, the same results are obtained.

2.3. Completely Price Inelastic Demand for Land

Next, consider the case where the compensated demand for land is completely price inelastic: $e = 0$ for any u and R. This case is obtained, for example, if the utility function is a Leontief type, so that land and the consumer good are always consumed in fixed proportions.

As we mentioned in subsection 2.1, the side effect due to the adjustment of consumption decisions vanishes in this case,

$$\eta(x) = -T(x),$$

and the market benefit coincides with the social rent,

$$\mu(x) = B(x), \qquad 0 \leq x \leq \bar{x}.$$

Since (2.26) is satisfied at all levels of $R(x)$, the level of $R(x)$ is indeterminate. This can be understood as follows. Suppose that the optimum is obtained by the rent function, $R^*(x)$. Consider the effect of raising the rent function by an arbitrary amount c everywhere in the city. Since the utility level cannot be higher than the optimal level, if we can show that the optimal utility level is attained even when the market rent is $R^*(x) + c$, we can conclude that the market rent is indeterminate at the optimum.

When the utility level is given, the assumption of completely inelastic demand implies that lot sizes are constant regardless of the market rent. This has two implications: the lot size is the same everywhere in the city, and it does not change when the rent profile rises to $R^*(x) + c$. In our model differential rent is returned to residents as an equal subsidy, so the income of households rises by ch^*, where h^* is the optimal lot size. Households, therefore, can afford the optimal bundle at the higher rent level, and the optimum utility level is attained with the new market rent profile,

$R^*(x)+c$. The market rent is thus indeterminate if $e=0$.

One important implication of this indeterminacy is that the optimal solution can be achieved without having a jump in the rent function even if we add the constraint that the market rent be equal to the rural rent at the boundary. After solving for the optimal allocation without the constraint, we simply lower the market rent curve until the rent at the boundary equals the rural rent. This observation yields the following proposition which is the result obtained by Solow and Vickrey (1971), and Kanemoto (1975).

Proposition 2: If the compensated demand for land is completely price inelastic, and if the market rent equals the rural rent at the edge of the city, then at the optimum the market benefit equals the market rent at the edge of the city and is greater in the rest of the city.

This proposition is illustrated in Figure 5. Note that the second best optimum coincides with the first best optimum, since, when demand for land is completely price inelastic, the only difference between them is the market rent that does not affect consumption decisions of households.

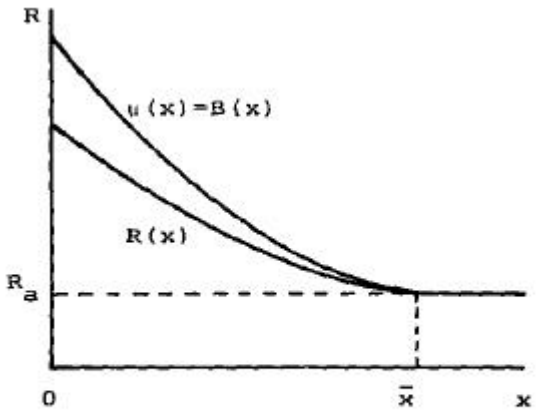

Figure 5
Completely Price Inelastic Case
with $R(\bar{x}) = R_a$

The proposition suggests that there is a strong tendency towards overinvestment in roads when $e=0$. Considering the results obtained in the preceding section, however, the proposition is somewhat misleading. As long as compensated demand for land is not completely price inelastic, the market rent is not indeterminate and we obtain a situation like the one depicted in Figure 1, where the social rent is higher than the market rent near the center and lower near the edge. Although the market benefit approaches the social rent as the elasticity tends to zero, the relationship between the market rent and the social rent remains basically the same as long as the elasticity is positive, since (2.26) is effective even when the elasticity is very small. How the relationship among the market rent, the social rent, and the market benefit changes as the elasticity becomes smaller is illustrated in Figure 6. If the elasticity is greater than

1, the market benefit is greater than the market rent at the edge of the city, as in Figure 6a (which reproduces Figure 4). If the elasticity is between 0 and 1, the market benefit falls below the market rent but is still higher than the social rent at $x = \bar{x}$, as in Figure 6b (or Figure 3). As the elasticity approaches zero, the market benefit tends to the social rent, but the market rent remains higher than the social rent at $x = \bar{x}$. In the limit we obtain the case, depicted in Figure 6c, in which the market benefit is less than the market rent near the edge of the city. Thus Figure 5 and hence Proposition 2 cannot approximate the case where the elasticity is close to, but not exactly, zero.

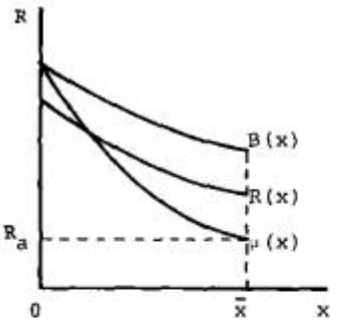

Figure 6a. e>1

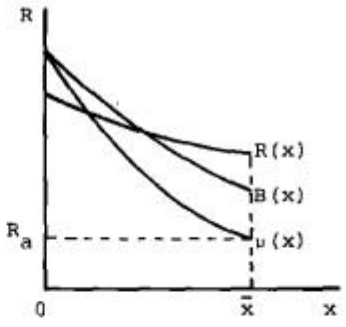

Figure 6b. 0<e<1

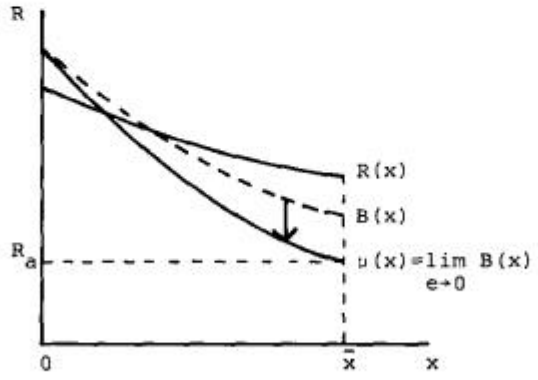

Figure 6c. e→0

Figure 6. Completely Price Inelastic Case as a Limit as e → 0.

The conclusion that the naive benefit-cost criterion has a tendency toward overinvestment is nevertheless correct, since the market city has a wider road than the optimum city, as shown in Kanemoto (1975). The main reason is that at the second best optimum the market rent is higher than the rural rent at the edge of the city. This tends to make the market city larger than the second best city. In the models in Solow and Vickrey (1971) and Kanemoto (1975), where a fixed amount of land is required for nontransportation use, the city can grow only if the road is widened.

3. An Open City

Next, consider an open and small city in which the utility level is given from

outside: $u = \bar{u}$. This time we consider the, absentee-landlord case. The income of a household is given by the value of marginal productivity of labour: $y = w$. These two conditions replace the population constraint (2.1) and the resource constraint (2.2) in a closed city.

The bid rent function (1.4) and the compensated demand functions, (1.7) and (1.8), become

$$R(x) = R(w - t(x), \bar{u}), \tag{3.1}$$

$$z(x) = z(R(x), \bar{u}), \tag{3.2}$$

$$h(x) = h(R(x), \bar{u}). \tag{3.3}$$

The net product of the city after the cost of maintaining the given utility level of residents,

$$\int_0^{\bar{x}} \{[w - z(x) - t(x)]N(x) - R_a\theta(x)\}dx, \tag{3.4}$$

is maximized. The Hamiltonian and the Lagrangian for this problem are

$$\begin{aligned}\Phi = &\frac{w - z[R(w - t(x), \bar{u}), \bar{u}] - t(x) - \lambda(x)}{h[R(w - t(x), \bar{u}), \bar{u}]} L_H(x) - R_a\theta(x) \\ &+ \eta(x)g(T(x), L_T(x))\end{aligned} \tag{3.5}$$

and

$$\Psi = \Phi + \mu(x)[\theta(x) - L_H(x)L_T(x)], \tag{3.6}$$

where $\lambda(x)$ and $\eta(x)$ are respectively the adjoint variables associated with (1.11) and (1.1), and $\mu(x)$ is a Lagrange multiplier for (1.13).

The control variables are $L_H(x)$ and $L_T(x)$, and the control parameters are $\bar{x}$, $t(\bar{x})$ and $T(0)$. We assume that a city planner can determine the boundary of the city regardless of the level of the market rent there. Under this assumption there is no constraint on $t(\bar{x})$.

The first order conditions are

$$\mu(x) = R(x) - \lambda(x)/h(x), \tag{3.7}$$

$$\mu(x) = B(x) - g_L(T, L_T)\int_x^{\bar{x}} \frac{\mu - R}{R} eN dx', \tag{3.8}$$

$$\lambda'(x) = -\mu g_T / g_{L'} \tag{3.9}$$

$$\lambda(0) = 0, \tag{3.10}$$

$$\mu(\bar{x}) = R_a, \tag{3.11}$$

where $B(x)$ is defined by (2.24), $\mu(x)$ satisfies (2.10), and e is the price elasticity of compensated demand for land as defined by (2.18). These conditions are similar to

those obtained for a closed city and have similar interpretations.[6]

Calculations of the correct benefit and cost are the same as in the closed city except for the boundary conditions. From (3.7) through (3.10), the difference between the shadow rent and the market rent, $r(x)$, satisfies the differential equation

$$r'(x)h(x)+r(x)h'(x)=\left[T+\int_x^{\bar{x}}(r/R)eNdx'\right]g_T(T,L_T), \tag{3.12}$$

with the boundary condition

$$r(0)=0.$$

When this differential equation is solved, the social marginal cost of the road is given by $r(x)+R(x)$, and the social marginal benefit is

$$B(x)-g_L\int_x^{\bar{x}}(r/R)eNdx'.$$

Next, we compare the market benefit, $B(x)$, and the market rent, $R(x)$, at the second best optimum to see whether the naive benefit-cost criterion results in overinvestment in roads. In order to do so, we first compare the market rent, $R(x)$, and the social rent, $\mu(x)$. Since congestion tolls are not imposed, the social transportation costs are greater than the private trans-portation costs. The social rent, therefore, tends to be steeper than the market rent. In the open city, however, both rents are equal at the center by the transversality condition (3.10). Thus the social rent is lower than the market rent everywhere in the city except at the center where they are equal, and the following theorem is obtained.

Theorem 3:

$$\mu(0)=R(0),$$

and

$$\mu(x)<R(x), \qquad 0<x\leq\bar{x}.$$

We omit the proof, which is quite simple. Notice that this theorem holds even if the compensated demand for land is completely price inelastic.

Next, we compare the market benefit and the social rent. The market benefit differs from the social rent by the indirect effect through consumption decisions. A reduction in transportation costs at a radius has a tendency to raise the market rent beyond that radius. Since, by Theorem 3, the market rent is higher than the social rent, raising the market rent increases the gap. The indirect effect of a reduction in transportation costs thus causes a social loss, and the social benefit is smaller than the

---

[6] As in the closed city, $\eta(x)$ may become positive, and a jump in $t(x)$ may occur. However, the following theorems and proposition hold even if $t(x)$ has a jump.

market benefit.

Theorem 4: If $e > 0$ for all x, then

$$\mu(x) < B(x), \qquad 0 \leq x \leq \bar{x}.$$

For $x < \bar{x}$, the Theorem is immediately obtained from (3.8) and Theorem 3. At $x = \bar{x}$, *L'Hôpital's* rule yields the inequality as in the proof of Theorem 2.

The above two theorems show that the market benefit is greater than the market rent at least near the center. The naive benefit-cost analysis, therefore, has a tendency to overinvest in roads near the center. The following proposition shows that this pattern is reversed near the edge of the city if the elasticity of demand for land is less than one.

Proposition 3: Suppose the compensated demand for land is not completely price inelastic. Then the market benefit is greater than the market rent near the center. If, further, the price elasticity of compensated demand for land is less (greater) than one, the market benefit is smaller (greater) than the market rent near the edge of the city.

The proof is the same as that of Proposition 1. Figure 7 depicts the case of inelastic demand. Figure 8 the case of elastic demand. Notice that relative positions of the market benefit and the market rent are the same as in a closed city though their relationships with the social rent are different.

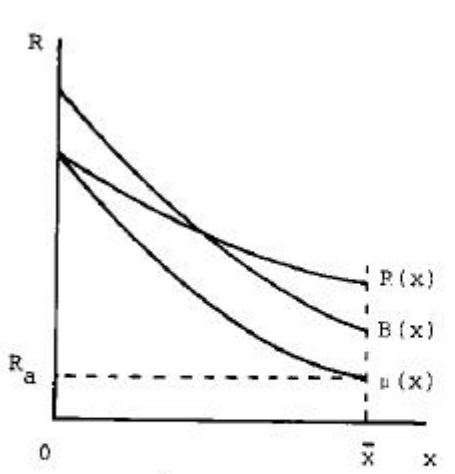

Figure 7
Inelastic Demand:
An Open City

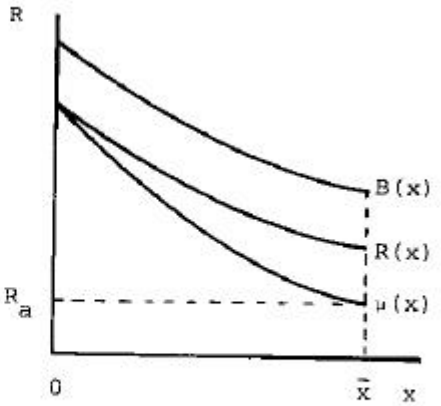

Figure 8
Elastic Demand:
An Open City

In a closed city the market benefit equaled the social rent at the center, but in an open city the market benefit exceeds the social rent everywhere. In a closed city the market

rent crossed the social rent at some intermediate point, while in an open city the market rent is equal to the social rent at the center.

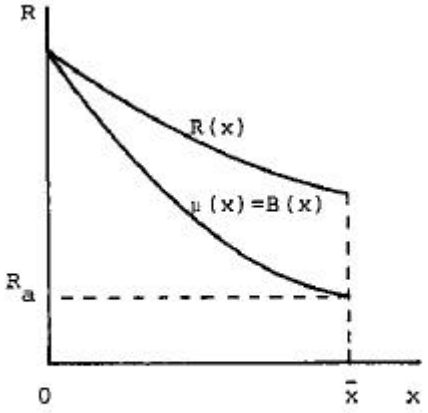

Figure 9
Completely Price Inelastic Demand:
An Open City

When compensated demand for land is completely price inelastic, the second term on the RHS of (3.8) vanishes. The market benefit, therefore, coincides with the social rent and we obtain the following proposition which is illustrated in Figure 9.

Proposition 4: If compensated demand for land is completely price inelastic, then the market benefit is equal to the market rent at the center and is smaller than the market rent in the rest of the city.

Thus, in sharp contrast to Proposition 2 in a closed city, there is a tendency to underinvest in roads everywhere in the city. Since the market rent is higher than the rural rent at the edge of the city, however, the market city tends to be bigger than the optimum city. This increases the total population of the city and hence the total traffic, which works in the direction of widening the road. In Kanemoto (1975), the road is shown to be wider in the market city than in the optimum city.

4. An Economy with Many Cities

In this section we consider an economy consisting of many cities. The model is the same as that in section 4 of the preceding chapter. The population constraint is

$$P = nP_c \tag{4.1}$$

where P, P_c, and n are respectively the population of the economy, the population of a city, and the number of cities. The resource constraint is

$$\int_0^{\bar{x}} \left\{ \frac{z[R(y-t(x),u),u]+t(x)}{h[R(y-t(x),u),u]} + R_a\theta(x) \right\} dx \leq F(P_c). \tag{4.2}$$

The aggregate production function, $F(P_c)$, has increasing returns to scale. The boundary condition for $T(x)$ at $x=0$ is

$$T(0)=P_c. \tag{4.3}$$

The common utility level is maximized under the constraints (1.1), (1.2), (1.11), (1.12), (1.13), (4.1), (4.2), and (4.3). The control variables are $L_H(x)$ and $L_T(x)$, and the control parameters are P_c, n, y, u, $\bar{x}$, and $t(\bar{x})$. The Hamiltonian for this problem is

$$\Phi=-\lambda(x)\frac{L_H(x)}{h[R(y-t(x),u),u]}+\eta(x)g(T(x),L_T(x))$$
$$-\delta\left\{\frac{z[R(y-t(x),u),u]+t(x)}{h[R(y-t(x),u),u]}L_H(x)+R_a\theta(x)\right\}, \tag{4 4}$$

and the Lagrangian is

$$\Psi=\Phi+\mu(x)[\theta(x)-L_H(x)-L_T(x)], \tag{4.5}$$

where $\lambda(x)$, $\eta(x)$, and δ are adjoint variables associated with (1.11), (1.1), and (4.2) respectively, and $\mu(x)$ is a Lagrange multiplier for (1.13).

After dividing $\lambda(x)-\lambda(0)$, $\eta(x)$, and $\mu(x)$ by δ and denoting the obtained variables by $\lambda(x)$, $\eta(x)$, and $\mu(x)$ respectively, the first order conditions become (2.19), (2.23), (2.25), (2.26),

and

$$\lambda(x)=-\int_0^x\eta(x')\,g_T(T(x'),L_T(x'))dx', \tag{4.6}$$

$$\frac{1}{h(x)}[F'(P_c)-z(x)-t(x)-\lambda(x)]=\mu(x), \tag{4.7}$$

$$\int_0^{\bar{x}}\frac{1}{V_I}Ndx-\frac{1}{\delta}=\int_0^{\bar{x}}[\mu(x)-R(x)]\frac{\hat{h}_R}{h}\frac{N}{V_I}dx. \tag{4.8}$$

(4.6), (4.7), and (4.8) correspond to (2.6), (2.20), and (2.27). As before, $-\eta(x)$ is the social cost of increasing commuting costs of all households passing through x by one unit. $-\eta(x)g_T$ is, therefore, the social cost of an increase in congestion caused by a unit increase in the traffic at x, and $\lambda(x)$ is the social congestion costs that a resident at x imposes on other travelers by commuting from x to the center.

Multiplying (4.7) by $h(x)N(x)$ and integrating from 0 to $\bar{x}$ yields

$$P_cF'(P_c)=\int_0^{\bar{x}}[(z+t)N+\lambda N+\mu L_H]dx.$$

Comparing this equation with the resource constraint (4.2) and noting that the resource constraint holds with equality at the optimum, we obtain

$$-\left[F(P_c)-P_cF'(P_c)\right]=\int_0^{\bar{x}}\left[\lambda N+\mu L_H-R_a\theta\right]dx. \tag{4.9}$$

Thus *the operating loss of a producer at the optimum equals the total social congestion costs, plus the total social rent of residential land, minus the total payment of the rural rent*. This is similar to the result obtained in the previous chapter: the operating loss of a firm equals the total congestion tolls, plus the total rent of residential land, minus the total payment of the rural rent. The difference is that there are no tolls capturing the social congestion costs in this chapter and the social rent does not equal the market rent. It is quite natural that the same relationship holds for social values instead of market values.

As shown in subsection 2.1, if we assume constant returns to scale in transportation technology, the social congestion costs equal the total shadow rent of roads at each radius:

$$\lambda'(x)T(x)=\mu(x)L_T(x), \qquad 0\le x\le\bar{x}.$$

Then by integration by parts, (4.9) becomes

$$-\left[F(P_c)-P_cF'(P_c)\right]=\int_0^{\bar{x}}\left[\mu(x)-R_a\right]\theta(x)dx. \tag{4.10}$$

This is again similar to the relationship obtained in Chapter IV. The operating loss of a producer equals the difference between the total social rent and the total payment of the rural rent, where the total social rent includes the rent on the road. Note that this relationship does not in general hold for the market rent, since (2.26) requires that the sums of the market and social rents be equal when they are weighted by *eN/R* which equals $\theta(x)$ only by chance.

It is easy to see that the social benefit and cost can be calculated exactly in the same way as in the closed city. The relationships among the social rent, the market rent, and the market benefit are also the same as in the closed city.

5. Concluding Remarks

The analysis in this and the preceding chapters are centered on the interaction between pricing of traffic congestion and the investment decision of roads. If congestion is optimally priced, the investment decision is quite straightforward. The allocation of land between roads and residence must be determined in such a way that the marginal *social* benefits of widening the road equals the marginal social cost at each radius. The marginal social benefit at a radius is simply the marginal direct saving in transportation costs with the volume of traffic there fixed; the marginal social cost is the market rent of the residential land.

This simplicity in the benefit-cost criterion is the general property of the first best world where all goods are priced properly. Since all prices reflect the true social marginal values of the goods, prices may stand in for social values in the calculation of benefits and costs. Thus the marginal social cost of widening the road is given by the market rent in our model.

The fact that all prices reflect the social marginal values has another important implication. When the road is widened, commuting costs decrease and hence the land

rent rises. This induces a change in the allocation of the entire city through a change in the consumption bundles of households. The change however, can be ignored in the calculation of the marginal benefit and cost. The reason is that the *social* values of the induced change is zero, since the *market* value of the induced change is zero due to the envelope property, and the market value equals the social value when all prices equal the social marginal values. This is the reason why the marginal social benefit equals the marginal direct saving in transportation costs with the fixed traffic volume.

The simplicity disappears if traffic congestion is not properly priced. Prices no longer reflect the marginal social values of goods accurately, and in particular, the market rent does not equal the social marginal value of residential land. Accordingly, the cost side of the benefit-cost criterion must be changed. The benefit side also becomes more complicated since the induced change in the consumption decisions has a nonzero social value or loss. The naive benefit-cost analysis usually adopted by policy makers, therefore, gives rise to an inefficient land use.

Unfortunately, the correct benefit-cost criterion is difficult to calculate. Furthermore, boundary conditions that must be used to calculate the benefit-cost criterion are different between closed and open cities. The correct benefit cost criterion is, therefore, unlikely to be practical, at least until we know more. Meanwhile, it would be useful to know whether the naive benefit-cost analysis leads to too wide a road.

The results in Chapter IV suggest that the road in the city with the naive benefit-cost analysis is usually wider than that it in the first best optimum where congestion tolls are levied and roads are optimally built. This comparison, however, may not be useful, since it is difficult to levy congestion tolls because of very high administrative costs. The analysis in this chapter is a partial attempt at the comparison with the second best optimum in which roads are built optimally under the constraint that congestion tolls are impossible. We compared the benefit and the cost in the erroneous benefit-cost criterion at the second best optimum and showed that the benefit exceeds the cost near the center and that the benefit exceeds the cost also near the edge in the case of price elastic demand for land and is less than the cost in the price inelastic case. This implies that, if the erroneous benefit-cost criterion is adopted only in a very narrow ring near the center, overinvestment in roads will result. If it is adopted near the edge underinvestment will result in the inelastic case and overinvestment in the elastic case.

Unfortunately, the analysis is not conclusive if the erroneous benefit-cost criterion is adopted everywhere in the city. It seems, however, more likely that the naive benefit-cost criterion leads to overinvestment in roads. The major reason is that the market rent is higher than the rural rent at the second best optimum and the market city with the benefit-cost criterion tends to be bigger, which is made possible only by building wider roads and lowering commuting costs. The results obtained in somewhat different models by Wheaton (1978), Pines and Sadka (1979), and Wan (1979) also support this conjecture.

Notes

The analysis in this chapter originates in Solow and Vickrey (1971). They

formulated a transportation cost minimization problem in a long narrow city framework and asked the question whether or not the cost-benefit analysis based on the market rent yields too wide a road. To see this, they compared the benefit from widening the road with the market rent at the optimum configuration.

They, in effect, made the following three assumptions. First, the city was assumed to be closed in the sense that the total production (or the total population when interpreted as a residential model) in the city was fixed. Second, they assumed that only production required land, that production required only land, and that the price elasticity of demand for land was zero so that demand for space was not affected by the level of land rent. Third, the market rent was constrained to be equal to the rural rent (in their case, zero rent) at the boundary of the city. Their model, therefore, corresponds to the case of subsection 2.3 in this chapter. Naturally, they obtained exactly the same conclusion as in Proposition 2 - that the benefit is greater than the market rent everywhere in the city - and concluded that the cost-benefit analysis based on market rent has a tendency to overinvest in roads.

Kanemoto (1975) introduced an open city facing a given export price, and compared it with a closed city. The model is essentially the same as the completely-price-inelastic case of the open city in this chapter. The relationship between the market benefit and the market rent at the optimum allocation of land is the same as that in Proposition 4.

Since these models assume completely price inelastic demand for land, the first best allocation coincides with the second best allocation. The second best allocation differs from the first best allocation if substitution between land and other goods is possible. Solow (1975) first considered this type of a second best problem in a spatial equilibrium framework. He maximized the utility level of households within the class of linear road width functions in a closed city. According to his numerical calculations, the market benefit from widening the road is greater than the market rent. He explained this result as follows. Since congestion tolls are not levied, the market rent is flatter than the social rent. But the two rents are equal to the rural rent at the edge of the city. The market rent is therefore lower than the social rent, and the value of land is underestimated in the naive cost-benefit calculations.

Our analysis indicates that this explanation fails to notice the following two aspects of the second best allocation. First, though the social rent is steeper than the market rent, the two are not in general equal at the edge of the city. Our analysis shows that the market rent is higher than the social rent at the edge of the city. Second, the market benefit from widening the road does not correctly reflect the social benefit. The market benefit is greater than the social benefit because the adjustment of consumption caused by a decrease in transportation costs involves social costs when congestion tolls are not levied.

Kanemoto (1976) considered a production city with substitutability between labour and land in an open city framework. The results are parallel to those in section 3. The analysis of a closed city is based on Kanemoto (1977a).

Wheaton (1978) considered a similar problem in a nonspatial framework with more than one type of roads. He also analyzed the problem of finding the optimal uniform congestion tax which is constrained to have the same tax rate on all roads regardless of different degrees of congestion.

Arnott (1979) extended our analysis to the case where the road is of arbitrary width. Arnott and MacKinnon (1978) obtained the numerical solution of using the fixed point algorithm. Wan (1979) applied the perturbation method to the second best problem and also obtained numerical solutions.

Pines and Sadka (1979) considered a discrete model in which a city is divided into two rings. Assuming that the areas of the two rings are fixed, they showed that there is more investment in roads in the market city with the naive benefit-cost analysis than in the second best city.

REFERENCES

Arnott, R., (1979), "Unpriced Transportation Congestion, " *Journal of Economic Theory*, forthcoming.

Arnott, R. and J. MacKinnon, (1978), "Market and Shadow Land Rents with Congestion, " *American Economic Review* 68, 588-600.

Kanemoto, Y., (1975), "Congestion and Cost-Benefit Analysis in Cities, " *Journal of Urban Economics* 2, 246-264.

Kanemoto, Y., (1976), "Optimum, Market and Second-Best Land Use Patterns in a von Thünen City with Congestion, " *Regional Science and Urban Economics* 6, 23-32.

Kanemoto, Y., (1977a), "Cost-Benefit Analysis and the Second Best Land Use for Transportation, " *Journal of Urban Economics* 4, 483-503.

Kanemoto, Y., (1977b), *Theories of Urban Externalities*, Ph.D. thesis, Cornell University.

Pines, D. and E. Sadka, (1979), "Optimum, Second-Best and Market Allocations of Resources within an Urban Area, " *Journal of Urban Economics*, forthcoming.

Solow, P.M., (1973), "Congestion Cost and the Use of Land for Streets, " *The Bell Journal of Economics and Management Science* 4, 602-618.

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

Wan, F.Y.M., (1979), "Accurate Solutions for the Second Best Land Use Problem, " Technical Report No. 79-30, Institute for Applied Mathmatics and Statistics, University of British Columbia.

Wheaton, W.C., (1978), "Price Induced Distortion in American Highway Investment, " *Bell Journal of Economics* 9, 622-632.

CHAPTER VI

NEIGHBOURHOOD EXTERNALITIES AND A CUMULATIVE DECAY PROCESS

Whether we like it or not, people often believe they suffer external costs from the presence of some other type of people in their neighbourhood: the rich may fear heavier taxes if poorer households live in the same municipality; whites may not like to live close to blacks; Greeks may believe that their daughters are not safe if there are too many scots in a neighbourhood; and so on. Whether real or imaginary, such externalities raise many issues, some of which are more political or moral than economic. One of the fundamental issues that arise in the context of externalities between different races is whether we approve preferences of individuals who are racially prejudiced: some societies do not, and force individuals to act against their preferences. A typical example is the "busing" regulation in American cities, where school children in a racially segregated area are "bused" to a school at a distant location in order to have racially mixed schools.

Although these issues are extremely important, we concentrate on the economic consequences of the externalities and avoid moral or political judgements. We also restrict ourselves to what might be termed passive discrimination: the well being of discriminators is affected by the locational decisions of others, but discriminators are unable to influence the decisions of others. The reader must be aware that the problem analyzed in this chapter has other important aspects.

We first examine the stability of spatial residential patterns. We find that externalities introduce a tendency toward segregation by type: individuals who suffer an externality from the presence of individuals of another group tend to cluster together to avoid the externality.

We next consider a special kind of a dynamic problem which arises in a city with externalities between different types of households. This analysis is motivated by the experience of American cities in 1960's and 70's. American cities have experienced extensive migration of the middle class households from central cities to the suburbs. Explanations of this phenomenon can be roughly classified into the following two types. The first type sees the migration as an equilibrium process. As the income level rises

and commuting costs fall due to technological progress in transportation, the population density gradient becomes flatter in equilibrium. The population in the suburbs, therefore, increases relative to that in central cities. The population increase in the suburbs consists of wealthier families because, for a variety of reasons, richer families have a tendency to live farther from the center.[1]

The second type focuses on the deterioration of central cities that accompanied the out-migration of the middle class. This type explains the process as one of cumulative decay: the deterioration of central cities drives out wealthier residents and so lowers per capita income, and the reduction of per capita income leads to further deterioration. The central city deteriorates cumulatively until it eventually reaches a new equilibrium state. The process of middle class out-migration is thus viewed as a disequilibrium rather than equilibrium process.

In our treatment the decay process appears as a problem of the stability of the boundary between rings of different types of households. When the previously stable boundary becomes unstable as a result of a change in some exogenous factor, a rapid movement of the boundary occurs. The shift to a new stable equilibrium can be interpreted as the cumulative decay process: an increase of one type of households increases the external costs for the other type, causing them to move away and inducing a further increase of the first type.

In section 1 we formulate a model with two types of households, one of which receives a higher income than the other, and also suffer an external cost from the presence of the other. Set up this way, the model can be used to explore the spatial behaviour of 'rich' and 'poor'. Stability of different spatial patterns is examined in section 2. In section 3 we analyze stability of the boundary between the two types, allowing for migration into and out of the city. The possibility of a cumulative process is considered in section 4 and several examples are examined in section 5.

1. The Model

Consider a single-centered city whose residents consist of two types of households that we can call discriminators and nondiscriminators. Discriminators suffer external diseconomy if they live close to nondiscriminators.

---

[1] For example, since there are more newer houses in the suburbs, the quality of housing is better in the suburbs. A trade-off between commuting costs and housing also works in favour of the suburban locations of richer households, as seen in Chapter I.

In contrast to our method in previous chapters, we assume that the city stands ready built: houses with certain qualities and lot sizes are already built in the city and the characteristics of houses do not change during the time interval relevant to our analysis.

It is not difficult to relax this assumption and consider the case of malleable housing capital: although the analysis becomes quite tedious, the results are basically the same. The present formulation is preferable because housing capital is in fact quite durable and we are concerned with short-run phenomena. The only serious problem arises at the boundary of the city, where new houses must be built when the city expands. Since we assume that houses are readily available even outside the current boundary of the residential zone, the expansion of the boundary occurs instantaneously in our model. In reality, however, new construction takes time and our results should not be taken too literally. We discuss the problem in the end of section 4.

$h(x)$ denotes the services provided by a house and lot at distance x from the center. Since houses are usually larger farther from the center,

$$h'(x) > 0. \tag{1.1}$$

Note that in this chapter $h(x)$ denotes the services from both land and buildings, rather than the lot size as in previous chapters.

There are $N(x)\,\mathrm{dx}$ houses in the ring between x and $x+dx$, where we assume that $N(x)$ does not decrease as distance from the center increases:

$$N'(x) \geq 0. \tag{1.2}$$

This assumption requires that the width of the residential zone, $L_H(x)$, increases faster than the lot size with distance from the center. It precludes the case of a linear city when the lot size increases with distance.

The opportunity cost of a unit of housing services is assumed to be a constant R_a. In equilibrium the rent at the edge of the city must equal R_a:

$$R(\bar{x}) = R_a. \tag{1.3}$$

Since we assumed that ready-built houses are standing outside the edge of the city, we may take R_a equal to zero. In order to include other possibilities, however, we do not specify the value of R_a in the following analysis.

We want to know how the two groups of households distribute themselves over the ready-built houses when there is externality between the two groups. For the sake of simplicity, we analyze an externality that operates in only one direction.

Discrimination may in fact be extremely complex, but this assumption leads to useful insights about the effect of discrimination on city form. The discriminators, denoted by superscript d, suffers external diseconomies from the presence of nondiscriminators, denoted by superscript n: the nondiscriminators do not experience any externality. Thus the utility function of a discriminator at x is

$$u^d(z^d(x), h(x), A(x)) \quad , \tag{1.4}$$

where $A(x)$ denotes the external diseconomy suffered by the discriminator as a result of living near nondiscriminators, and $z^d(x)$ is the consumption of the consumer good. A nondiscriminator at x has a utility function with no externality term:

$$u^n(z^n(x), h(x)) . \tag{1.5}$$

We assume positive marginal utilities of the consumer good and housing for both, and a negative marginal utility of the externality for the discriminator:

$$u_z^d(z^d, h, A) > 0, \quad u_h^d(z^d, h, A) > 0, \tag{1.6}$$

$$u_z^n(z^n, h) > 0, \qquad u_h^n(z^n, h) > 0, \tag{1.7}$$

$$u_A^d(z^d, h, A) < 0, \tag{1.8}$$

where the subscripts z, h, and A denote partial derivatives.

The externality given by a nondiscriminator living at x' to a discriminator at x is $a(|x - x'|)$. The function a ( ) is nonnegative and nonincreasing,

$$a(|x - x'|) \geq 0, \tag{1.9}$$

$$a'(|x - x'|) \leq 0, \tag{1.10}$$

and $|x - x'|$ is the absolute value of $x - x'$. The total external diseconomies received by a discriminator at x is the sum of diseconomies generated by all nondiscriminators:

$$A(x) = \int_0^\infty a(|x - x'|) N^n(x') dx', \tag{1.11}$$

where $N^n(x')dx'$ is the population of nondiscriminators between x' and $x' + dx'$. If

we imagine that the residential zone is circular, (1.11) implies that a nondiscriminator at the same radius, but on the opposite side of the city, induces a larger externality than one very near by but at a slightly different radius. Although this oddity disappears in a linear city, it may affect the generality of the results that follow.

We can now analyze the city forms arising from discrimination if we specify the budget constraints of discriminators and nondiscriminators. Choosing the case which is most common, and probably therefore most interesting, we assume that discriminators are richer than nondiscriminators. A rich discriminator earns an income y^d and pays the commuting costs $t^d(x)$. The budget constraint is

$$y^d = z^d(x) + R(x)h(x) + t^d(x), \tag{1.12}$$

where $R(x)$ is the rent of a unit amount of housing services at x. A poorer nondiscriminator earns a lower income y^n and pays lower commuting costs $t^n(x)$:

$$y^d > y^n \tag{1.13}$$

$$t^{d\prime}(x) > t^{n\prime}(x), \qquad 0 \le x \le \bar{x}, \tag{1.14}$$

$$t^d(0) = t^n(0) = 0. \tag{1.15}$$

Lower commuting costs for a nondiscriminator may be considered as representing lower time costs. Introducing different transportation costs complicates the analysis slightly, as the discussion of the assumption expressed by equation (1.30) below shows. There are, however, gains in realism and in generality which compensate for the additional complexity. The budget constraint for a nondiscriminator is

$$y^n = z^n(x) + R(x)h(x) + t^n(x). \tag{1.16}$$

We assume that neither a discriminator nor a nondiscriminator owns a house in the city. Our model, therefore, corresponds to the absentee-landlord case in Chapter I, with landlords that do not discriminate.

By spatial arbitrage, all households in each group receive equal utility levels in equilibrium:

$$u^d = u^d(z^d(x), h(x), A(x)), \tag{1.17}$$

$$u^n = u^n(z^n(x), h(x)). \tag{1.18}$$

By the assumption of positive marginal utilities, (1.6) and (1.7), these equations can be uniquely solved for z^d and z^d to obtain demand functions for the consumer good,

$$z^d(x) = z^d(h(x), u^d, A(x)), \tag{1.19}$$

$$z^n(x) = z^n(h(x), u^n), \tag{1.20}$$

where

$$z_h^d(h(x),u^d,A(x)) = -u_h^d / z_z^d < 0, \tag{1.21}$$

$$z_u^d(h(x),u^d,A(x)) = 1/u_z^d > 0, \tag{1.22}$$

$$z_A^d(h(x),u^d,A(x)) = -u_A^d / u_z^d > 0, \tag{1.23}$$

and

$$z_h^n(h(x),u^n) = -u_h^n / u_z^n < 0, \tag{1.24}$$

$$z_n^n(h(x),u^n) = 1/u_z^n > 0. \tag{1.25}$$

Substituting (1.19) and (1.20) into (1.12) and (1.16) respectively, we obtain the *bid rent functions*:

$$\begin{aligned} R^d(x) &= \frac{1}{h(x)}[y^d - z^d(h(x),u^d,A(x)) - t^d(x)] \\ &\equiv R^d[I^d(x),u^d,h(x),A(x)], \end{aligned} \tag{1.26}$$

$$\begin{aligned} R^n(x) &= \frac{1}{h(x)}[y^n - z^n(h(x),u^n) - t^n(x)] \\ &\equiv R^n[I^n(x),u^n,h(x)], \end{aligned} \tag{1.27}$$

where

$$I^d(x) \equiv y^d - t^d(x), \tag{1.28}$$

$$I^n(x) \equiv y^n - t^n(x). \tag{1.29}$$

The bid rent functions in this chapter are slightly different from those in other chapters, since $h(x)$ appears in the bid rent functions. A household must take the amount of housing services as given and the only variable a household can choose is the location of a house. It is important to notice that this implies the marginal rate of substitution between housing and the consumer good need not equal the bid rent.

Since the externality $A(x)$ depends on how the nondiscriminators are distributed over space, we must know the locational patterns of the nondiscriminators to obtain the bid rent of the discriminator. The bid rent function of the discriminators, however, influences the spatial distribution of the nondiscriminators. This spatial interrelationship is the only complication in our model.

The following assumption plays a crucial role in determining the stable residential pattern:

$$R_I^n I^{n\prime}(x) + R_h^n h'(x) = t^{n\prime}(x) + z_h^n(h(x), u^n)h'(x)$$

$$< t^{d\prime}(x) + z_h^d(h(x), u^d, A(x))h'(x)$$

$$= R_I^d I^{d\prime}(x) + R_h^d h'(x), \qquad (1.30)$$

for any relevant range of x, A, u^d, and u^n. This assumption is made to ensure that, if the rich did not discriminate, they would have a flatter bid rent curve and live farther from the center than the poor, as in Chapter I.

From (1.21) and (1.24), we can rewrite (1.30) as

$$\frac{u_h^d(z^d(x), h(x), A(x))}{u_z^d(z^d(x), h(x), A(x))} - \frac{u_h^n(z^n(x), h(x))}{u_z^n(z^n(x), h(x))}$$

$$> \frac{1}{h'(x)}[t^{d\prime}(x) - t^{n\prime}(x)] > 0, \qquad (1.31)$$

where the last inequality is obtained from (1.14). The condition can now be interpreted in terms of two opposing forces. First, since discriminators have higher transportation costs, they tend to live closer to the center. Second, if they have a higher marginal rate of substitution between housing and the consumer good than nondiscriminators - if they are willing to give up more of the consumer good for a marginal increase in housing services -, then there is an opposing tendency for discriminators to live in larger houses farther from the center of the city. Our assumption requires that the latter tendency overwhelm the former.

The difference between the marginal rates of substitution between housing and the consumer good is closely related to the normality of housing. Roughly speaking, condition (1.31) is satisfied if housing is a normal good and the normality is strong enough to offset the greater transportation costs of discriminators.[2]

---

[2] This statement is precisely true if we assume a utility function which is separable and can be written

$$u^d(z^d, h, A) = U(u^n(z^d, h), A).$$

Given the above functional form, a discriminator has exactly the same preferences over housing and the consumer good as a nondiscriminator, and the preferences are not affected by the externality. Consider

2. Stability of Spatial Patterns

In the absence of externalities, assumption (1.30) assures that the bid rent of the rich will be flatter than that of the poor, and the poor therefore live closer to the center of the city. It can be shown that when the rich suffer external diseconomies, the pattern is unaffected *if the number of housesper unit distance is constant*. This qualification is required because our externality function (1.11) employs only radial distances. If the number of houses per unit distance increases with distance, the assumption (1.30) must be strengthened.

When the number of houses per unit distance is constant,

$$N'(x) = 0 \quad , \qquad\qquad 0 \leq x \leq \bar{x} \, . \qquad\qquad (2.1)$$

We assume there is no *active* discrimination in the housing market: neither discriminators nor landlords try to influence where nondiscriminators live.

To see that only the central location of nondiscriminators is stable, we examine each of the possible configurations. The pattern where both the rich discriminators and the poor discriminators live at a same distance from the center is unstable.

---

a hypothetical problem of choosing both h and z under the budget constraint, $I = z + Rh$. Because of the separability, the choice of a discriminator is the same for any level of externality. Moreover, both types behave in exactly the same way, and have the same uncompensated demand function for housing, $\hat{h}(I,R)$. As in (I.2.7), the uncompensated demand functions satisfy

$$\begin{aligned} \hat{h}_I(I,R) &= \frac{u^i_z}{D}\left[u^i_{hz}u^i_z - u^i_h u^i_{zz}\right] \\ &= \frac{(u^i_z)^3}{D}\frac{\partial(u^i_h/u^i_z)}{\partial z}, \end{aligned}$$

where

$$D = 2u^i_{hz}u^i_z u^i_h - (u^i_h)^2 u^i_{zz} - (u^i_z)^2 u^i_{hh} \geq 0, \qquad\qquad i = n, d.$$

Now if housing is a normal good, we have

$$\frac{\partial(u^n_h/u^n_z)}{\partial z} > 0.$$

Since $z^d(x) > z^n(x)$ from $y^d > y^n$, this implies that

$$\frac{u^d_h(z^d(x),h(x),A(x))}{u^d_z(z^d(x),h(x),A(x))} = \frac{u^n_h(z^d(x),h(x))}{u^n_z(z^d(x),h(x))} > \frac{u^n_h(z^n(x),h(x))}{u^n_z(z^n(x),h(x))},$$

and if the normality is strong enough, (1.31) is satisfied.

Consider a zone at radius x where both the rich and the poor locate. Under the assumption that there is no active discrimination in the housing market, both must pay the same rent at x, and therefore their bid rents must be equal there. Since, by assumption (1.11), the strength of the externality depends only on the radial distance between a discriminator and all nondiscriminators, any increase in the number of nondiscriminators at x drives down the bid rent of the discriminators by increasing the externality. This induces a further increase of the number of nondiscriminators because the bid rent of the nondiscriminators remains the same, and the nondiscriminators outbid the discriminators. The process continues until the zone is filled with nondiscriminators.[3]

It is convenient to introduce a formula which will tell us the relative levels of bid rents of discriminators and nondiscriminators at x'' if we know their relative positions at x'. Since it is simpler to work with the bid rent on a house and lot, $E(x) = R(x)h(x)$, than with the bid rent per unit amount of housing services, $R(x)$, we rewrite (1.26) and (1.27) as

$$E^d(x) \equiv R^d(x)h(x) = y^d - z^d(h(x), u^d, A(x)) - t^d(x),$$

and

$$E^n(x) \equiv R^n(x)h(x) = y^n - z^n(h(x), u^n) - t^n(x).$$

In order to isolate the effect of the externality, we consider the difference between the slopes of $E^d(x)$ and $E^n(x)$ at x between x' and x'', fixing the level of the externality at $A(x'')$:

$$H(x; x'') \equiv z_h^n(h(x), u^n)h'(x) + t^{n\prime}(x)$$

$$-[z_h^d(h(x), u^d, A(x''))h'(x) + t^{d\prime}(x)]$$

$$> 0, \tag{2.2}$$

where the inequality follows from assumption (1.30). We then obtain

$$[E^d(x'') - E^n(x'')] - [E^d(x') - E^n(x')]$$

$$= h(x'')[R^d(x'') - R^n(x'')] - h(x')[R^d(x') - R^n(x')]$$

---

[3] Note that this result crucially depends on our assumption (1.11) that the strength of the externality depends only on the radial distance. It is still an open question whether the result carries over to the case where the externality depends also on circumferential distance from a nondiscriminator.

$$= \int_{x'}^{x''} H(x; x'')dx + J(x', x''), \tag{2.3}$$

where $J(x', x'')$ captures the effect of the difference in the externality between x' and x'':

$$J(x', x'') = z^d[h(x'), u^d, A(x')] - z^d[h(x'), u^d, A(x'')]. \tag{2.4}$$

From (1.23), $J(x', x'')$ satisfies

$$J(x', x'') \begin{matrix} > \\ = \\ < \end{matrix} 0 \quad \text{as} \quad A(x') \begin{matrix} > \\ = \\ < \end{matrix} A(x''). \tag{2.5}$$

Now consider the case illustrated in Fig.1 where the zone of nondiscriminators extends from x^* to x^{**}, between two zones of rich discriminators. In equilibrium the bid rent of the two groups must be equal at the two borders, since there is no price discrimination in the housing market. Suppose that two bid rents are equal at the inner boundary, x^*, as in Fig.1. From (2.1) the external diseconomy is the same at two boundaries:

$$A(x^*) = A(x^{**}).$$

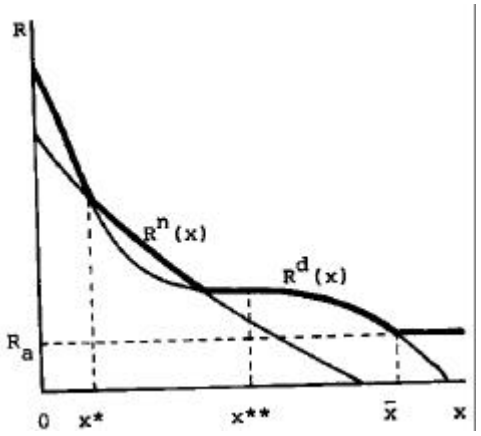

Figure 1
The intermediate location of nondiscriminators

If we set $x' = x^*$ and $x'' = x^{**}$, (2.3) becomes

$$h(x^{**})\left[R^d(x^{**}) - R^n(x^{**})\right] = \int_{x^*}^{x^{**}} H(x; x^{**})dx > 0,$$

which implies that the bid rent of discriminators is higher than that of nondiscriminators at the boundary. Therefore, discriminators outbid nondiscriminators in the neighbourhood of the outer boundary and the boundary moves closer to the center.

Thus the intermediate location of nondiscriminators cannot be an equilibrium.

The same reasoning can also be applied to a city which has only two zones, nondiscriminators living in the outer zone and discriminators living in the inner zone.

When nondiscriminators live in more than one zone, denote the borders of nondiscriminators' zone farthest from the center by $x*$ and $x**$. Suppose the two bid rents are equal at $x*$. The externality is stronger at the inner boundary than at the outer boundary, since the inner boundary is closer to other zones of nondiscriminators. Hence, $J(x*, x**)$ is positive and

$$h(x^{**})\left[R^d(x^{**}) - R^n(x^{**})\right]$$
$$= \int_{x^*}^{x^{**}} H(x; x^{**})dx + J(x^*, x^{**}) > 0.$$

This case is not an equilibrium, either.

Finally, consider the case of the central location of nondiscriminators. Let $x*$ be the boundary between the zones of nondiscriminators and rich discriminators as in Fig.2. In equilibrium, the bid rents are the same at the boundary:

$$R^d(x^*) = R^n(x^*).$$

For any $x' < x*$, we have

$$A(x') \geq A(x^*).$$

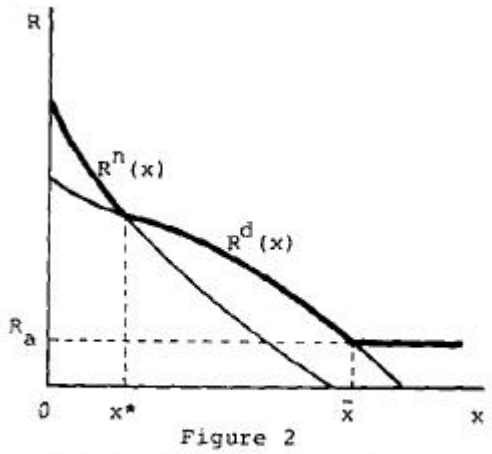

Figure 2
The central location of nondiscriminators

Substituting $x*$ for x'' in (2.3), we obtain

$$-h(x')\left[R^d(x')-R^n(x')\right]=\int_{x'}^{x^*}H(x;x^*)dx+J(x',x^*)>0.$$

Hence, $R^d(x')<R^n(x')$ for any $x'<x^*$ and nondiscriminators outbid discriminators inside the boundary.

At any point, x', outside the boundary, discriminators have a higher bid rent:

$$h(x')\left[R^d(x')-R^n(x')\right]=\int_{x^*}^{x'}H(x;x')dx+J(x'',x')>0,$$

since $A(x')\leq A(x^*)$.

Thus the central location of nondiscriminators is an equilibrium. Since the cases considered here exhaust all the possibilities, the central location of nondiscriminators is the only stable market equilibrium under the assumption (1.30) and (2.1). This result shows that the existence of externalities does not alter the spatial pattern when (1.30) and (2.1) hold. No discrimination is, therefore, necessary to confine nondiscriminators in the central part of the city. Moreover, the external diseconomy makes the segregated pattern more stable since the bid rent curve of discriminators becomes flatter. Note that it is not the strength of the externality that makes the bid rent curve of discriminators flatter, but the fact that externality *diminishes* with distance. It is easy to see that if externality is uniform in the city, no change in the slope of the bid rent curve occurs.

We have shown that passive discrimination of the sort we have modeled can explain the spatial distribution of racial groups, blacks in American cities for example, when the group discriminated against is uniformly poorer than the discriminators. This result does not suggest that *active* discrimination does not exist. Recent studies support the view that there is in fact active discrimination in the housing market of American cities.

If the number of houses per unit distance increases with distance from the center, $N'(x)>0$, the above result must be modified. In order for the central location of nondiscriminators to be a unique stable configuration, the inequality (1.30) must be strengthened to

$$z_h^d h'(x)+t^{d\,\prime}(x)>z_h^n h'(x)+t^{n\,\prime}(x)+\varepsilon\,, \tag{2.6}$$

for some large enough $\varepsilon>0$. The problem arises because our externality function (1.11) employs only radial distance. If there are more households per unit distance at larger radii, the externality will be higher at the outer boundary than at the inner

boundary, which causes an additional tendency to lower the bid rent of discriminators at the outer boundary. The inequality (1.30), therefore, must be strong enough to offset this effect. In the rest of the chapter, (2.6) is assumed to hold for a sufficiently large ε so that the only stable configuration is the central location of nondiscriminators.

3. The Boundary Bid Rent Curves

In section 2 we established the existence of a single boundary between two types of households, with the rich discriminators living farthest from the center. In section 4 we will examine the stability of the boundary between the two zones, but in order to do so we develop an additional concept, the *boundary bid rent curve.* The boundary bid rent curve is the bid rent at x when the boundary is at x.

Assume that the city is open: migration into and out of the city is free and costless. The utility levels of rich discriminators and poor nondiscriminators in the city, u^d and u^n, then equal the corresponding utility levels in the rest of the world, V^d and V^n, respectively. The utility levels, however, are not necessarily fixed. An increase in the population of the city is accompanied by a decrease in the population of the outside world, which causes a rise in the utility level in the outside world because of diminishing returns. We assume that the general utility level of discriminators is a nondecreasing function of the population of discriminators in the city, and, that the same is true for nondiscriminators.[4]

$$u^d = V^d(P^d), \tag{3.1}$$

$$u^n = V^n(P^n), \tag{3.2}$$

where

$$V^{d\prime}(P^d) \geq 0, \tag{3.3}$$

$$V^{n\prime}(P^n) \geq 0, \tag{3.4}$$

and P^d and P^n are respectively the populations of discriminators and

[4] In general, the utility level of discriminators (and also nondiscriminators) depends on the populations of both discriminators and nondiscriminators. For simplicity, we assume that the population of one type has no effect on the utility level of the other type. We make a similar assumption for income levels in (3.5) and (3.6) below.

nondiscriminators in the city. $V^d{}'(P^d)$ and $V^n{}'(P^n)$ are almost zero if there are so many people of each type in the outside world that an additional individual does not cause any significant change in allocation there. Since our formulation implicitly assumes that the population of the city is small enough for an additional individual to matter within the city, this in effect requires that the city is small compared with the rest of the world. Roughly speaking, therefore, $V^d{}'(P^d)$ and $V^n{}'(P^n)$ are zero if the city is small, and increase for cities which are larger relative to the rest of the world.

The income of a city resident also depends on the population of the city. This reflects two factors. First, if prices of products are constant, the wage rate falls due to diminishing returns as the population increases. Second, when the population increases, production expands, which reduces prices of products in the world market. This also causes a decrease in wage rate. We therefore assume that the income of each type of household is a nonincreasing function of the population of that type in the city,

$$y^d = y^d(P^d), \tag{3.5}$$

$$y^n = y^n(P^n). \tag{3.6}$$

where

$$y^d{}'(P^d) \leq 0, \tag{3.7}$$

$$y^n{}'(P^n) \leq 0. \tag{3.8}$$

$y^d{}'(P^d)$ and $y^n{}'(P^n)$ are smaller in absolute value in a smaller city, since the effects on the world prices are smaller by the same argument as we applied to the case of $V^d{}'(P^d)$ and $V^n{}'(P^n)$.

Let x^* denote the boundary between the zones of discriminators and nondiscriminators. Then

$$P^n(x^*) = \int_0^{x^*} N(x)dx, \tag{3.9}$$

$$P^d(x^*) = \int_{x^*}^{\bar{x}} N(x)dx, \tag{3.10}$$

where $\bar{x}$ is determined so that the highest bid rent equals the rural rent, R_a, at the

edge of the city.

Now, we express bid rent functions as functions of x^* using (3.9) and (3.10). The bid rent, $R^n(x;x^*)$, of a nondiscriminator at x when the boundary is at x^* is

$$R^n(x;x^*)=\frac{1}{h(x)}\left\{y^n(P^n(x^*))-z^n\left[h(x),V^n(P^n(x^*))\right]-t^n(x)\right\}. \tag{3.11}$$

The slope of the bid rent curve is

$$\begin{aligned} R_x^n(x;x^*) &\equiv \frac{\partial}{\partial x}R^n(x;x^*) \\ &= -\frac{1}{h(x)}\left[z_h^n h'(x)+t^{n\prime}(x)\right]-R^n(x)\frac{h'(x)}{h(x)} \end{aligned} \tag{3.12}$$

The location of the boundary enters this formulation, not because nondiscriminators discriminate, but because the location of the boundary determines P^n, which affects income and utility levels.

The bid rent of discriminators depends in addition on the externality that they suffer from nondiscriminators. The externality received by a discriminator at x is

$$A(x;x^*)=\int_0^{x^*} a(|x-x'|)N(x')dx'. \tag{3.13}$$

The bid rent of discriminators is then

$$\begin{aligned} &R^d(x;x^*) \\ &=\frac{1}{h(x)}\left\{y^d(P^d(x^*))-z^d\left[h(x),A(x;x^*),V^d(P^d(x^*))\right]-t^d(x)\right\} \end{aligned} \tag{3.14}$$

Its slope is

$$R_x^d(x;x^*)=-\frac{1}{h(x)}\left[z_h^d h'(x)+t^{d\prime}(x)+z_A^d A_x\right]-R^d(x)\frac{h'(x)}{h(x)}. \tag{3.15}$$

A_x is defined as

$$A_x\equiv\frac{\partial}{\partial x}A(x;x^*)=\int_0^{x^*}\operatorname{sgn}(x-x')a'(|x-x')N(x')dx', \tag{3.16}$$

where

$$\operatorname{sgn}(x-x')=\begin{cases}+1 & \text{if} \quad x-x'\geq 0,\\ -1 & \text{if} \quad x-x'<0.\end{cases}$$

A_x is nonpositive at least when x is greater than or equal to x^*. The externality, therefore, tends to make the bid rent curve of discriminators flatter. It follows from

assumption (1.30) that the bid rent curve of discriminators is flatter than that of nondiscriminators at the boundary:

$$\begin{aligned} R_x^d(x^*;x^*) &= -\frac{1}{h(x^*)}\left[z_h^d h'(x^*) + t^d{}'(x^*) + z_A A_x\right] - R^d(x^*)\frac{h'(x^*)}{h(x^*)} \\ &\geq -\frac{1}{h(x^*)}\left[z_h^d h'(x^*) + t^d{}'(x^*)\right] - R^d(x^*)\frac{h'(x^*)}{h(x^*)} \\ &\geq -\frac{1}{h(x^*)}\left[z_h^n h'(x^*) + t^n{}'(x^*)\right] - R^n(x^*)\frac{h'(x^*)}{h(x^*)} \\ &= R_x^n(x^*;x^*) \end{aligned} \tag{3.17}$$

This confirms the result in the preceding section that nondiscriminators live closer to the center.

At the edge of the city, the bid rent of discriminators must equal the rural rent:

$$R^d(\bar{x};x^*) = R^a, \tag{3.18}$$

which determines $\bar{x}$ as a function of $x*$ and hence $P^d(x^*)$ in (3.10).

Next, we introduce the concept of the *boundary bid rent curve*, which is the bid rent at x when the boundary is at x. The boundary bid rent curves will play a crucial role in the analysis of a cumulative process. For nondiscriminators it is

$$\hat{R}^n(x) = R^n(x;x), \tag{3.19}$$

and from (3.11) its slope is

$$\begin{aligned} \hat{R}^n{}'(x) &= R_x^n(x;x) + R_{x^*}^n(x;x) \\ &= R_x^n(x;x) - \left[z_u^n v^n{}'(P^n) - y^n{}'(P^n)\right]\frac{N(x)}{h(x)} \\ &\leq R_x^n(x;x), \end{aligned} \tag{3.20}$$

where

$$R_{x^*}^n(x;x^*) \equiv \partial R^n(x;x^*)/\partial x^*.$$

Thus the boundary bid rent curve is steeper than the bid rent curve. An expansion of the boundary is possible only if the population of nondiscriminators increases in the city. This raises the utility level in the outside world and lowers the income level in the city, causing a fall in the bid rent curve. The relationship between the bid rent

curve and the boundary bid rent curve is illustrated in Figure3.

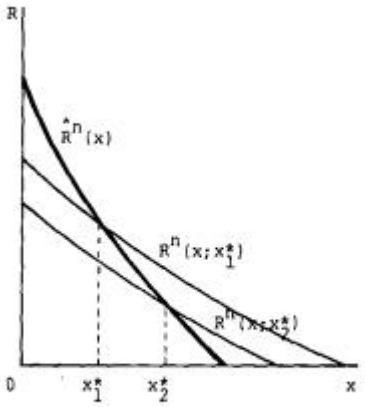

Figure 3
The relationship between the boundary bid rent curve and the bid rent curves

The boundary bid rent of discriminators is

$$\hat{R}^d(x) = R^d(x;x), \tag{3.21}$$

with a slope

$$\begin{aligned}&\hat{R}^{d\prime}(x)\\&= R_x^d(x;x) - \frac{1}{h(x)}\left\{z_A^d A_{x^*}(x;x) + \left[z_u^d v^{d\prime}(P^d) - y^{d\prime}(P^d)\right]P^{d\prime}(x)\right\}\end{aligned} \tag{3.22}$$

where

$$A_{x^*}(x;x^*) \equiv \frac{\partial}{\partial x^*} A(x;x^*) \tag{3.23}$$

and hence

$$A_{x^*}(x;x) = a(0)N(x) \geq 0\,. \tag{3.24}$$

Whether the slope of the boundary bid rent curve of the discriminators is steeper than that of the bid rent curve is not clear. An outward movement of the boundary acts on the bid rent curve of discriminators in two opposing ways. The increased population of nondiscriminators drives up the externality causing the bid rent to fall. As will be shown, however, $P^{d\prime}(x)$ is usually negative: the population of discriminators decreases as the boundary moves outward, lowering the utility for discriminators in the outside world, increasing their income in the city and tending to

cause their bid rent to rise. The boundary bid rent curve of discriminators is therefore either flatter or steeper than the bid rent curve depending on which tendency is stronger.

The first term in the square bracket of (3.22), $z_A^d A_{x^*}(x;x)$, is positive from (3.24), since z_A^d is positive from (1.23). The second term is more complicated. $P^d{}'(x)$ can be obtained by differentiating (3.10) and (3.18).

$$P^d{}'(x) = -\frac{z_A^d\left[N(\bar{x})A_{x^*}+N(x)A_x\right]+N(x)\left[t^d{}'(\bar{x})+(R_a+z_h^d)h'(\bar{x})\right]}{\left[z_u^d v^d{}'(P^d)-y^d{}'(P^d)\right]N(\bar{x})+t^d{}'(\bar{x})+z_A^d A_x(\bar{x};x)+(R_a+z_h^d)h'(x)} \quad . \quad (3.25)$$

The first square bracket on the numerator is positive under the assumption that $N'(x)\geq 0$, since

$$\begin{aligned} & N(\bar{x})A_{x^*}(\bar{x};x)+N(x)A_x(\bar{x};x) \\ & = N(x)\left[a(\bar{x})N(\bar{x})+\int_0^x a'(|\bar{x}-x'|)(N(x')-N(\bar{x}))dx'\right] \qquad (3.26) \\ & > 0. \end{aligned}$$

The first term in the second square bracket of the numerator of (3.25) is positive but the second term may be negative. The second term is zero if the marginal rate of substitution between housing and the consumer good equals the bid rent, which is the case if $h(x)$ can be freely chosen. If houses are newly constructed at the edge of the city, $h(\bar{x})$ may be optimize It is, therefore, plausible to assume that the magnitude of the second term is small. Thus, the numerator tends to be positive.

The first two terms of the denominator are positive. The third term is nonpositive but the magnitude is small since the externality is weak at the edge of the city. The fourth term is also small since $R_a + z_h^d$ is small as argued above. Therefore, the denominator also tends to be positive and $P^d{}'(x)$ is likely to be negative.

The reason for this result is roughly as follows. If the zone of nondiscriminators expands, the city must expand to accommodate the same population of discriminators. Consider the effects on a discriminator at the edge of the city. There are three major effects: commuting costs increase, the strength of the externality increases since there are more nondiscriminators in the city, and the boundary shifts outward to where houses are larger, by the assumption that $h'(x) > 0$. The first two effects tend to lower the utility level of the discriminator, but the direction of the third effect depends on whether houses are larger or smaller than the optimum at the edge of the city. If houses are smaller than the optimum, the third effect tends to raise the utility level. Since the

third effect disappears when $h(x)$ is optimized, the first two effects are likely to be dominant, and the utility level declines as x^* increases. This induces emigration of discriminators, resulting in a decrease in the population of discriminators in the city.

4. Stability of the Boundary and a Cumulative Process

Next, we examine stability of the boundary between the zones of rich discriminators and poor nondiscriminators. It is easy to show that, if the boundary bid rent of discriminators is less steep than that of nondiscriminators, the boundary is stable, and if steeper, the boundary is unstable. Consider the situation represented by Fig.4b. The *boundary* is at x^*, and beyond x^* the discriminators outbid the nondiscriminators. The boundary bid rent of the discriminators is steeper, however, as illustrated in Fig.4a. Notice that, if the boundary x^* is to be an equilibrium, the boundary bid rents must be equal there as well as the bid rents.

Figure 4. Stability of Boundaries

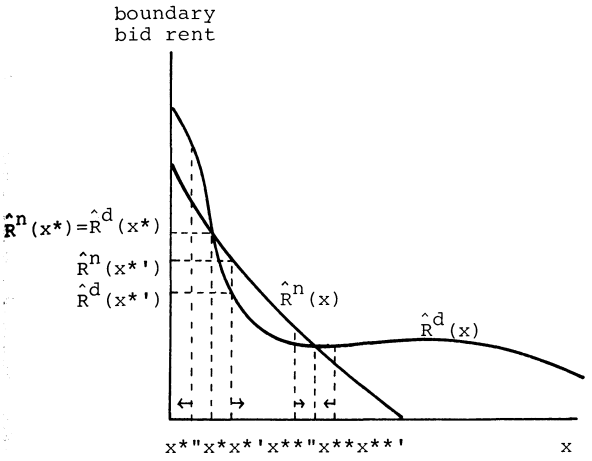

Figure 4a

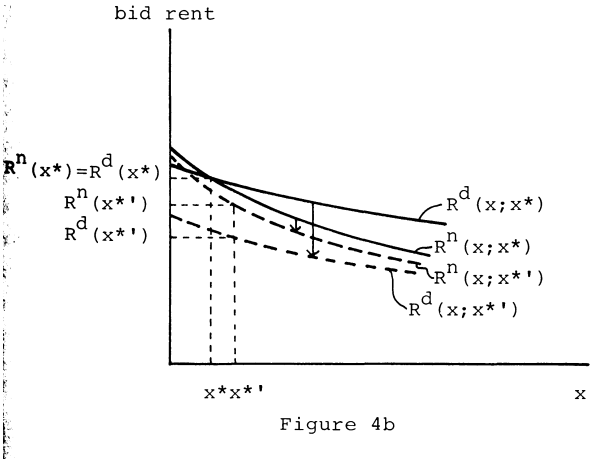

Figure 4b

Now imagine that the boundary shifts outward to $x^{*\prime}$ because of some random disturbances. The bid rent of discriminators falls farther than the bid rent of nondiscriminators, as in Fig.4b. Nondiscriminators outbid discriminators at the new boundary and the boundary moves farther outward. The process continues until the boundary reaches x^{**}.

If the boundary had shifted inward, discriminators would have outbid nondiscriminators, causing the boundary to move inward until it reached the center, x^* is therefore unstable. The same argument applied at x^{**} will show that the boundary is stable at that point.

We have seen that the bid rent curve of discriminators is flatter than that of nondiscriminators at the boundary. As shown in the preceding section, the boundary bid rent curve of nondiscriminators is steeper than their bid rent curve, and the boundary bid rent curve of discriminators is flatter than their bid rent curve if the externality is weak. In order to have an unstable equilibrium, therefore, the externality must be strong.

We next examine the condition for an unstable equilibrium more carefully. The difference between the slopes of the two boundary bid rent curves is

$$\begin{aligned}\hat{R}^{d\prime}(x)-\hat{R}^{n\prime}(x) &= -\frac{1}{h(x)}\{-H(x;x)+z_A^d\hat{A}'(x) \\ &\quad +[(z_u^dV^{d\prime}-y^{d\prime})p^{d\prime}(x)-(z_u^nv^{n\prime}-y^{n\prime})N(x)]\}\end{aligned} \quad , \qquad (4.1)$$

where

$$\begin{aligned}\hat{A}'(x) &\equiv \frac{d}{dx}A(x;x) \\ &= a(x)N(x)+\int_0^x a'(|x-x'|)(N(x')-N(x))dx' > 0\end{aligned} \quad . \qquad (4.2)$$

The first and third terms in the brace of (4.1) are negative, and the second term positive. Therefore, if the second term is greater than the absolute value of the sum of the first and third terms, the boundary bid rent curve of discriminators is steeper than that of nondiscriminators, and the boundary is unstable. This is more likely to occur if

(a) $H(x;x)$ is smaller: the tendency of the poor to live closer to the center in the absence of the externality is smaller;

(b) $(z_u^dV^{d\prime}-y^{d\prime})P^{d\prime}(x)-(z_u^nV^{n\prime}-y^{n\prime})N(x)$ is smaller: the city is smaller in

comparison with the rest of the world;

(c) z_A^d is bigger: the marginal disutility of the external diseconomy is larger;

(d) $a(x)$ is bigger, which is true if x is smaller, that is, the boundary is closer to the center, or if the externality diminishes less rapidly with distance.[5]

Now consider a historical process in which bid rent shift up due to some exogenous factors such as technological progress. We assume that the bid rent of nondiscriminators rises more rapidly than that of discriminators. This assumption does not necessarily mean that the income of nondiscriminators rises more rapidly than that of discriminators. Even if the income of discriminators were to rise faster than that of nondiscriminators, the bid rent of nondiscriminators might rise faster if the utility level of nondiscriminators attainable in the rest of the world was increasing more slowly. To make our analysis easier, we fix the boundary bid rent of discriminators and allow the boundary bid rent of nondiscriminators to rise over time.

Since the boundary bid rent curve depends on the choice of utility and transportation cost functions and on other parameters of the model, we cannot say much, *a priori*, about its shape. Instead, we illustrate a few examples. If the boundary bid rent curve of nondiscriminators is steeper than

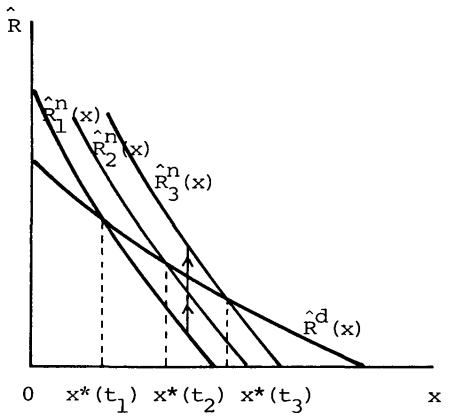

Figure 5. No Cumulative Process

that of discriminators everywhere in the city, we obtain Fig.5. In this case, $x^*(t_1), x^*(t_2)$ and $x^*(t_3)$ are all stable and the boundary gradually shifts outward as the bid rent of nondiscriminators rises.

---

[5] The integral in (4.2) is greater when $N'(x)$ is greater. However, if $N'(x)$ is large, H must be large enough to insure the central location of nondiscriminators, and the net effect is uncertain.

Figure 6 depicts the case where the boundary bid rent curve of discriminators is steeper at the center. At time t_1, $x^*(t_1)$ is an unstable equilibrium and $x^{**}(t_1)$ is a stable equilibrium. If the boundary were to the right of $x^*(t_1)$, it would move to a stable equilibrium at $x^{**}(t_1)$. With the boundary initially at $x=0$, however, no nondiscriminator would enter the city until time t_2, when the boundary bid rent at $x=0$ of nondiscriminators rose as high as that of discriminators. Then any small perturbation would induce a sudden outward shift of the boundary to $x^{**}(t_2)$. Thus a very rapid movement of discriminators to the suburbs occurs after the first nondiscriminator enters the city.

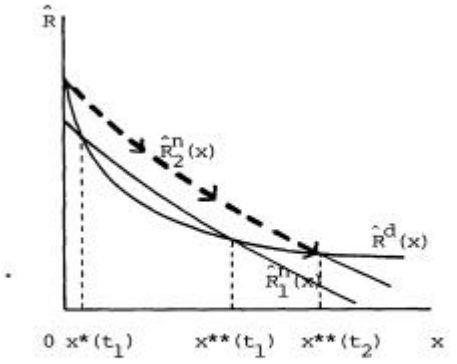

Figure 6
A Cumulative Process at the Center

Finally, consider the case where the boundary bid rent curve of discriminators is flatter than that of nondiscriminators near the center but becomes steeper at some point as in Figure 7. In this case, the boundary gradually moves outward until the bid rent of nondiscriminators becomes tangent to that of discriminators, and then jumps to x^{**}.

Figure 7b illustrates the corresponding bid rent curves. The rapid shift of the boundary is accompanied by a downward shift of both bid rent curves. The bid rent of nondiscriminators must fall because the population of nondiscriminators in the city rises, resulting in an increase in the external utility level by (3.4). Since $u^n = v^n(p^n)$, the utility level of nondiscriminators in the city must also rise, and for utility levels to rise rents must fall. Similarly, since the shifting boundary would usually drive out some discriminators, the utility level of discriminators falls, and rents are likely to rise. Paradoxically, then, the so-called deterioration of the city center may be desirable in terms of income distribution.

In the inner part of the zone of discriminators, the increased externality leads to a fall in the bid rent. In the outer part, however, the rent will usually rise.

The cumulative decay process analyzed by Baumol (1972a, b) and by Oates, Howrey, and Baumol (1971) can be viewed as a rapid movement of the boundary of the sort described in this chapter. If a small increase in the number of nondiscriminators lowers the utility level of discriminators, the discriminators move out to the suburbs, leading to a further deterioration of central cities. This process occurs only if the rent does not fall sufficiently to compensate discriminators for the increase in the external diseconomy, or in our model only if the boundary bid rent curve of nondiscriminators is flatter than that of discriminators.

As discussed in section 1, the fact that the cumulative process is instantaneous in our model depends on the assumption that houses are readily available even outside the current boundary of the city. In reality, however, houses cannot be constructed immediately, and the cumulative process may take

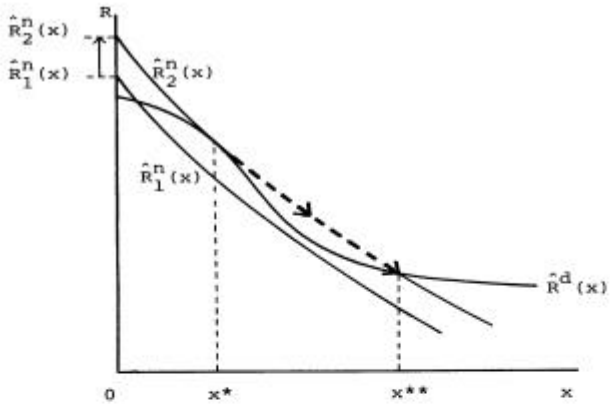

Figure 7a

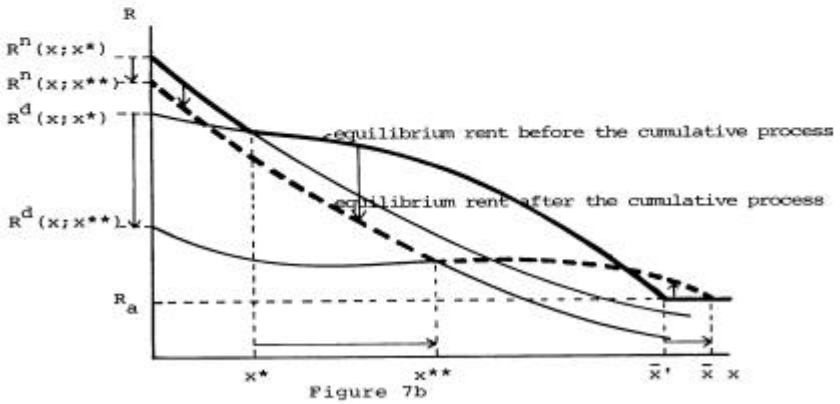

Figure 7b

Figure 7. A Cumulative Process at a Later Stage

quite a long time. It is not the rate of change that characterizes a cumulative decay process, however. The process is cumulative if it can be seen as a disequilibrium process moving towards a new equilibrium, like the boundary shift from x^* to x^{**} in

Figure 7, rather than an equilibrium process.

Notes

Baumol formalized a process of cumulative urban deterioration in section 7 of his paper (1972a). Policy implications of his model were further analyzed by Baumol (1972b) and Oates, Howrey and Baumol (1971). The process of decay is described by two difference equations. One equation embodies the mechanism through which an increase in deterioration leads to a reduction in income per capita in the subsequent period as a consequence of induced emigration to the suburbs, while the other equation describes how the fall in income induces further deterioration. For a suitable set of parameters, these equations obviously have a solution which converges to an equilibrium point and the process toward an equilibrium can be viewed as a cumulative process of urban deterioration.

The weakness of this argument is that individual behaviour and market adjustments are not explicitly considered. For example, we immediately face the following question. Why does the land rent in the central city not fall to keep the wealthier people in the center? If the land rent falls sufficiently, wealthier people will remain even in the deteriorated central city. For a cumulative deterioration process to occur, therefore, something must prevent the land rent from falling enough.

Obviously the rent cannot fall below zero. If it reaches zero, therefore, a cumulative process occurs. In this case deterioration results in vacant houses. Alternatively, poorer households might support the rent. This case can occur in two ways. One is through an increase in *per capita* housing demand by the poorer households and the other is through migration of the poorer households from other areas. Our model in this chapter formalizes the latter case.

Kanemoto (1978) considered the same problem in a simpler model with three discrete regions: the city center, the suburbs, and the rest of the world. The paper explores the case of fiscal burden and the case where one type receives an external *economy* from the other type while generating an external diseconomy. The model in this chapter can easily be extended to include these cases.

We chose not to formulate an explicit dynamic adjustment model because exposition would become tedious, and because the basic results can be explained heuristically, as done in this chapter. Schelling was the first to formulate dynamic models of segregation in the housing market in his papers (1971) and (1972), following his earlier work (1969). Miyao (1978a) extended his analysis, explicitly including the

individual choice of space and location within a city. Kanemoto (1978) also considered a model of dynamic adjustment. Miyao (1978b) considered the same problem in the framework of a probabilistic model of locational choice. These analyses correspond to that in section 2.

Yellin (1974), Rose-Ackerman (1975), (1977), Yinger (1976a, b), and Courant and Yinger (1977) provide static analyses of an externality between different types of households. Yellin has the most general formulation of the externality which we adopted in this chapter.

Although we did not use any results from mathematical theory of catastrophe, our analysis may be cast in that framework. In section 4 we examined how the phase portrait changes as various parameters change. A cumulative process occurs at what is called in catastrophe theory a bifurcation point, where a basic change in the phase portrait occurs: a stable equilibrium becomes an unstable equilibrium.

REFERENCES

Baumol, W.J., (1972a), "Macroeconomics of Unbalanced Growth, " *American Economic Review* 62, 415-426.

Baumol, W.J., (1972b), "The Dynamics of Urban Problems and Its Policy Implications, " in: Peston and Corry (eds.). *Essays in Honour of Lord Robbins*, (International Arts and Sciences Press, New York).

Courant, P.N. and J. Yinger, (1977), "On Models of Racial Prejudice and Urban Residential Structure, " *Journal of Urban Economics* 4, 272-291.

Kanemoto, Y., (1978), "Externality, Migration and Urban Crises, " *Journal of Urban Economics*, forthcoming.

Miyao, T., (1978a), "Dynamic Instability of a Mixed City in the Presence of Neighbourhood Externalities, " *American Economic Review* 68, 454-463.

Miyao, T., (1978b), "A Probabilistic Model of Location Choice with Neighbourhood Effects, " *Journal of Economic Theory* 19, 357-368.

Oates, W.E., E.P. Howrey and W.J. Baumol, (1971), "The Analysis of Public Policy in Urban Dynamic Models, " *Journal of Political Economy* 79, 142-153.

Rose-Ackerman, S., (1975), "Racism and Urban Structure, " *Journal of Urban Economics* 2, 85-103.

Rose-Ackerman, S., (1977), "The Political Economy of a Racist Housing Market, " *Journal of Urban Economics* 4, 150-169.

Schelling, T.C., (1969), "Models of Segregation, " *American Economic Review* 56, 488-493.

Schelling, T.C., (1971), "Dynamic Models of Segregation, " *Journal of Mathematical Sociology* 1, 143-186.

Schelling, T.C., (1972), "A Process of Residential Segregation: Neighbourhood Tipping, " in: Pascal, A.H. (ed.). *Racial Discrimination in Economic Life*, (Lexington Books, Lexington, MA).

Yellin, J., (1974), "Urban Population Distribution, Family Income and Social Prejudice, " *Journal of Urban Economics* 1, 21-47.

Yinger, J., (1976a), "A Note on the Length of the Black-White Border, " *Journal of Urban Economics* 3, 370-382.

Yinger, J., (1976b), "Racial Prejudice and Racial Residential Segregation in an Urban Model, " *Journal of Urban Economics* 3, 383-396.

CHAPTER VII

OPTIMAL GROWTH OF CITIES

There have been very few works on the mathematical theory of urban growth. Recently, however, this area has begun to attract the attention of theorists. Miyao (1977a,b) analyzed capital accumulation in urban transportation. Rabenau (1976) considered the optimal growth of a small and open factory town with durable housing stock. Fujita (1976a,b) studied accumulation of more than one kind of durable housing capital. These works, however, are concerned only with growth of a certain city, despite the fact that in a modern economy the migration of households and firms is not difficult. Limiting the analysis to a single city prevents us from examining the interaction among cities. Isard and Kanemoto (1976) made an attempt to consider the optimal growth of an economy consisting of many cities and their hinterlands, though the model there is too complicated to go beyond the derivation and interpretation of first order conditions. This motivates the drastic simplifying assumptions of the model in this chapter.

For the first time, productive capital appears in our economy. Like capital in simple neoclassical growth models, it has a number of convenient features: it does not depreciate; it can be applied to any task; and if it is not needed for production, it can be eaten. In addition, because we are considering an economy with many cities, we also require capital that can be moved between and within cities costlessly.

The time dimension must be added to analyze capital accumulation. Since we already have the spatial dimension, the model becomes quite complicated. To keep the model manageable, we make the following drastic simplifying assumptions.

a. The economy consists of cities only: there is no rural sector (except possibly for the constant rural rent).

b. Capital accumulation occurs only in the urban production sector and there is no capital accumulation in the transportation sector.

c. Capital is perfectly mobile: capital can be moved between an within cities instantaneously and without cost.

d. Households are perfectly mobile.

e. All cities are identical. This assumption can be made only when capital is perfectly mobile: otherwise, a new city has zero capital stock initially, and cannot be identical with older cities. Under the assumption of mobility, capital stock in other cities can be instantaneously moved to the new city, and all cities can be made identical.

We assume that there is a Marshallian externality of the sort discussed in Chapter II. At the optimal city size declining production costs, which result from increasing city size, exactly match increasing transportation costs.

The utility level that households achieve in our economy has been determined by the amounts of land and of the consumer good they received. Now that the economy contains capital, some part of output can be invested in physical capital.

The problem of determining the optimal path of our economy may be solved in two stages. At each instant of time, all cities must maintain the optimal spatial allocation, as in Chapter I. The key difference is that when there is capital, the optimization is performed using the part of the product allocated to current consumption, rather than the entire product. The maximum utility level achievable in each city is then obtained as a function, $U(c, P)$, of current consumption, c, and the population of the city, P. In section 1, the model from Chapter I is reformulated with capital, and in section 2 the static spatial optimum is derived given the level of consumption.

At this point the inhabitants of each city know how to allocate their consumption, but not how much of their total product to consume. We assume that they choose to maximize the undiscounted sum of utilities over an infinite time horizon. That is, they are exactly as concerned about the welfare of their most remote descendants as they are about their own. We chose this assumption mainly for the sake of simplicity, but also because we see no moral justification for discounting the welfare of future generations. At any rate, it is quite easy to extend our analysis to the discounted case.

In maximizing the undiscounted sum over an infinite time horizon, we encounter a well-known difficulty: the undiscounted sum of future utilities is infinite, and we are left attempting to compare infinities. Economists have, of course, found several ways of avoiding this problem. In this chapter, we adopt a version of the *Ramsey device* used by Koopmans (1965). This approach changes the origin of the instantaneous

utility function, taking the utility level of the *optimal steady state* as zero, where the optimal steady state, or the Golden Rule, is the balanced growth path which maximizes the utility level among all feasible balanced growth paths. If $u(x)$ denotes the utility level at time t, and u^* that at the optimal steady state, the sum of the difference over infinite time horizon,

$$\int_0^\infty [u(t) - u^*]dt,$$

is maximized. The new objective function turns out to be bounded from above, and the difficulty of comparing infinities disappears. In section 3, the objective function is maximized with respect to the paths of consumption and population of a city.

In optimal growth theory, it is usually assumed that the utility function is concave. In our model, however, the maximum utility function, $U(c, P)$, may not be concave, although the original utility function over the consumer good and land is assumed to be concave. As it turns out, the maximum utility function is not even quasi-concave in most cases. This does not create serious difficulties for our analysis, if we assume that the concavity of the per capita production function is strong enough.

In section 4, a phase diagram analysis is carried out to determine whether a city grows during the process of capital accumulation. Section 5 contains remarks on the limitations of the model, and speculations on how the results might be modified if the model is extended.

1. The Model

Consider the growth of an economy consisting of cities. Let capital accumulation occur in the urban production sector and the number of cities change in the process of growth. Assume there is no non-urban sector and the total population of the whole economy is partitioned into cities. This assumption is clearly unrealistic and precludes the analysis of the evolution of an economy through different stages, for example, from the rural stage to the urban stage, as analyzed by Isard and Kanemoto (1976). Considering the complexity of the problem and the dominance of the urban sector in a modern economy, however it seems worthwhile to start with this simple formulation.

As discussed in Chapter II, economic factors which cause cities can be classified into three categories: concentration of immobile factors, increasing returns to scale, and

externalities. In this chapter we consider cities based on a Marshallian externality of the kind analyzed in section 5 of Chapter II. Instead of starting from the production function of an individual firm, we simply assume that the aggregate production function of a city can be written as

$$F(P,K,P), \tag{1.1}$$

where P and K are respectively the population and the aggregate capital stock of the city. The production function is homogeneous of degree one with respect to the first two terms and the derivative with respect to the third term is positive. The production function, therefore, exhibits increasing returns to scale if the third term is taken into account.[1]

Because it is easier to work with the capital-labour ratio and consumption per capita, $k = K/P$ and c, than with the absolute quantities, we want a per capita production function $f(k,P)$. By the homogeneity assumption, the per capita production function is

$$f(k,P) = F(1,k,P) = F(1,K/P,P), \tag{1.2}$$

where

$$f_P > 0. \tag{1.3}$$

We assume that the per capita production function is strictly concave and

$$f_k > 0. \tag{1.4}$$

As in previous chapters, we assume that all cities are identical. If at time t the population of the whole economy is $\overline{P}(t)$, the capital stock for the whole economy is $\overline{K}(t) = \overline{P}(t)k(t)$, and per capita consumption of the produced good is $c(t)$, then the output available for capital accumulation after consumption is

$$\dot{\overline{k}}(t) = \overline{P}(t)f(k(t),P(t)) - \overline{P}(t)c(t). \tag{1.5}$$

---

[1] It is not difficult to show that if an individual firm has a production function $\tilde{f}(\ell,k,P)$ where ℓ and k are respectively labour and capital inputs, the aggregate production function can be written as (1.1) when the number of firms is optimal.

If we assume that the population growth rate is a constant

$$\lambda = \dot{\bar{P}}(t) / \bar{P}(t), \tag{1.6}$$

then (1.5) can be rewritten as

$$\dot{k}(t) = f(k(t), P(t)) - \lambda k(t) - c(t). \tag{1.7}$$

The spatial structure of a city is the same as in the previous chapters. $\theta(x)dx$ of land is available in the ring between x and $x+dx$, where x is the distance from the center of a city. A household living at x at time t has a lot size $h(x,t)$. Then there are $(\theta(x)/h(x,t))dx$ households between x and $x+dx$ at time t. A household at x at time t consumes $z(x,t)$ of the produced good and spends $T(x)$ on commuting costs expressed in terms of the produced good. For simplicity, we assume that there is no capital accumulation or no technological progress in the transportation sector. Note that we have changed the notation for commuting costs and that t now denotes time. A city uses $c(t)P(t)$ of the produced good for consumption, which includes direct consumption, commuting costs, and the payment of the rural rent R_a. The *resource constraint* for a city is then

$$c(t)P(t) = \int_0^{\bar{x}(t)} \{[z(x,t)+T(x)]/h(x,t) + R_a\}\theta(x)dx, \tag{1.8}$$

where $\bar{x}(t)$ is the edge of the city at time t.

The *population constraints* are

$$P(t) = \int_0^{\bar{x}(t)} [\theta(x)/h(x,t)]dx, \tag{1.9}$$

and

$$\bar{P}(t) = n(t)P(t), \tag{1.10}$$

where $n(t)$ is the number of cities at time t. We shall ignore the constraint that $n(t)$ be an integer and take $n(t)$ as a continuous variable.

The utility function is $u(z,h)$, and we impose the constraint that the utility level be equal everywhere at each instant of time. The utility level may vary over time.

The *equal-utility constraint* can be written

$$u(t) = u[z(x,t), h(x,t)]. \tag{1.11}$$

Having set up the model, our problem is to maximize the undiscounted sum over an infinite time horizon:

$$\int_0^{\infty} [u(t) - u^*]dt, \tag{1.12}$$

subject to the constraints (1.7) through (1.11), and the initial condition,

$$k(0) = k_0, \tag{1.13}$$

where u^* is the utility level in the optimal steady state. The problem is solved in two stages.

2. Optimal Spatial Structure

In this section the first stage optimization is carried out for given $c(t)$ and $P(t)$, and the properties of the maximum utility function $U(c(t), P(t))$ are examined. This problem is exactly the same as the one in subsection 2.1 of Chapter I if we substitute $c(t)$, $P(t)$, and $T(x)$ for w, P, and $t(x)$. The utility level is maximized under the resource constraint, the population constraint, and the equal-utility constraint, which are in this case, (1.8), (1.9), and (1.11) respectively. Control variables are $z(x)$ and $h(x)$, and control parameters are $\bar{x}$ and u. The time variable t is suppressed in this section, since it plays no role in the optimization.

The first order conditions can be rewritten

$$u_h(z(x), h(x)) / u_z(z(x), h(x)) = R(x), \tag{2.1a}$$

$$y = z(x) + R(x)h(x) + T(x), \tag{2.1b}$$

$$R(\bar{x}) = R_a, \tag{2.1c}$$

after simple manipulations. As in Chapter I, the optimal solution can be achieved as a competitive equilibrium if all households receive the same income y. The solution, therefore, can be described by using the bid rent function $R(I(x), u) = R(y - T(x), u)$ defined in Equation (1.12) of Chapter I:

$$y = c + s, \tag{2.2a}$$

$$sP = \int_0^{\bar{x}} [R(y - T(x), u) - R_a]\theta(x)dx, \tag{2.2b}$$

$$P = \int_0^{\bar{x}} R_1(y - T(x), u)\theta(x)dx, \tag{2.2c}$$

$$R(y - t(\bar{x}), u) = R_a. \tag{2.2d}$$

s is the social dividend each household receives and is equal to the total differential rent divided by the population of the city. (2.2a) and (2.2b) correspond to (1.28) in Chapter I. (2.2c) is a restatement of the population constraint using the property of the bid rent function: $R_1 = 1/h$.

If c and P are given, the four equations, (2.2a)-(2.2d), determine the four variables, y, s, $\bar{x}$, and u. The utility level which is obtained can then be written as a function $U(c, P)$ Total differentiation of (2.2) yields the partial derivatives of the maximum utility function:

$$U_P(c, P) = s / \int_0^{\bar{x}} R_u\theta(x)dx < 0, \tag{2.3}$$

$$U_c(c, P) = -P / \int_0^{\bar{x}} R_u\theta(x)dx > 0, \tag{2.4}$$

where subscripts P and c denote partial derivatives with respect to P and c respectively. Thus an increase in the population of a city, given the consumption of resources per capita, lowers the utility level which can be attained in the city. An increase in per capita consumption given the population raises the utility level. The marginal rate of substitution between P and c is equal to the negative of the social dividend divided by the population:

$$S(c, P) \equiv U_P(c, P)/U_c(c, P) = -s/P. \tag{2.5}$$

Further properties of the maximum utility function are difficult to derive in the general case. The following results for four cases have been obtained by tedious calculations. The cases are

(i) the Leontief utility function $u(z, h) = [\min(z/\alpha, h)]^{1/\gamma}, \gamma > 1$, in a linear city, $\theta(x) = \theta$;

(ii) the Leontief utility function in a pie-slice city, $\theta(x) = \theta x$;

(iii) the Cobb-Douglas utility function $u(z,h)=(z^{\alpha}h^{1-\alpha})^{1/\gamma}$, $\gamma>1$, in a linear city; and

(iv) the Cobb-Douglas utility function in a pie-slice city.

In all cases, linear commuting costs are assumed: $T(x)=Tx$. The results are

(1) U_{cc} is negative in the linear city cases (i) and (iii).
In circular cities (ii) and (iv), U_{cc} is positive if γ is close to 1 and is negative for a large enough γ (in case (iv) we have proven this only in the case $R_a=0$).

(2) U_{PP} is positive in all cases (in the Cobb-Douglas cases we have proven this only in the case of $R_a=0$). This shows that $U(c,P)$ is not concave.

(3) $U(c,P)$ is not usually quasi-concave. In order for U to be quasi-concave, $\Delta=2U_{cP}U_cU_P-U_{cc}U_P{}^2-U_{PP}U_c{}^2$ must be nonnegative. In the case of the Leontief utility function, Δ equals zero in a linear city, and Δ is negative if $c>T\bar{x}$ in a circular city. In the Cobb-Douglas case, Δ is negative in a linear city and in the case of $R_a=0$ in a circular city.

(4) $S_c(c,P)$ is negative in all cases. This implies that -P would be a normal good if $U(c,P)$ were quasi-concave. (Note that P is a 'bad' and hence $-P$ is a good.)

(5) $S_P(c,P)$ is positive in all cases (in case (iv) we have proven this only in the case of $R_a=0$). This implies that c would be an inferior good if $U(c,P)$ were quasi-concave.

These results show that even if the original utility function $u(z,h)$ is concave, the maximum utility function $U(c,P)$ is not usually well behaved: $U(c,P)$ is usually neither concave nor quasi-concave. As it turns out, however, this does not cause a serious difficulty in the second stage optimization if the concavity of the production function (1.2) is strong enough.

3. Optimal Growth of Cities

In the second stage of our optimization procedure, the undiscounted sum over an infinite time horizon,

$$\int_0^\infty [U(c(t), P(t)) - u^*]dt, \tag{3.1}$$

is maximized subject to (1.7), (1.10), and (1.13). $U(c, P)$ is the maximized utility level from section 2 and u^* is the utility level at the optimal steady state. Since $n(t)$ appears only in the constraint (1.10) and is taken as a continuous variable, the problem is equivalent to the one of maximizing (3.1) under the constraints (1.7) and (1.13) with respect to $c(t)$ and $P(t)$. Although the population of a city $P(t)$ must be greater than one, we ignore this constraint, assuming that it is always satisfied along the optimal path.

Before solving this problem, we first examine the optimal steady state, at which the utility level is maximized among all feasible steady states. The optimal steady state is therefore the solution to the problem of maximizing

$$U(c, P)$$

subject to

$$f(k, P) - \lambda k - c = 0, \tag{3.2}$$

with respect to c, P, and k.

First order conditions for an interior optimum are

$$f_k(k, P) = \lambda, \tag{3.3a}$$

$$\frac{U_P(c, P)}{U_c(c, P)} + f_P(k, P) = 0\,. \tag{3.3b}$$

The first equation is the usual condition that the system operate at the biological rate of interest: the marginal productivity of capital must equal the population growth rate. The second equation requires that the population of a city be determined so that the *per capita* marginal external benefit on the production side equals the marginal rate of substitution between population and resource consumption per capita. From (2.5) and (2.2b), this is equivalent to

$$P[Pf_P(k, P)] = \int_0^{\bar{x}} [R(x) - R_a]\theta(x)dx, \tag{3.4}$$

which may be interpreted as the condition obtained in Chapter II that the total differential rent equals the total Pigouvian subsidy. An additional worker in a city produces $f(k, P)$ of the product himself, but at the same time increases the population

of the city and raises the production of other workers by Pf_P. The latter is the marginal external benefit, and the Pigouvian subsidy must equal Pf_P to achieve an efficient allocation. The left side of (3.4) is, therefore, the total amount of the Pigouvian subsidy in the city, which must equal the total differential rent when the number of cities is optimal.

The second order conditions are as follows.

$$f_{kk} \leq 0, \tag{3.5a}$$

$$S_P - S \cdot S_c + f_{PP} - (f_{kP})^2 / f_{kk} \leq 0, \tag{3.5b}$$

where $S(c,P)$ is the marginal rate of substitution between P and c and is defined in (2.5). The first two terms are

$$\begin{aligned} S_P - S \cdot S_c &= -\frac{1}{U_c^3}\left[2U_{cP}U_cU_P - U_{cc}U_P^2 - U_{PP}U_c^2\right] \\ &\equiv -\Delta / U_c^3. \end{aligned}$$

Since $U(c,P)$ is not usually quasi-concave as seen in section 2, $S_P - S \cdot S_c$ is usually positive. $f_{PP} - (f_{kP})^2 / f_{kk}$ is, however, negative if $f(k,P)$ is concave. (3.5b) can, therefore, be satisfied if the concavity of the production function is strong enough. (3.5a) is satisfied because we assumed that the production function is concave. We henceforth assume that (3.5a) and (3.5b) are satisfied with strict inequalities. We also assume that the optimal steady state is unique.

The following two observations can be immediately obtained from the first order conditions (3.3). First, unlike usual one sector growth models, the optimal steady state depends on the shape of the utility function. The population of a city serves as a link between the consumption side and the production side, and the capital-labour ratio at the optimal steady state is affected by the shape of the utility function. Second, at the optimal steady state, the configuration of a city remains exactly the same, and the number of cities increases at the same rate as the population growth.

Now, let us go back to the original problem of maximizing (3.1) with respect to $c(t)$ and $P(t)$ subject to (1.7) and (1.14). As shown in section 2 of the appendix on optimal control, the Hamiltonian for this problem is

$$\Phi = U(c(t), P(t)) + q(t)\left[f(k(t), P(t)) - \lambda k(t) - c(t)\right], \tag{3.6}$$

where $q(t)$ is an adjoint variable associated with the constraint (1.7). $q(t)$ satisfies the adjoint equation:

$$-\dot{q}(t) = q(t)\left[f_k(k(t), P(t)) - \lambda\right], \tag{3.7}$$

and the Hamiltonian must be maximized with respect to $c(t)$ and $P(t)$. The first order conditions for the maximization are

$$U_c(c(t), P(t)) = q(t), \tag{3.8a}$$

$$U_P(c(t), P(t)) + q(t) f_P(k(t), P(t)) = 0, \tag{3.8b}$$

and the second order conditions are

$$U_{cc} \le 0, \tag{3.9a}$$

$$U_{PP} + qf_{PP} \le 0, \tag{3.9b}$$

$$U_{cc}U_{PP} - (U_{cP})^2 + qU_{cc}f_{PP} \ge 0. \tag{3.9c}$$

As seen in section 2, (3.9a) is satisfied if the concavity of the original utility function $u(z,h)$ is strong enough. For (3.9b) to be satisfied, f_{PP} must be negative and its absolute value must be greater than U_{PP}/q, since U_{PP} is usually positive, and by (3.8a), q is also positive. In (3.9c) the sum of the first two terms is usually negative. Again, f_{PP} must be negative with a large absolute value.

Combining (3.7) and (3.8a) yields the differential equation:

$$U_{cc}\dot{c}(t) + U_{cP}\dot{P}(t) = U_c\left[\lambda - f_k\right], \tag{3.10}$$

and from (3.8a) and (3.8b) we obtain

$$S(c(t), P(t)) + f_P(k(t), P(t)) = 0. \tag{3.11}$$

Using (2.5), (3.11) becomes

$$s(t) = P(t) f_P(k(t), P(t)). \tag{3.12}$$

Thus the social dividend equals the Pigouvian subsidy at each point of time along the optimal path. In other words, the total amount of the Pigouvian subsidy for residents of the city must always equal the total differential rent.

Since there is no constraint on $k(t)$ at terminal time $t=\infty$, the transversality condition must be obtained to determine the value of $k(t)$ at $t=\infty$. If we can show that the optimal path converges to the optimal steady state, the transversality condition must be

$$\lim_{t\to\infty} q(t)k(t) = q^*k^*, \tag{3.13}$$

where $q^* = U_c(c^*, P^*)$, and asterisks denote the optimal steady state values of the variables.

We prove that the optimal path converges to the optimal steady state in two steps.[2] In the rest of this section, we show that the optimal path visits any arbitrarily small neighbourhood of the optimal steady state. This result still allows the possibility that the optimal path enters a neighbourhood of the optimal steady state but leaves there eventually. In the next section, we examine the behaviour of the optimal path near the steady state, and show that the steady state is a saddle point. Since this means that all paths except the one which converges to the saddle point diverge, the only path that visits an arbitrarily small neighbourhood of the optimal steady state is the convergent one. Thus the optimal path must converge to the optimal steady state, and (3.13) is in fact the required transversality condition.

To establish that the optimal path must visit an arbitrary neighbourhood of the steady state, we observe that the Kuhn-Tucker Theorem shows that when the constraint qualification is satisfied[3], there exists a multiplier q^* such that the optimal steady state maximizes the Lagrangian

$$U(c,P) + q^*\left[f(k,P) - \lambda k - c\right].$$

Thus the optimal steady state (k^*, c^*, P^*) satisfies

---

[2] This approach is similar to the one used by Scheinkman (1976) in the discounted case with many stocks.

[3] See, for example, Mangasarian (1969). See also section 3 in the appendix on optimal control theory for the explanation of constraint qualification.

$$\begin{aligned} &U(c^*,P^*)+q^*\left[f(k^*,P^*)-\lambda k^*-c^*\right] \\ &\quad \geq U(c,P)+q^*\left[f(k,P)-\lambda k-c\right], \quad \text{for any } k, c, \text{ and } P. \end{aligned} \tag{3.14}$$

It is not difficult to show that under some regularity conditions this inequality can be strengthened to the following: if $|k-k^*|>\varepsilon$ for any positive ε, then there exists $\rho>0$ such that

$$\begin{aligned} u^* &= U(c^*,P^*)+q^*\left[f(k^*,P^*)-\lambda k^*-c^*\right] \\ &> U(c,P)+q^*\left[f(k,P)-\lambda k-c\right]+\rho, \quad \text{for any } c \text{ and } P. \end{aligned} \tag{3.15}$$

If a path does not visit an arbitrarily small neighbourhood of the optimal steady state, there exists some $\varepsilon>0$ such that $|k(t)-k^*|>\varepsilon$ for any t. Inequality (3.15) then holds for any t and we can integrate it from 0 to ∞ to obtain

$$\int_0^\infty \left[U(c,P)-u^*\right]dt < -q^*\left[k(\infty)-k_0\right]-\int_0^\infty \rho\, dt. \tag{3.16}$$

Since $k(\infty)\geq 0$, the right side of the inequality is minus infinity. Thus the value of the criterion function of any path that does not visit an arbitrarily small neighbourhood of the optimal steady state is minus infinity.

Now it is easy to construct feasible paths which have values of the criterion greater than $-\infty$. For example, consider a path which approaches the optimal steady state with a constant $\dot{k}(\neq 0)$ and stops there. Such a path always exists if the initial capital-labour ratio, k_0, is larger than k^*, since we can determine $c(t)$ in such a way that $\dot{k}(t)$ is negative and constant until $k(t)$ reaches k^*. Even if the initial capital-labour ratio is smaller than k^*, such a path exists as long as $f(k-P)-\lambda k$ is positive for any k between k_0 and k^*.

Since $\dot{k}$ is constant and is not equal to zero, k^* will be reached within a finite length of time. The value of the criterion up to that time is then finite, and after that time the value can be made equal to zero by setting $c(t)=c^*$ and $P(t)=P^*$. Thus there exists a feasible path with a finite value of the criterion, and any path that does not visit an arbitrarily small neighbourhood of the optimal steady state cannot be optimal.

4. Phase Diagram Analysis

Now we examine the local behaviour of the optimal path near the optimal steady state. The optimal path satisfies differential equations (1.7) and (3.10), and equation (3.11) which must hold at each instant of time. The dynamic system contains three variables: k, c, and P. In order to work with a two-dimensional phase diagram, we use (3.11) to express c as a function, $c(k,P)$, of k and P, and obtain differential equations of k and P. Then implicit differentiation of (3.11) yields derivatives of $c(k,P)$:

$$c_k(k,P) = -f_{Pk}/S_c, \tag{4.1}$$

$$c_P(k,P) = -(S_{PP} + f_{PP})/S_c. \tag{4.2}$$

Observing

$$\dot{c}(t) = c_k\dot{k} + c_P\dot{P},$$

we can rewrite (3.10) as follows using (4.1), (4.2) and (1.7):

$$\dot{P}(t) = \frac{1}{D(k,P)}\{\phi(k,P)[\lambda - f_k(k,P)] + \varphi(k,P)[f(k,P) - \lambda k - c(k,P)]\}, \tag{4.3}$$

where

$$D(k,P) = U_{cc}(S_P + f_{PP}) - U_{cP}S_c, \tag{4.4a}$$

$$\phi(k,P) = -U_cS_c, \tag{4.4b}$$

$$\varphi(k,P) = -U_{cc}f_{Pk}. \tag{4.4c}$$

The differential equation (1.7) can also be rewritten using $c(t) = c(k(t), P(t))$,

$$\dot{k}(t) = f(k,P) - \lambda k - c(k,P). \tag{4.5}$$

(4.3) and (4.5) describe the paths that $k(t)$ and $P(t)$ must follow. The optimal steady state is the rest point of (4.3) and (4.5) since (3.2), (3.3a), and (3.3b) hold at the rest point.

To construct the phase diagram, we must know the signs of D, ϕ and φ. By simple manipulations, D becomes as follows.

$$D(k,P) = U_{cc}f_{PP} + \left[U_{cc}U_{PP} - (U_{cP})^2\right]/U_c$$
$$\geq 0,$$

(4.6)

where the inequality is obtained from (3.8a) and (3.9c). In order to determine the signs of ϕ and φ, we assume

$$f_{Pk} \geq 0, \tag{4.7}$$

$$S_c \leq 0. \tag{4.8}$$

The first inequality implies that capital and population are complementary in production. As mentioned in section 2, the second inequality is satisfied in all the examples we have calculated. Since $U_{cc} \leq 0$ from the second order condition (3.9a), ϕ and φ are both nonnegative under these assumptions:

$$\phi(k,P) \geq 0, \tag{4.9}$$

$$\varphi(k,P) \geq 0. \tag{4.10}$$

These assumptions also imply that

$$c_k(k,P) > 0. \tag{4.11}$$

We now construct the phase diagram of (4.3) and (4.5). Following the usual procedure, we first examine the loci of $\dot{P} = 0$ and $\dot{k} = 0$. The locus of $\dot{k} = 0$ is

$$f(k,P) - \lambda k - c(k,P) = 0, \tag{4.12}$$

and the slope of the locus is

$$\frac{dk}{dP}\Big|_{\dot{k}=0} = \frac{f_P - c_P}{c_k + \lambda - f_k}. \tag{4.13}$$

Since by (3.3a) we have $f_k = \lambda$ at the optimal steady state, the slope there is

$$\frac{dk}{dP}\Big|_{\dot{k}=0} = \frac{f_P - c_P}{c_k}. \tag{4.14}$$

The denominator is positive by (4.11). By (4.2) the numerator is

$$f_P - c_P = \frac{1}{S_c}[f_{PP} + S_P - S \cdot S_c],$$

which is also positive from (4.6), (3.5b) and the strict concavity of $f(k,P)$. Thus the $\dot{k} = 0$ locus is upward sloping at the Golden Rule:

$$\frac{dk}{dP}\Big|_{\dot{k}=0} > 0 \qquad \text{at} \qquad f_k = \lambda. \tag{4.15}$$

Since we have

$$\frac{\partial}{\partial P}\left[f(k,P)-\lambda k-c(k,P)\right]=f_P-c_P>0, \tag{4.16}$$

$\dot{k}=f-\lambda k-c$ is negative above the $\dot{k}=0$ locus and positive below the locus as illustrated in Figure 1.

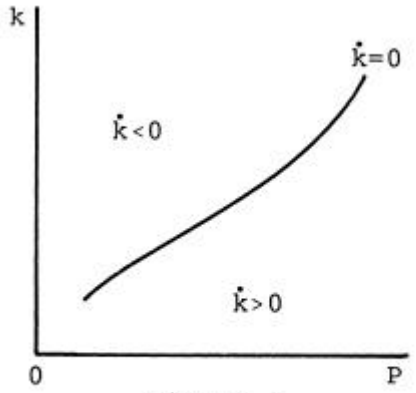

Figure 1
The Locus of $\dot{k}=0$

Next, consider the locus of $\dot{P}=0$. From (4.3) it is a combination of the $\dot{k}=0$ locus and the locus of

$$\lambda-f_k(k,P)=0. \tag{4.17}$$

The slope of the locus of (4.17) is

$$\left.\frac{dk}{dP}\right|_{f_k=\lambda}=-f_{kP}/f_{kk}>0, \tag{4.18}$$

where the inequality is the result of concavity and complementarity. The locus of (4.17) is, therefore, upward sloping. Since

$$\frac{\partial}{\partial k}\left[\lambda-f_k(k,P)\right]=-f_{kk}$$
$$>0,$$

$\lambda-f_k$ is positive above the locus of $\lambda=f_k$ and negative below the locus. This is illustrated in Figure 2.

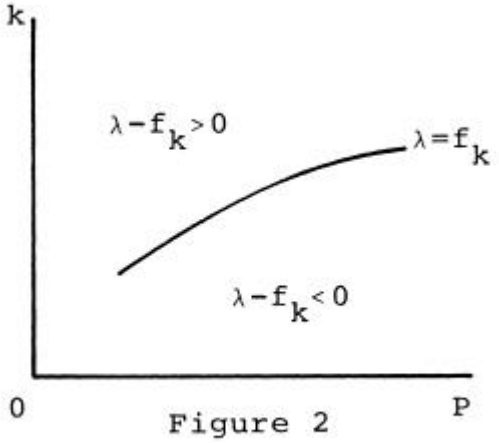

Figure 2
The Locus of $\lambda=f_k$

The locus of $\lambda = f_k$ intersects with the $\dot{k}=0$ locus at the Golden Rule. The $\dot{k}=0$ locus is steeper than the locus of $\lambda = f_k$ at the intersection point since the following inequality holds there:

$$\begin{aligned}
&\frac{dk}{dP}\Big|_{\dot{k}=0} - \frac{dk}{dP}\Big|_{\lambda=f_k} \\
&= \frac{f_P - c_P}{c_k} + \frac{f_{kP}}{f_{kk}} \\
&= -\frac{S_P - SS_c + f_{PP} - (f_{kP})^2 / f_{kk}}{f_{Pk}} \\
&\geq 0,
\end{aligned} \tag{4.20}$$

where we used (3.5b) and (4.7). Figure 3 illustrates the relationship between the two loci. Since D, ϕ and φ are all nonnegative, the $\dot{P}=0$ locus passes through regions (A) and (C) in Figure 3, and $\dot{P}$ is positive on the side of region (D). There are three possibilities:

(i) the $\dot{P}=0$ locus is downward sloping,

(ii) the $\dot{P}=0$ locus is upward sloping but flatter than the $\lambda = f_k$ locus, and

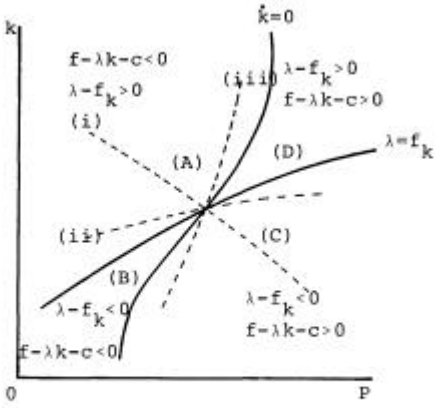

Figure 3. The Loci of $\dot{k}=0$, $\lambda=f_k$, and $\dot{P}=0$

(iii) the $\dot{P}=0$ locus is upward sloping and steeper than the $\dot{k}=0$ locus.

The slope of the $\dot{P}=0$ locus is, at the Golden Rule,

$$\frac{dk}{dP}\Big|_{\dot{P}=0}=\frac{Df_{kP}}{U_{cc}f_{kk}[(-1/f_{kk})(f_{pk})^2-(-U_c/U_{cc})(S_c)^2]}$$
$$\begin{matrix}>\\=\\<\end{matrix}0 \qquad \text{as} \qquad (-1/f_{kk})(f_{Pk})^2 \begin{matrix}>\\<\end{matrix} (-U_c/U_{cc})(S_c)^2 \tag{4.21}$$

and

$$\frac{dk}{dP}\Big|_{\dot{P}=0}-\frac{dk}{dP}\Big|_{\dot{k}=0}$$
$$=\frac{U_c(S_c)^2[S_P-SS_c+f_{PP}-(f_{kP})^2/f_{kk}]}{U_{cc}f_{Pk}[(-1/f_{kk})(f_{kP})^2-(-U_c/U_{cc})(S_c)^2]}$$
$$\begin{matrix}>\\=\\<\end{matrix}0 \qquad \text{as} \qquad (-1/f_{kk})(f_{Pk})^2 \begin{matrix}>\\<\end{matrix} (-U_c/U_{cc})(S_c)^2. \tag{4.22}$$

These relationships imply that

$$\frac{dk}{dP}\Big|_{\dot{P}=0} > \frac{dk}{dP}\Big|_{\dot{k}=0} \qquad \text{if} \qquad (-1/f_{kk})(f_{Pk})^2 > (-U_c/U_{cc})(S_c)^2 \tag{4.23}$$

$$\frac{dk}{dP}\Big|_{\dot{P}=0} < 0 \qquad \text{if} \qquad (-1/f_{kk})(f_{Pk})^2 < (-U_c/U_{cc})(S_c)^2 \tag{4.24}$$

Thus case (i) is obtained if $(-1/f_{kk})(f_{Pk})^2 < (-U_c/U_{cc})(S_c)^2$, and case (iii) otherwise, but case (ii) never occurs.

In case (i), we obtain a phase diagram depicted in Figure 4. The optimal steady state is a saddle point and all paths except for the two stable branches diverge. Since it was shown in the preceding section that the optimal path must visit any arbitrarily small neighbourhood of the optimal steady state, the optimal path must be one of the stable branches. The diagram also shows that at least in the neighbourhood of the steady state the optimal path is either in the region where $\dot{k} > 0$ and $\dot{P} < 0$ or in the region where $\dot{k} < 0$ and $\dot{P} < 0$. *The population of a city therefore declines as capital accumulates*. Notice,

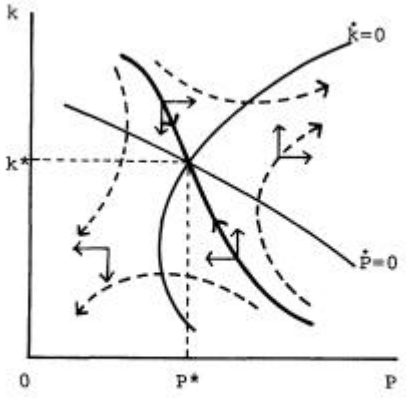

Figure 4. A Phase Diagram

however, that this conclusion may not hold globally as Figure 5 illustrates.

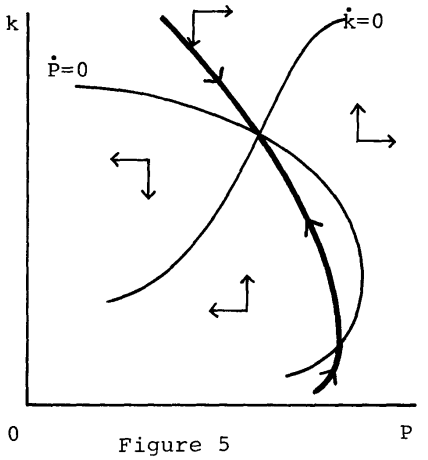

Figure 5
Global Behaviour Which Is Different from Local Behaviour

In case (iii), we obtain a phase diagram like Figure 6. The optimal steady state is a saddle point in this case as well and the optimal solution is either of the stable branches. It can be seen from the diagram that the population of a city rises as capital accumulates in the neighbourhood of the Golden Rule.

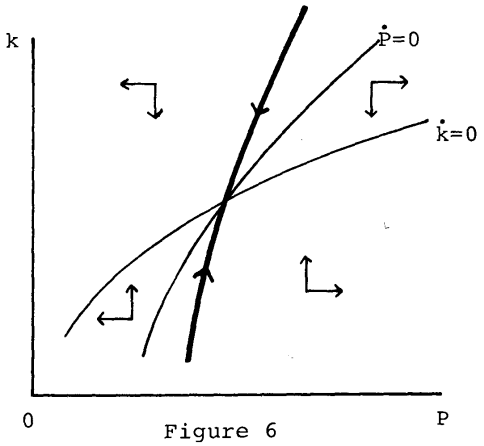

Figure 6
An Upward Sloping $\dot{P}=0$ Locus

These results are summarized in the following theorem:

Theorem 1: Suppose $S_c \leq 0$ *and* $f_{Pk} \geq 0$. *If* $(-1/f_{kk})(f_{Pk})^2 < (-U_c/U_{cc})(S_c)^2$, *then the population of a city falls as capital accumulates in the neighbourhood of the optimal*

steady state ; and if $(-1/f_{kk})(f_{Pk})^2 > (-U_c/U_{cc})(S_c)^2$, *then the population rises.*

The assumption of complementarity, $f_{Pk} \geq 0$, is rather arbitrary although the assumption is satisfied in most widely used production functions such as the Cobb-Douglas and CES functions. If capital and population are anticomplementary, the following result is obtained.

Theorem 2: Suppose $S_c \leq 0$ *and* $f_{Pk} \leq 0$. *Then the population of a city falls as capital accumulates in the neighbourhood of the optimal steady state.*

Labour augmenting technical progress, or Harrod neutral technical progress, can be incorporated in this analysis quite easily although other types of technical progress are not easy to handle. When the rate of labour augmenting technical progress is σ, the same result as in the case without the technical progress is obtained if λ is replaced by $\lambda+\sigma$ and P by the population in terms of efficiency labour, $Q = Pe^{\sigma t}$. Since *P* and *Q* have the relationship:

$$\frac{\dot{P}}{P} = \frac{\dot{Q}}{Q} - \sigma, \tag{4.25}$$

the rate of increase of the population of a city is smaller by the technical progress rate than the case without the technical progress. Thus labour augmenting technical progress introduces a tendency for city size to decline over time.

The reason why the sign and the magnitude of f_{Pk}, are crucial in Theorems 1 and 2 must be obvious. The population size is determined in such a way that $S + f_P = 0$, i.e., the marginal cost of having a bigger population on the consumption side balances the marginal externality benefit on the production side. If f_{Pk}, is positive, an increase in the capital-labour ratio increases the marginal benefit and tends to increase the population of a city. This tendency would be offset if the marginal cost on the consumption side rises. As capital-labour ratio rises, per capita consumption

usually increases. If S_c is nonpositive as assumed in the Theorems, the marginal cost 2 also rises. Theorem 1 states that when $(-U_c/U_{cc})(S_c)^2$ is greater than $(-1/f_{kk})(f_{Pk})^2$, this effect overwhelms the effect of the rise in the marginal benefit. If f_{Pk}, is negative, both effects work in the same direction and the population of a city always declines in the process of capital accumulation as in Theorem 2.

5. Concluding Remarks

We have characterized the condition required for capital accumulation to be accompanied by an increase in the population of a city. It was shown that the population growth tends to occur if capital and the external economy are complementary in production and that the population tends to decline if the marginal rate of substitution between the population and consumption becomes greater in absolute value as the consumption increases. It is believed that ordinary factors of production

are usually complementary, although it is not clear whether this is true if there are externalities. In examples that we have calculated, the marginal rate of substitution between the population of a city and consumption rises in absolute value as the consumption increases. Empirical studies are therefore necessary to determine whether capital accumulation favours bigger cities.

It is quite obvious that our model is too simple to capture the complexity of modern cities. It does not deal with the following important aspects of real cities.

First, we do not have a hierarchy of cities. Rather, our cities are identical. More than one kind of good has to be introduced to obtain a hierarchy of cities.

Second, the production function is assumed to remain the same over time (except for the possibility of labour augmenting technical change). It might have been shifting to increase the benefits of bigger cities.

Third, perfect mobility and malleability of capital is not a realistic assumption, and there are costs involved in building a new city, which tends to reduce the number of cities and hence to increase the size of a city.

Fourth, there is a good reason to believe that a market economy has a very different growth path from the optimal one. As shown in Chapter II, the market equilibrium is not unique and a city size greater than the optimum may well be an equilibrium.

Fifth, technical progress and capital accumulation in transportation sector has worked to reduce the cost of bigger cities.

REFERENCES

Fujita, M., (1976a), "Spatial Patterns of Optimal Growth: Optimum and Market, " *Journal of Urban Economics* 3, 193 -208.

Fujita, M., (1976b), "Toward a Dynamic Theory of Urban Land Use, " *Papers of Regional Science Association* 37, 133-165.

Isard, W. and Y. Kanemoto, (1976), "Stages in Space-Time Development, " *Papers of Regional Science Association* 37, 99-131.

Koopmans, T.C., (1965), "On the Concept of Optimal Economic Growth, " in: *Study Week on Econometric Approach to Planning, Pontificiae Scientiarum Scripta Varia* XXVIII, (Rand McNally, Chicago).

Mangasarian, O.L., (1969), *Nonlinear Programming*, (McGraw-Hill, New York).

Miyao, T., (1977a), "The Golden Rule of Urban Transportation Investment," *Journal of Urban Economics* 4, 448-458.

Miyao, T., (1977b), "A Long-Run Analysis of Urban Growth Over Space, " *Canadian Journal of Economics* 10, 678-686.

Rabenau, B., (1976), "Optimal Growth of a Factory Town, " *Journal of Urban Economics* 3, 97-112.

Ramsey, F.P., (1928), "A Mathematical Model of Saving, " *Economic Journal* 38, 543-559.

Scheinkman, J.A., (1976), "On Optimal Steady States of n-Sector Growth Models when Utility is Discounted, " *Journal of Economic Theory* 12, 11-30.

APPENDIX I

EQUALITY AND THE BENTHAMITE SOCIAL WELFARE FUNCTION

In this appendix we explain the reason why utility levels differ between different locations at the Benthamite optimum. The basic reason is that the utility possibility frontier is skewed in favour of households living farther from the center. This can be illustrated by considering a rectangular city consisting only of two households. For notational simplicity, the width of the city is assumed to be 1, i.e., $\theta(x)=1$. Household i consumes z_i of the consumer good and h_i of space. Both households are assumed to commute to the center of the city from the center of their properties, i.e., commuting costs of household 1 and 2 are respectively $t(\frac{1}{2}h_1)$ and $t(h_1+\frac{1}{2}h_2)$, where household 1 lives closer to the center. This city is illustrated in Figure 1.

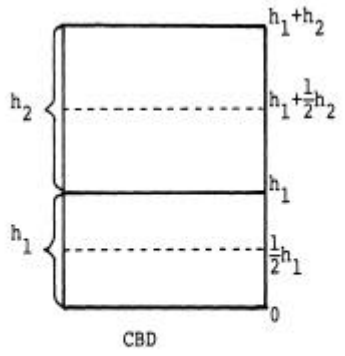

Figure 1. A Rectangular Two-Household City

The resource constraint for the city is given by

$$z_1+z_2+t(\frac{1}{2}h_1)+t(h_1+\frac{1}{2}h_2)+R_a(h_1+h_2)=Y. \quad (1)$$

Given the resource constraint, we can obtain the set of feasible utility levels of the two households. The frontier of the set is called the *utility possibility frontier* and

depicted by the curve LL' in Figure 2. The utility possibility frontier expresses the maximum utility that household 1 can achieve at every possible utility level for household 2. It is obtained by maximizing u_1, subject to the resource constraint and to

$$u(z_2, h_2) \geq u_2. \tag{2}$$

The Lagrangian is therefore

$$\begin{aligned}\Lambda = u(z_1, h_1) + \delta\left[Y - z_1 - z_2 - t(\frac{1}{2}h_1) - t(h_1 + \frac{1}{2}h_2) - R_a(h_1 + h_2)\right] \\ + \lambda\left[u(z_2, h_2) - u_2\right].\end{aligned} \tag{3}$$

The first order conditions can be summarized as

$$\lambda = \frac{u_z(z_1, h_1)}{u_z(z_2, h_2)} = \frac{u_h(z_1, h_1)}{u_h(z_2, h_2)} - \frac{1}{2}\delta\,\frac{t'(\frac{1}{2}h_1) + t'(h_1 + \frac{1}{2}h_2)}{u_h(z_2, h_2)}. \tag{4}$$

Now it can be shown that the utility possibility frontier; is skewed as in Figure 2, so that its slope is flatter than minus 1 when the two households obtain the same utility level.

By the Envelope Theorem in Appendix III, the slope of the utility possibility frontier is

$$\frac{du_1}{du_2} = \frac{\partial \Lambda}{\partial u_2} = -\lambda = -\frac{u_z(z_1, h_1)}{u_z(z_2, h_2)}. \tag{5}$$

Since the last term on the RHS of (4) is negative, we have

$$\frac{u_h(z_1, h_1)}{u_z(z_1, h_1)} > \frac{u_h(z_2, h_2)}{u_z(z_2, h_2)}. \tag{6}$$

Thus the slope of an indifference curve is steeper at (z_1, h_1) than at (z_2, h_2). Due to the convexity of indifference curves, this implies, as shown in Figure 3, that $z_1 > z_2$

and $h_1 > h_2$ if $u_1 = u_2$. But if land is a normal good, the following inequality is obtained from (I.2.7) and (I.2.8):

$$\begin{aligned}\left.\frac{du_z(z,h)}{dz}\right|_{u=const.} &= u_{zz} + u_{zh}\left.\frac{dh}{dz}\right|_{u=const.} \\ &= u_{zz} - u_{zh}\frac{u_z}{u_h} \\ &= -D(u_h/u_z)\hat{h}_I < 0.\end{aligned}$$

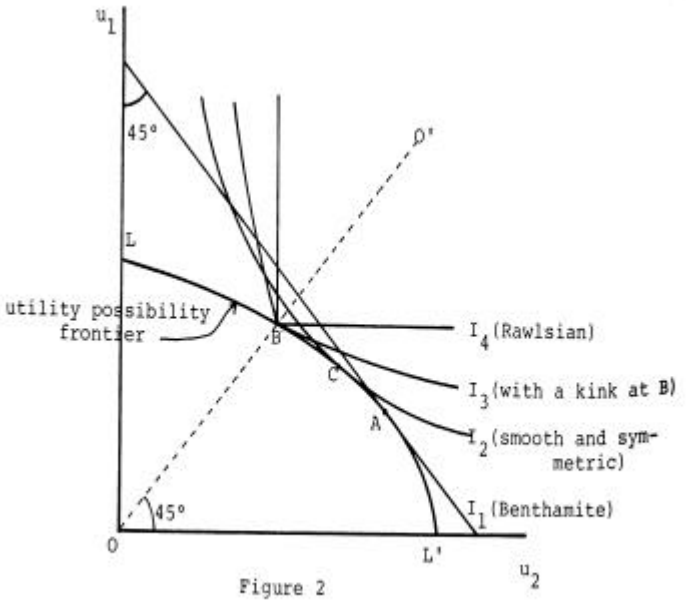

Figure 2

Hence, u_z decreases as z increases along an indifference curve and we finally obtain

$$\left.\frac{du_1}{du_2}\right|_{u_1=u_2} = -\frac{u_z(z_1,h_1)}{u_z(z_2,h_2)} > -1.$$

As the simple sum of utilities is maximized in the Benthamite case, the Benthamite optimum is point A in Figure 2 at which the 45° line I_1 is tangent to the utility possibility frontier. Since the utility possibility frontier is flatter than 45° when

utility levels are equal, the optimum must lie below the equal-utility line 00'.[1] Thus household 2 receives a higher utility level than household 1.

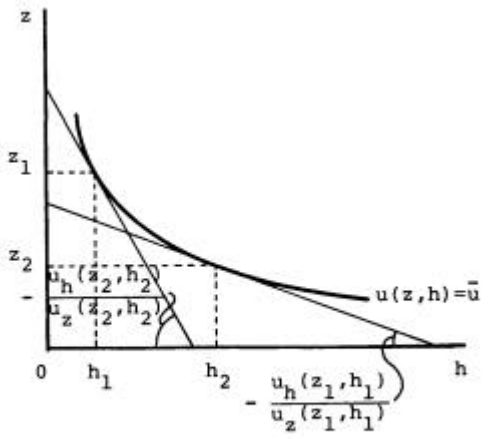

Figure 3

This result generalizes to any *smooth* symmetric quasiconcave social welfare function, $W(u_1,u_2)$, represented by indifference curves like I_2 since all indifference curves of a smooth symmetric social welfare function must have slope -1 along the equal-utility line 00'. A symmetric quasi-concave social welfare function yields equal utility levels only if indifference curves have kinks along the 45° line, as I_3 does. One example is the Rawlsian case represented by I_4.

The skewed utility possibility frontier is a result of the so-called concealed nonconvexity. In our model, it is assumed that a household must choose only one location and cannot live at more than one location at a time. This assumption can be interpreted in two ways. First, it may be considered as a restriction on the consumption set. For example, in Figure 4, which describes housing consumptions at two locations x and x', the consumption set is limited to the two axes, and any point within the first quadrangle cannot be chosen. In this case, the consumption set is not obviously convex. Second, the assumption may be a consequence of nonconvex preferences. If indifference curves are concave to the origin as in the Figure 4, a household, given a linear budget constraint, always chooses one of the corners.

---

[1] It is implicitly assumed that the utility possibility set is convex. This is true if the transportation cost function is convex and the utility function is concave.

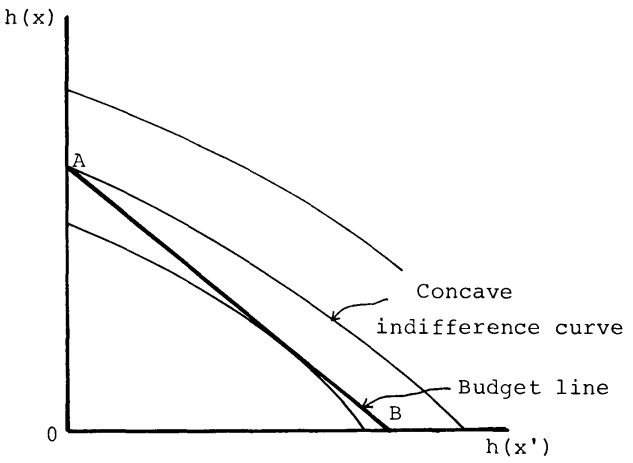

Figure 4. Concealed Nonconvexity

In our model, this nonconvexity is not harmful for the existence and efficiency of competitive equilibrium, since enough smoothness is obtained by introducing a population density function. The crucial assumption is that the population density at each distance can be any real number. This is not true in a model with several regions in which the population in each region must be an integer. In such a case demand for land in one region is discontinuous at the price level where an individual moves in or out of the region. Hence, there may never be a price vector that equilibriate the market for land. If, however, the population in a region can be any real number, such discontinuity will not occur and the existence of competitive equilibrium will be guaranteed. Schweizer, Varaiya and Hartwick (1976) proved that competitive equilibrium exists in a model with the concealed nonconvexity if a population density can be any real number.

This result is analogous to the well-known result in general equilibrium theory (due to Star (1969) and others) that nonexistence of equilibrium caused by nonconvexities of individual units disappears as the economy becomes larger relative to individual economic units. In particular, it is parallel to the work of Aumann (1966) which shows that in a model with a continuum of households, each of infinitesimal endowment, the existence of competitive equilibrium can be proven without making any assumption about convexity.

Although the nonconvexity does not introduce any difficulty concerning the existence and efficiency of competitive equilibrium, it causes inequality in utility levels. The asymmetry in the utility possibility frontier arises since housed holds living near the center, for example, are not allowed an access to land in the suburbs. In such a

case, households living at different locations face different opportunity sets. If, however, households can live at more than one location, they all face exactly the same budget constraint and there is no difference between households. The utility possibility frontier is then symmetric and all households receive the same utility level at the Benthamite optimum.

Some economists prefer the Benthamite case on the grounds that the Rawlsian social welfare function must be assumed to obtain the equal-utility optimum. As can be seen from Fig. 2 however, this claim is not true. Utility levels are equal at the optimum if the social welfare indifference curves have sufficiently strong kinks along the equal-utility line.

Any symmetric indifference curve with no kinks has a slope -1 along the equal utility line. This implies that in the neighborhood of the equal utility line the social welfare function behaves in the same way as the Benthamite social welfare function. Thus at least locally the aggregate utility is maximized and the social welfare function exhibits no preference for equality of utility levels. If local preference for equality is assumed at the point where utility levels are equal, indifference curves will have kinks and utility levels may be equal at the optimum.

REFERENCES

Aumann, R.J., (1966), "Existence of Competitive Equilibria in Markets with a Continuum of Traders," *Econometrica* 34, 1-17.

Schweizer, U., P. Varaiya, and J. Hartwick, (1976), "General Equilibrium and Location Theory," *Journal of Urban Economics* 3, 285-303.

Starr, R., (1969), "Quasi-equilibria in Markets with Nonconvex Preferences," *Econometrica* 37, 25-38.

APPENDIX II

LOCAL PUBLIC GOODS IN A MORE GENERAL MODEL

In this appendix the analysis in Chapter III is extended to the case of two factors of production. It is assumed that there is more than one kind of consumer goods and that land as well as labour is used in producing the consumer goods. The production function of the *i*-th consumer good is written as

$$F^i(L_i, H_i) \qquad i = 1, \ldots, k \tag{1}$$

where L_i and H_i are respectively labour and land inputs. Assuming that the production function is homogeneous of degree one, we obtain the per-unit-land production function, $f^i(l_i)$:

$$F^i(L_i, H_i) = H_i F^i(\frac{L_i}{H_i}, 1) = H_i f^i(l_i) \tag{2}$$

where l_i is the labour-land ratio, L_i / H_i.

The utility function of city residents is

$$u(z(x), h(x), X), \tag{3}$$

where $z(x) = (z_1(x), \ldots, z_i(x), \ldots, z_k(x))$, is the vector of consumer goods.

In contrast to our procedure in Chapter III, we assume a vector of transportation costs $t(x) = (t_1(x), \ldots, t_i(x), \ldots, t_k(x))$, for consumer goods within a city. Each city now has a port, or perhaps a railroad station at the center, where goods are bought at prices $p = (p_1, \ldots, p_i, \ldots, p_k)$ for distribution throughout the economy. Cities are small, so that prices are effectively parametric, and producers at x face the net price vector

$$p(x) = p - t(x). \tag{4}$$

If good *i* is produced at *x*, we obtain the following equations by profit maximization:

$$P_i(x) f^{i\prime}(l_i(x)) = w(x) \tag{5}$$

$$P_i(x)[f^i - l_i(x) f^{i\prime}(l_i(x))] = R(x) \tag{6}$$

where $w(x)$ and $R(x)$ are respectively wage rate and land rent at x.

We assume that there is also a retail market at the center of the city. In buying the consumer goods, residents in the city are assumed to incur transportation costs from the market to the place of residence. Therefore, households living at x face the price vector of the consumer good:

$$q(x) = p + t(x). \tag{7}$$

In each city one developer collects land rent and pays the rural rent and the costs of the public good. The profit is distributed equally among all households in the economy. If we assume that there are many identical cities, a household receives dividends from many developers and a change in one city does not significantly affect the total dividend, s, that a household receives. A household working at x' receives the wage, $w(x')$, and the dividend. If the household lives at x, the budget constraint is

$$w(x') + s = q(x) \cdot z(x) + \left[t_h(x) - t_h(x')\right] + R(x)h(x) \tag{8}$$

where $t_h(x)$ is the commuting costs from x to 0 and hence $t_h(x) - t_h(x')$ is the commuting costs from x to x'.

We assume that all households have the same skill and the same utility function. Then all households receive the same utility level in equilibrium. This implies that all households living at the same location must receive the same net income after commuting costs wherever they work. Therefore, we obtain

$$w(x') = w - t_h(x'), \tag{9}$$

where $w \equiv w(0)$.

A household's utility maximization yields

$$\frac{u_h}{u_{z_i}} = \frac{R(x)}{q_i(x)} \qquad I = 1, \ldots, k. \tag{10}$$

Using (9), we can rewrite (5) and (8) as

$$p_i(x) f^{i\,\prime}(l_i(x)) = w - t_h(x) \tag{11}$$

$$w + s = q(x) \cdot z(x) + t_h(x) + R(x)h(x) \tag{12}$$

Totally differentiating (11) and (6), we obtain

$$dw = p_i(x) f^{i\,\prime\prime}(l_i(x)) dl_i(x) \tag{13}$$

$$-p_i(x) l_i(x) f^{i\,\prime\prime}(l_i(x)) dl_i(x) = dR(x). \tag{14}$$

Combining these two equations, the following simple relationship can be obtained:

$$dR(x) = -l_i(x)dw \tag{15}$$

Totally differentiating (3) and (12), and noting the small city assumption that the utility level is given, we obtain

$$0 = u_z dz(x) + u_h dh(x) + u_x dX$$

$$dw = q(x) \cdot dz(x) + R(x)dh(x) + h(x)dR(x).$$

From these two equations we have

$$h(x)dR(x) = \frac{u_x}{u_z}dx + dw. \tag{16}$$

From (15) and (16) , the change of the total rent in the city due to an increase of the public good is equal to the social benefit of the public good:

$$\begin{aligned}
&\int_0^{\bar{x}} \frac{dR(x)}{dx}\theta(x)\,dx \\
&= \sum_{i=1}^{k} \int_{\underline{x}}^{\bar{x}} -\frac{dw}{dx} l_i(x)\theta(x)\,dx + \int_{\underline{x}}^{\bar{x}} \frac{dw}{dx} N(x)\,dx + \int_{\underline{x}}^{\bar{x}} \frac{u_x}{u_z} N(x)\,dx \\
&= \frac{dw}{dx}\left[\int_{\underline{x}}^{\bar{x}} N(x)\,dx - \sum_{i=1}^{k} \int L_i(x)\,dx \right] + \int_{\underline{x}}^{\bar{x}} \frac{u_x}{u_z} N(x)\,dx \\
&= \int_{\underline{x}}^{\bar{x}} \frac{u_x}{u_z} N(x)\,dx .
\end{aligned} \tag{17}$$

The last equality is obtained using the fact that the total labour force must be equal to the population of the city. Thus, even if there are more than one factor of production and more than one consumer good, the benefit of the public good is reflected in the increase of land rent in a small city.

We can also see that the profit maximization of a city developer leads to an efficient supply of the public good. A city developer maximizes

$$\int_0^{\bar{x}} [R(x) - R_a]\theta(x)dx - C(x),$$

where $C(X)$ is the cost of producing the public good. Then

$$\begin{aligned}
&\frac{d}{dx}\left\{ \int_0^{\bar{x}} [R(x) - R_a]\theta(x)dx - C(x) \right\} \\
&= \int_0^{\bar{x}} \frac{dR(x)}{dX}\theta(x)dx - C'(x) \\
&= \int_{\underline{x}}^{\bar{x}} \frac{u_x}{u_z} N(x)dx - C'(x) = 0.
\end{aligned}$$

As in section 3 of Chapter III it can be seen that the last equality is the condition for an efficient supply of the public good.

Notice that this result does not depend on the number of commodities produced in the city, or on whether different goods are produced in different zones. We used only the conditions for a small city: given utility level; given price vector of consumer goods; and constant returns to scale in production. Although in general the wage rate changes as the supply of the public good changes, it does not affect the conclusion, since the effects on the production side and the consumption side cancel out each other.

This result can be interpreted in the same way as in section 1 of Chapter III. The benefits of the public good must accrue to somebody or become a deadweight loss. But there is no deadweight loss if there are no distortions in the rest of the economy. Therefore, all the benefit must be received by somebody. By the assumption of a small city, the residents cannot benefit from the public good. Because of constant returns to scale there is no profit in equilibrium. Thus the land rent is the only place the benefit appears.

This argument suggests that if returns to scale are constant, the sum of land rent and the profits (or losses) of producers reflects the benefit of the public good. It is not difficult to show that this is indeed true.

APPENDIX III

THE ENVELOPE PROPERTY

Optimization imposes a very strong structure on the problem considered. This is the reason why neoclassical economics, which assumes optimizing behaviour, has been the most successful of social sciences. One of its important aspects is the envelope property discussed in this Appendix.

The envelope property is concerned with the rate of change of the maximum (or minimum) value of a criterion function caused by a change in some parameter; for example, a change in the maximum utility level of a household caused by a change in income, a change in the minimum cost of production caused by a change in the output level, and so on. A change in a parameter in general induces a change in the optimum levels of choice variables. According to the envelope property, however, the induced change in the choice variables may be ignored in calculating the effect of a change in a parameter on the maximum value if the change is very small. In other words, a change in the maximum value caused by a marginal change in a parameter, which also induces a change in the choice variables, is equal to a change in criterion function with choice variables fixed.

In section 1, the envelope property is explained in the simplest possible case. The Envelope Theorem is stated and proved in section 2. In section 3 properties of the indirect utility function and the expenditure function are derived as applications of the Envelope Theorem.

1. The Simplest Case

The essence of the envelope property may be explained using the following simple maximization problem. Consider the problem of maximizing the criterion function, $V(x,b)$, with respect to x for a given parameter b. An interior maximum is obtained at the point where the derivative of the criterion function with respect to x is zero,

$$\frac{\partial V(x,b)}{\partial x} = 0.$$

The maximizing value of x changes as the parameter b changes: from x^* to $x^{*\prime}$ as b moves to b'. *The envelope property states that the total effect of an infinitesimal change in the parameter on the maximized value of the criterion function (including the effect of an induced change in the optimum value of x) equals the partial effect on the criterion function with the level of x fixed.* In Figure 1 the former is the movement from V to V'; and the latter from V to $\tilde{V}$. Since the criterion function is

approximately flat near the optimum point, the difference between the two, $\tilde{V}V'$, is very small compared with $V\tilde{V}$. As the change in the parameter approaches zero, the difference becomes negligible and the envelope property can be invoked.

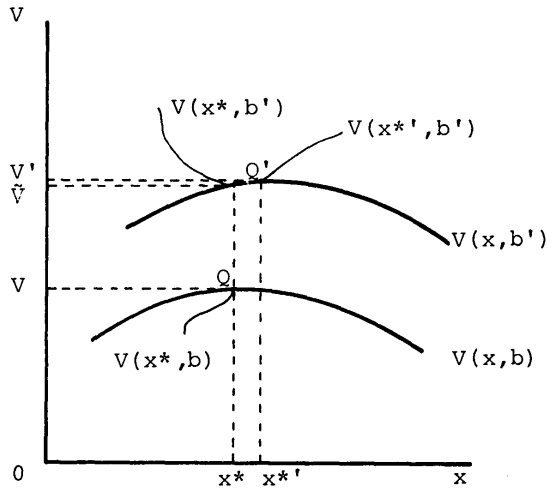

Figure 1.

The envelope property can be derived by mechanically differentiating the criterion function at the maximum. Since the optimum value of x depends on b, it can be described as a function, $x^*(b)$, of b. Then the total effect including a change in x^* is

$$\frac{dV(x^*(b),b)}{db} = \frac{\partial V}{\partial x}\frac{dx^*}{db} + \frac{\partial V}{\partial b}$$

and the partial effect is

$$\frac{\partial V(x^*,b)}{\partial b}.$$

The two are equal since $\partial V / \partial x = 0$ at the optimum.

Figure 2 illustrates why this property is called the envelope property. The heavy curve represents the maximum value, $V^*(b) = V(x^*(b),b)$, of the criterion function corresponding to different values of the parameter. The lighter curves describe the value of the criterion achieved with fixed values, $\bar{x}$ (and $\bar{x}'$) of x, as b is varied. The values and the slopes of the two types of curves, $V(x^*(b),b)$ and $V(\bar{x},b)$, are equal at the value of b for which x is optimal, that is, where $\bar{x} = x^*(b)$. The two curves are tangent at that point, and $V(\bar{x},b)$ is below $V^*(b)$ everywhere else, since

$V^*(b)$ is the maximum: $V^*(b) > V(\bar{x},b)$ if $\bar{x} = x^*(b)$. This holds for any $\bar{x}$ and the curve $V^*(b)$ is the envelope of the curves $V(\bar{x},b)$.

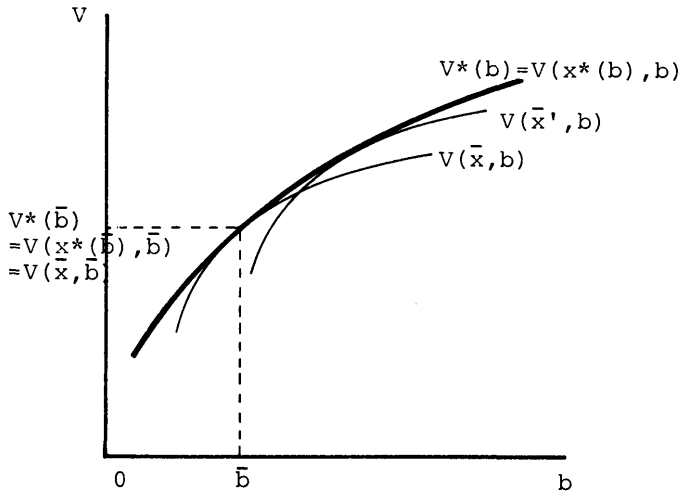

Figure 2.

Figure 2 suggests another way of proving the envelope property. Since $V^*(b)$ is maximum, $V^*(b) \geq V(\bar{x},b)$ for any b and $V^*(b) = V(\bar{x},b)$ if $\bar{x} = x^*(b)$. This implies that $V(\bar{x},b)$ lies below $V^*(b)$ everywhere and the two coincide at the value of b for which $\bar{x}$ is optimal. If the two curves are smooth, this is possible only when the two curves are tangent at this point, which proves the envelope property: $dv^*(b)/db = \partial V(\bar{x},b)/\partial b$ if $\bar{x} = x^*(b)$.

The envelope property appears in many areas of economics. Probably the most famous application is the relationship between the long-run cost curve and the short-run cost curve. The short-run cost curve is obtained when only a subset of factors are optimally chosen, and the long-run cost curve when all factors are chosen optimally. In the short run some factor inputs are fixed whereas in the long run they become variable and can be chosen optimally. Cost curves describe the minimized cost as functions of the output. The argument in the last proof of the envelope property can be applied to show that the long-run cost curve is an envelope of short-run cost curves. [1]

Another important example is concerned with benefits of a public good. Consider a household with the utility function, $u(z,h,X)$, where z is the composite consumer good and the numeraire, h is the lot size, and X is the supply of a public good. For a given consumption bundle the marginal benefit of the public good is $\partial u(z,h,X)/\partial X$. When the consumption bundle is optimally chosen, the maximum utility level depends on the income, *I*, the land rent, *R*, and the level of the public good,

[1] See Dixit (1976).

X, and it can be described by the indirect utility function, $v(R,I,X)$. For the optimum consumption bundle the marginal benefit of the public good is $\partial v(R,I,X)/\partial X$. The envelope property implies that $\partial u(z,h,X)/\partial X = \partial v(R,I,X)/\partial X$ if z and h are optimal given R, I, and X. This result is used in Chapter III.

2. The Envelope Theorem

Consider the problem of maximizing the criterion function $f(x,b)$ subject to the constraints $g_j(x,b)=0,\ j=1,2,\ldots,m,$ with respect to the vector $x=(x_1,x_2,\ldots,x_n)$ for a fixed vector of parameters $b=(b_1,b_2,\ldots,b_q)$. Let $x^*(b)$ be the optimal choice for this problem. Then granted a certain regularity condition[2] there exists the vector of Lagrange multipliers $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_m)$ such that $x^*(b)$ maximizes the Lagrangian

$$\Phi(x,\lambda,b) = f(x,b) + \lambda \cdot g(x,b) \tag{2.1}$$

without any constraint, where $g(x,b)=(g_1(x,b), g_2(x,b),\ldots, g_m(x,b))$, and the dot between λ and $g(x,b)$ denote the inner product so that

$$\lambda \cdot g(x,b) = \sum_{j=1}^{h} \lambda_j g_j(x,b).$$

If $f(x,b)$ and $g(x,b)$ are differentiable with respect to x, the optimal choice, $x=x^*(b)$, satisfies the first order necessary conditions,

$$\partial f(x,b)/\partial x_i + \lambda \cdot \partial_g(x,b)/\partial x_i = 0, \qquad i=1,2,\ldots,n. \tag{2.2}$$

The *Envelope Theorem* describes a relationship between the *maximum value* function $f^*(b)=f(x^*(b),b)$ and the Lagrangian $\Phi(x,\lambda,b)$.

---

[2] The condition is called the *Jacobian condition*, and requires that the Jacobian matrix of first order partial derivatives of constraint functions,

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1}, & \frac{\partial g_1}{\partial x_2}, & \ldots, & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1}, & \frac{\partial g_2}{\partial x_2}, & \ldots, & \frac{\partial g_2}{\partial x_n} \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots\ldots & \ldots\ldots \\ \frac{\partial g_m}{\partial x_1}, & \frac{\partial g_m}{\partial x_2}, & \ldots, & \frac{\partial g_m}{\partial x_n} \end{bmatrix}$$

be of full row rank m at the optimum. In nonlinear programming which deals with the more general case which includes inequality constraints, a similar condition, called the *constraint qualification*, must be satisfied.

The Envelope Theorem: Assume that $f^*(b)$ and $\Phi(x,\lambda,b)$ are continuously differentiable in b. Then at $x = x^*(b)$,

$$\partial f^*(b)/\partial b_k = \partial \phi(x,\lambda,b)/\partial b_k, \qquad k = 1,2,\ldots,q. \tag{2.3}$$

Proof:

Since $x^*(b)$ satisfies the constraint $g(x^*(b),b) = 0$ for any b, we have

$$\sum_{i=1}^{n} (\partial g/\partial x_i)(\partial x_1^*(b)/\partial b_k) + \partial g/\partial b_k = 0, \qquad k = 1,2,\ldots,q. \tag{2.4}$$

By the definition of the maximum value function and the first order condition (2.2), we obtain

$$\begin{aligned} \partial f^*(b)/\partial b_k &= \sum_{i=1}^{n} (\partial f/\partial x_i)(\partial x_1^*/\partial b_k) + \partial f/\partial b_k \\ &= -\lambda \cdot \sum_{i=1}^{n} (\partial g/\partial x_i)(\partial x_1^*/\partial b_k) + \partial f/\partial b_k \end{aligned} \tag{2.5}$$

(2.4) now yields the desired result:

$$\begin{aligned} \partial f^*(b)/\partial b_k &= \lambda \cdot \partial g/\partial b_k + \partial f/\partial b_k \\ &= \partial \phi(x,\lambda,b)/\partial b_k. \end{aligned}$$

Q.E.D.

3. Applications: Properties of Indirect Utility Function and the Expenditure Function

Consider a consumer with a utility function, $u(x)$, where x is the consumption vector, $x \equiv (x_1, x_2, ..., x_n)$. The consumer maximizes the utility function subject to the budget constraint,

$$p \cdot x = I, \tag{3.1}$$

where p is the price vector, $p \equiv (p_1, p_2, ..., p_n)$, I the money income, and

$$p \cdot x = \sum_{i=1}^{n} p_i x_i.$$

The Lagrangian for this maximization problem is

$$\Psi = u(x) + \delta[I - p \cdot x], \tag{3.2}$$

where δ is the Lagrange multiplier associated with the budget constraint (3.1). The first order conditions are

$$\partial u / \partial x_i = \delta p_i, \qquad i = 1,2,\ldots,n \tag{3.3}$$

The optimal consumption depends on income and prices, and can be written as $x(p,I)$. Substituting $x = x(p,I)$ into the utility function $u(x)$ yields the maximum utility level, $v(p,I) \equiv u(x(p,I))$, which can be achieved at the given values of income and prices. $v(p,I)$ is called the *indirect utility function*. The Envelope Theorem, (2.3), may then be applied to examine the effect of a change in prices and the income on the maximized utility level:

$$\partial v / \partial p_i = -\delta x_i, \; i = 1,2,\ldots,n, \tag{3.4}$$

$$\partial v / \partial I = \delta \; . \tag{3.5}$$

The latter equality shows that the Lagrange multiplier equals the marginal contribution to the maximum utility level made by an increase in income, or the marginal utility of income. The multiplier is, therefore, interpreted as the *shadow value* of the monetary income in utility terms.

If a dollar increase in income is all spent on good *i*, the increase in utility is given by

$$\frac{\partial u / \partial x_i}{P_i}.$$

This is equal to the marginal utility of income which is obtained when the increase in income can be optimally distributed among all goods, since by (3.3) a marginal increase in expenditures increases the utility by the same amount, whichever good is purchased. Thus

$$\frac{\partial u / \partial x_i}{P_i} = \partial v / \partial I \qquad i = 1,2,\ldots,n.$$

(3.4) has the following interpretation. If the price of the *i*-th good is raised by a dollar per unit and consumption of the *i*-th good is fixed, expenditure on that good must increase by x_i dollars, and expenditure on other goods must decrease by the same amount. The utility level would therefore decline by x_i times the marginal utility of income. By (3.3) it does not matter if substitution occurs: at the optimum all goods have the same marginal utility per dollar expenditure.

Combining (3.4) and (3.5) yields *Roy's Identity*:

$$\begin{aligned} x_i &= -(\partial v(p,I) / \partial p_i) / (\partial v(p,I) / \partial I) \\ &\equiv \hat{x}_i(p,I), \qquad\qquad i = 1,2,\ldots,n, \end{aligned} \tag{3.6}$$

which is derived in Chapter I without using the Envelope Theorem. $\hat{x}_i(p,I)$ is the uncompensated (or Marshallian) demand function. This result is quite useful: demand functions can be obtained simply by differentiating the indirect utility function.

Next, consider the problem of minimizing the expenditure necessary to achieve a given utility level. In this problem, $p \cdot x$ is minimized under the constraint,

$$u(x) = u, \tag{3.7}$$

for a given u. The minimum expenditure level is a function of prices and the utility level, $E(p,u)$, which is called the *expenditure function.*

If λ is the Lagrange multiplier, the Lagrangian is

$$\Phi = p \cdot x + \lambda[u - u(x)], \tag{3.8}$$

and

$$p_i = \lambda(\partial u / \partial x_i), \quad i = 1,2,\ldots,n. \tag{3.9}$$

By the Envelope Theorem, (2.3),

$$\lambda = \partial E(p,u) / \partial u, \tag{3.10}$$

$$x_i = \partial E(p,u) / \partial p_i \equiv x_i(p,u). \tag{3.11}$$

The latter equation is usually called *Shephard's Lemma* and gives the compensated demand function $x_i(p,u)$.

It can be easily shown that the expenditure function is concave as a function of prices for any fixed utility level. Let p and p' be two arbitrary price vectors and x^* and $x^{*\prime}$ be corresponding optimal consumption vectors. Then

$$E(p,u) = p \cdot x^*$$

and

$$E(p',u) = p' \cdot x^{*\prime}.$$

Consider a new price vector $\hat{p} = tp + (1-t)p'$ for an arbitrary t between *0* and *1*, and the corresponding consumption vector $\hat{x}^*$. The following inequalities hold:

$$p \cdot x^* \leq p \cdot \hat{x}^*,$$

and

$$p' \cdot x^* \leq p' \cdot \hat{x}^*.$$

Multiplying the first inequality by t and the second by $1-t$ and adding them yields

$$\begin{aligned} E(tp+(1-t)p',u) &= (tp+(1-t)p')\cdot \hat{x}^* \\ &\geq tp\cdot x^* + (1-t)p'\cdot x^{*\prime} \\ &= tE(p,u)+(1-t)E(p',u). \end{aligned}$$

Thus $E(p,u)$ is concave with respect to p. If E is twice differentiable, the concavity implies

$$\partial^2 E(p,u)/\partial p_i^2 = \partial x_i(p,u)/\partial p_i \leq 0. \tag{3.12}$$

This shows that price increase for any good does not increase the uncompensated demand for that good, i.e., the own substitution effect is nonpositive. This is used in Equation (I.1.20) of Chapter 1.

Now, we derive the *Slutsky equation*, describing the relationship between the uncompensated and compensated demand functions. For given prices and income, utility maximization yields the indirect utility function $v(p,I)$ and the uncompensated demand function $\hat{x}_i(p,I),\ i=1,2,\ldots,n$. Consider the expenditure minimization given the maximum utility level $u=v(p,I)$. Unless some prices are zero, in which case some technical difficulty appears, the optimal choices coincide and $I=E(p,u)$. The uncompensated demand function therefore satisfies

$$x_i(p,u)=\hat{x}_i(p,E(p,u)),\qquad i=1,2,\ldots,n. \tag{3.13}$$

Differentiation of this equation with respect to P_j yields

$$\begin{aligned} \partial x_i(p,u)/\partial p_j &= \partial \hat{x}_i(p,I)/\partial p_j + [\partial \hat{x}_i(p,I)/\partial I][\partial E(p,u)/\partial p_j] \\ &= \partial \hat{x}_i(p,I)/\partial p_j + [\partial \hat{x}_i(p,I)/\partial I]x_j(p,u), \end{aligned} \tag{3.14}$$

where the last term results from substituting according to (3.10). This is the Slutsky equation,

$$\left.\frac{\partial x_i}{\partial P_j}\right|_{u=const.} = \left.\frac{\partial x_i}{\partial P_j}\right|_{I=const.} + x_j\frac{\partial x_i}{\partial I}, \tag{3.15}$$

used in deriving (V.2.27)

Compensated and uncompensated demand functions satisfy another relationship which is also used in deriving (V.2.27). Following an argument similar to that which led to (3.13), we obtain

$$x_i(p,v(p,I))=\hat{x}_i(p,I),\qquad i=1,\ldots,n.$$

Taking a partial derivative with respect to *I*, we obtain

$$\begin{aligned}&[\partial x_i(p,u)/\partial u][\partial v(p,I)/\partial I]\\&=\partial \hat{x}_i(p,I)/\partial I, \quad i=1,...,n.\end{aligned} \tag{3.16}$$

Notes

Discussions in this Appendix owe very much to Dixit (1976). The Envelope Theorem in section 2 was proved by Afriat (1971) and can also be found in Takayama (1974).

References

Afriat, S.N., (1971), "Theory of Maxima and the Method of Lagrange, " *SIAM Journal of Applied Mathematics* 20,

Dixit, A.K., (1976), *Optimization in Economic Theory*, (Oxford University Press, Oxford).

Takayama, A., (1974), *Mathematical Economics*, (Dryden Press, Hinsdale, IL).

APPENDIX IV
OPTIMAL CONTROL THEORY

This appendix provides a concise review of *optimal control theory*. Many economic problems require the use of optimal control theory. For example, optimization over time such as maximizations of utility over an individual's life time and of profit and social welfare of a country over time and optimization over space such as the ones analyzed in this book fit in its framework.

Although these problems may be solved by the conventional techniques such as *Lagrange's method* and *nonlinear programming* if we formulate the problems in discrete form by dividing time (or distance) into a finite number of intervals, continuous time (or space) models are usually more convenient and yield results which are more transparent. Optimization over continuous time, however, introduces some technical difficulties. In the continuous time model, the number of choice variables is no longer finite: since decisions may be taken at each instant of time, there is a continuously infinite number of choice variables. The rigorous treatment of optimization in an infinite-dimensional space requires the use of very advanced mathematics. Fortunately, once proven, the major results are quite simple, and analogous to those in the optimization in a finite-dimensional space.

There are three approaches in the optimal control theory: *calculus of variations*, the *maximum principle* and *dynamic programming*. Calculus of variations is the oldest among the three and treats only the interior solution. In applications, as it turned out, choice variables are often bounded, and may jump from one bound to the other in the interval considered. The maximum principle was developed to include such cases. Roughly speaking, calculus of variations and the maximum principle are derived by using some appropriate forms of differentiation in an infinite-dimensional space. Dynamic programming however, exploits the recursive nature of the problem. Many problems including those treated by calculus of variations and the maximum principle have the property that the optimal policy from any arbitrary time on depends only on the state of the system at that time and does not depend on the paths that the choice variables have taken up to that time. In such cases the maximum value of the objective function beyond time t can be considered as a function of the state of the system at time t. This function is called the *value function*. The value function yields the value which the best possible performance from t to the end of the interval achieves. The dynamic programming approach solves the optimization problem by first obtaining the value function. Although the maximum principle and dynamic programming yield the same results, where they can both be applied, dynamic programming is less general than the approach based on the maximum principle, since it requires differentiability of the value function.

We first try to facilitate an intuitive understanding of control theory in section 1. In order to do so, a very simple control problem is formulated and the necessary conditions for the optimum are derived heuristically. Following the dynamic programming approach, Pontryagin's maximum principle is derived from the *partial*

differential equation of dynamic programming. As mentioned above, this approach is not the most general one, but it facilitates economic interpretation of the necessary conditions. In section 2 the results in section 1 are applied to an example taken from Chapter VII. Section 3 considers a more general form of the control problem (due to Bolza and Hestenes) and Hestenes' theorem, giving the necessary conditions for the optimum, is stated without proof. This theorem is general enough to include most problems that appear in this book. Finally, in section 4, Hestenes' theorem is used to solve the control problems in Chapter I.

1. A Simple Control Problem

Consider a dynamic process which starts at *inital time* t_0 and ends at *terminal time* t_1. Both t_0 and t_1 are taken as given in this section. For simplicity, the state of the system is described by only one variable, $x(t)$, called the *state variable.* In most economic problems the state variable is usually a stock, such as the amounts of capital equipments and inventories available at time t. In Chapters IV and V of our book the volume of traffic at a radius is a state variable.

The state of the system is influenced by the choice of *control variables,* $u_1(t), u_2(t), \ldots, u_r(t)$, which are summarized as the *control vector*,

$$u(t) = (u_1(t), u_2(t), \ldots, u_r(t)). \tag{1.1}$$

The control vector must lie inside a given subset of a Euclidean r-dimensional space, U:

$$u(t) \in U, \qquad t_0 \le t \le t_1, \tag{1.2}$$

where U is assumed to be closed and unchanging. Note that control variables are chosen at each point of time. The rate of investment in capital equipment is one of the control variables in most models of capital accumulation; the rate of inventory investment is a variable in inventory adjustment models; and the population per unit distance is a control variable for the models in this book. An entire path of the control vector, $u(t)$, $t_0 \le t \le t_1$, is a vector-valued function $u(t)$ from the interval $[t_0, t_1]$ into the r-dimensional space and is simply called a *control.* A control is *admissible* if it satisfies the constraint (1.2) and some other regularity conditions which will be specified in section 3.

The state variable moves according to the differential equation

$$\frac{dx}{dt} = \dot{x}(t) = f_1(x(t), u(t), t), \tag{1.3}$$

where f_1 is assumed to be continuously differentiable. Notice that the function f_1, is not the same as f_0. In this section the *initial state*, $x(t_0)$, is given,

$$x(t_0) = x^0, \tag{1.4}$$

where x^0 is some constant, but the *terminal state*, $x(t_1)$, is unrestricted. For example, the capital stock at initial time is fixed; the rate of change of the capital stock equals the rate of investment minus depreciation; and the capital stock at terminal time is not restricted.

The problem to be solved is that of maximizing the *objective functional*

$$J = \int_{t_0}^{t_1} f_0(x(t), u(t), t)dt + S_0(x(t_1), t_1) \tag{1.5}$$

with respect to the control vector, $u(t)$, $t_0 \le t \le t_1$, subject to the constraints (1.2), (1.3), and (1.4), where f_0 and S_0, the functions which make up the objective functional are continuously differentiable. A *functional* is defined as a function of a function or functions, that is, a mapping from a space of functions to a space of numbers. In the investment decision problem for a firm, for example, $f_0(x(t), u(t), t)dt$ is the amount of profit earned in the time interval $[t, t+dt]$ and $S_0(x(t_1), t_1)$ is the scrap value of the amount of capital $x(t_1)$ at terminal time t_1.

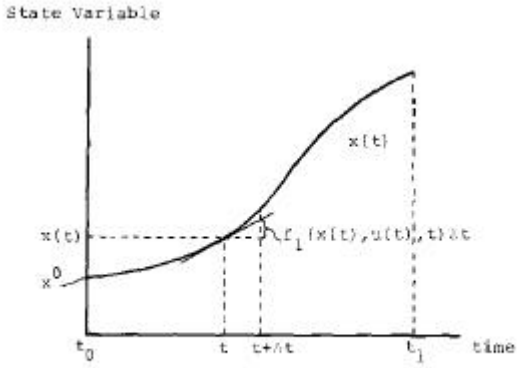

Figure 1a. A Trajectory of the State Variable.

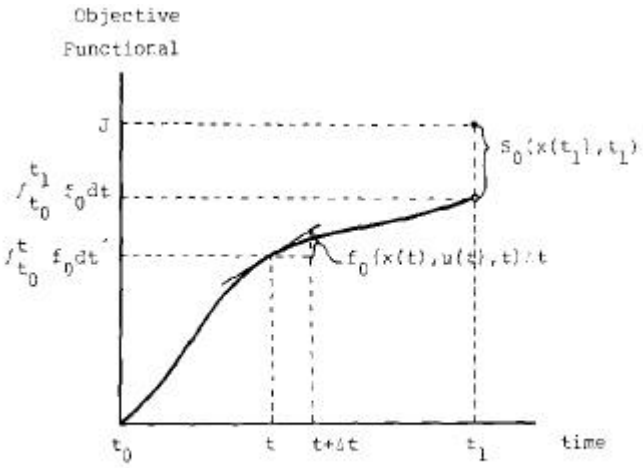

Figure 1b. The Objective Functional.

The problem is illustrated in Figure 1. In Fig.la, a possible trajectory of the state variable with the initial value x^0 is depicted. If the trajectory of the control vector is specified for the entire time horizon $[t_0, t_1]$, the trajectory of the state variable is completely characterized. The value of the state variable at time t and the choice of the control vector then jointly determine $f_0(x(t), u(t), t)$.

In Fig.1b we graph the part of the value of the objective functional which has been realized at any time t for the particular trajectory of the control vector f_0, therefore, appears as the slope in Fig.1b, while the value of the objective functional is the sum of the integral from t_0 to t_1, of f_0, and S_0, the scrap value at terminal time. Our problem is to obtain the trajectory of the control vector that maximizes the objective functional.

The major difficulty of this problem lies in the fact that an entire time path of the control vector must be chosen. This amounts to a continuously infinite number of control variables. In other words, what must be found is not just the optimal *numbers* but the optimal *functions*. The basic idea of control theory is to transform the problem of choosing the entire optimal path of control variables into the problem of finding the optimal values of control variables at each instant of time. In this way the problem of choosing an infinite number of variables is decomposed into an infinite number of more elementary problems each of which involves determining a finite number of variables.

The objective functional can be broken into three pieces for any time t – a past, a present and a future – :

$$\begin{aligned} J = &\int_{t_0}^{t} f_0(x(t)', u(t)', t')dt' \\ &+ \int_{t}^{t+\Delta t} f_0(x(t)', u(t)', t')dt' \\ &+ \int_{t+\Delta t}^{t_1} f_0(x(t)', u(t)', t')dt' + S_0(x(t_1), t_1). \end{aligned}$$

The decisions taken at any time have two effects. They directly affect the present term,

$$\int_{t}^{t+\Delta t} f_0(x(t)', u(t)', t')dt',$$

by changing f_0. They also change $\dot{x}$, and hence the future path of $x(t)$, through $\dot{x} = f_1(x(t), u(t), t)$. The new path of $x(t)$ changes the future part of the functional. For example, if a firm increases investment at time *t*, the rate at which profits are earned at that time falls because the firm must pay for the investment. The investment, however, increases the amount of capital available in the future and therefore profits earned in the future. The firm must make investment decisions weighing these two effects. In general, the choice of the control variables at any instant of time must take into account both the instantaneous effect on the current earnings $f_0 \Delta t$ and the indirect effect on the future earnings $\int_{t+\Delta t}^{t_1} f_0 dt' + S_0$ through a change in the state

variable. The transformation of the problem is accomplished if a simple way to represent these two effects is found.

This leads us to the concept of the *value function*, which might be used by a planner who wanted to recalculate the optimal policy at time t after the dynamic process began. Consider the problem of maximizing

$$\int_t^{t_1} f_0(x(t'),u(t'),t')dt' + S_0(x(t_1),t_1) \tag{1.6}$$

when the state variable at time t is x; $x(t) = x$. The maximized value is then a function of x and t:

$$J^*(x,t), \tag{1.7}$$

which is called the *value function*. The optimal value of the objective functional for the original problem (1.2)-(1.5) is

$$J^*(x^*(t),t) = J^*(x^0,t_0). \tag{1.8}$$

The usefulness of the value function must be obvious by now: it facilitates the characterization of the indirect effect through a change in the state variable by summarizing the maximum possible value of the objective functional from time t on as a function of the state variable at time t (and t).

The next step in the derivation of the necessary conditions for the optimum involves the celebrated *Principle of Optimality* due to Bellman. The principle exploits the fact that the value of the state variable at time t captures all the necessary information for the decision making from time t on: the paths of the control vector and the state variable up to time t do not make any difference as long as the state variable at time t is the same. This implies that if a planner recalculates the optimal policy at time t given the optimal value of the state variable at that time, the new optimal policy coincides with the original optimal policy. Thus if $u^*(t), t_0 \le t \le t_1$, is the optimal control for the original problem and $x^*(t), t_0 \le t \le t_1$, the corresponding trajectory of the state variable, the value function satisfies

$$J^* = \int_t^{t_1} f_0(x^*(t'),u^*(t'),t')dt' + S_0(x^*(t_1),t_1). \tag{1.9}$$

Applying the principle of optimality again, we can rewrite (1.9) as

$$\begin{aligned} J^*(x^*(t),t) &= \int_t^{t+\Delta t} f_0(x^*(t'),u^*(t'),t')dt' + \int_{t+\Delta t}^{t_1} f_0(x^*(t'),u^*(t'),t')dt' \\ &\quad + S_0(x^*(t_1),t_1) \\ &= \int_t^{t+\Delta t} f_0(x^*(t'),u^*(t'),t')dt' + J^*(x^*(t+\Delta t),t+\Delta t), \end{aligned} \tag{1.10}$$

for any t and $t+\Delta t$ such that $t_0 \le t \le t+\Delta t \le t_1$. This construction allows us to

concentrate on the decisions in the short interval from t to $t+\Delta t$ by summarizing the outcome in the remaining period in the value function, $J^*(x^*(t+\Delta t), t+\Delta t)$.

By the definition of the value function, any admissible control cannot do better than the value function if the initial state is the same. Consider the following special type of control, $u(t')$, $t \le t' \le t_1$: the control is arbitrary between time t and time $t+\Delta t$ and optimal in the remaining period given the state reached at time $t+\Delta t$. Then the corresponding value of the objective functional satisfies

$$J^*(x^*(t), t) \ge \int_t^{t+\Delta t} f_0(x(t'), u(t'), t')dt' + J^*(x(t+\Delta t), t+\Delta t) \qquad (1.11)$$

where $x(t')$, $t \le t' \le t_1$, is the state variable corresponding to the control $u(t')$ with the initial state $x(t) = x^*(t)$.

Combining (1.10) and (1.11) yields

$$\begin{aligned} J^*(x^*(t), t) &= \int_t^{t+\Delta t} f_0(x^*(t'), u^*(t'), t')dt' + J^*(x^*(t+\Delta t), t+\Delta t) \\ &\ge \int_t^{t+\Delta t} f_0(x(t'), u(t'), t')dt' + J^*(x(t+\Delta t), t+\Delta t) \end{aligned}$$

$$\text{for any } u(t') \in U, t \le t' \le t+\Delta t. \qquad (1.12)$$

This shows that the optimal control in the interval $[t, t+\Delta t]$ maximizes the sum of the objective functional in the interval and the maximum possible value of the functional in the rest of the period $[t+\Delta t, t_1]$. If both sides of the inequality are differentiable, Taylor's expansion around t yields[1]

---

[1] The details of Taylor's expansion here are as follows. Taylor's theorem states that if $F(t)$ is differentiable at $t=a$, then

$$F(t) = F(a) + (t-a)F(a) + o(t-a),$$

where $\lim_{t-a\to 0} \frac{o(t-a)}{t-a} = 0$. Noting that

$$F_0(t+\Delta t) \equiv \int_t^{t+\Delta t} f_0(t')dt'$$

satisfies

$$F_0'(t) = f_0(t),$$

we obtain

$$\begin{aligned} &\int_t^{t+\Delta t} f_0(x^*(t'), u^*(t'), t')dt' + J^*(x^*(t+\Delta t), t+\Delta t) \\ &= f_0(x^*(t), u^*(t), t)\Delta t + J^*(x^*(t), t) \\ &\quad + [(\partial J^*(x^*(t), t)/\partial x)\dot{x}^*(t) + \partial J^*(x^*(t), t)/\partial t]\Delta t + o(\Delta t), \end{aligned}$$

and

$$
\begin{aligned}
&-(\partial J^*(x^*(t),t)/\partial t)\Delta t \\
&= f_0(x^*(t),u^*(t),t)\Delta t + (\partial J^*(x^*(t),t)/\partial x) f_1(x^*(t),u^*(t),t)\Delta t + \ldots \\
&\geq f_0(x^*(t),u(t),t)\Delta t + (\partial J^*(x^*(t),t)/\partial x) f_1(x^*(t),u(t),t)\Delta t + \ldots,
\end{aligned}
$$

$$\text{for any} \quad u(t) \in U, \tag{1.13}$$

where ... represents higher order terms which become negligible as Δt tends to zero, since they approach zero faster than Δt. Note that we used $x(t) = x^*(t)$, $\dot{x}(t) = f_1(x(t),u(t),t)$ and $\dot{x}^*(t) = f_1(x^*(t),u^*(t),t)$.

Inequality (1.13) has a natural economic interpretation. For example, if a firm is contemplating the optimal capital accumulation policy, $f_0(x^*(t),u(t),t)\Delta t$, is approximately the amount of profits earned in the period $[t, t+\Delta t]$. $\partial J^*(x^*(t),t)/\partial x$ is the marginal value of capital, or the contribution of an additional unit of capital at time t; and $f_1(x^*(t),u(t),t)\Delta t = \dot{x}(t)\Delta t$ is approximately the amount of capital accumulated in period $[t, t+\Delta t]$. Thus $(\partial J^*/\partial x) f_1 \Delta t$ represents the value of capital accumulated during the period. (1.13), therefore, shows that the optimal control vector maximizes the sum of the current profits and the value of increased capital.

Dividing (1.13) by At and taking limits as At approaches zero, we obtain

$$
\begin{aligned}
&-\partial J^*(x^*(t),t)/\partial t \\
&= f_0(x^*(t),u^*(t),t) + (\partial J^*(x^*(t),t)/\partial x) f_1(x^*(t),u^*(t),t) \\
&\geq f_0(x^*(t),u(t),t) + (\partial J^*(x^*(t),t)/\partial x) f_1(x^*(t),u(t),t)
\end{aligned}
$$

$$\text{for any} \quad u(t) \in U. \tag{1.14}$$

Thus the optimal control vector $u^*(t)$ maximizes

$$f_0(x^*(t),u,t) + (\partial J^*(x^*(t),t)/\partial x) f_1(x^*(t),u,t) \tag{1.15}$$

at each instant of time, and we have finally transformed the problem of finding the optimal path to that of finding optimal numbers at each point in time. From the above discussion, it must be clear that (1.15) summarizes both the instantaneous effect and the indirect effect through a change in the state variable.

(1.14) can be rewritten as

$$
\begin{aligned}
&\int_t^{t+\Delta t} f_0(x(t'),u(t'),t')dt' + J^*(x(t+\Delta t),t+\Delta t) \\
&= f_0(x(t),u(t),t)\Delta t + J^*(x(t),t) \\
&\quad + [(\partial J^*(x(t),t)/\partial x)\dot{x}(t) + \partial J^*(x(t),t)/\partial t]\Delta t + o(\Delta t) \\
&= f_0(x^*(t),u(t),t)\Delta t + J^*(x^*(t),t) \\
&\quad + [(\partial J^*(x^*(t),t)/\partial x)\dot{x}(t) + \partial J^*(x^*(t),t)/\partial t]\Delta t + o(\Delta t),
\end{aligned}
$$

where we used $x(t) = x^*(t)$. Substituting these two equations into (1.12) yields (1.13).

$$-\partial J^*/\partial t = \max_{\{u \in U\}}\left[f_0(x^*(t),u,t) + (\partial J^*/\partial x) f_1(x^*(t),u,t)\right]. \tag{1.14'}$$

This equation holds for any x, not just $x^*(t)$, and can be considered a partial differential equation of $J^*(x,t)$. It is called the *partial differential equation of dynamic programming* or *Bellman's equation*.

In the dynamic programming approach, the right side of (1.14') is maximized with respect to u, yielding the partial differential equation. The partial differential equation is then solved with the boundary conditions. At the initial time $t_0, x(t_0) = x^0$, while at the terminal time t_1, the value function satisfies

$$J^*(x(t_1),t_1) = S_0(x(t_1),t_1) \tag{1.16}$$

for any $x(t_1)$. This equation is the *terminal boundary condition* associated with Bellman's equation. Since (1.16) holds for any $x(t_1)$, we have

$$\partial J^*(x(t_1),t_1)/\partial x = \partial S_0(x(t_1),t_1)/\partial x, \tag{1.17}$$

which is called the *transversality condition* at time t_1.

One of the disadvantages of the dynamic programming approach is that the partial differential equation is usually hard to solve. *Pontryagin's maximum principle*, which can be immediately derived from the partial differential equation of dynamic programming, is often more useful for economic applications. Furthermore, the method of dynamic programming employs the Taylor expansion in (1.13), which requires that the value function be differentiable. There are many problems for which the value function is not differentiable everywhere. The maximum principle, however, can be proven using a different and more general method. In this section we derive the maximum principle from Bellman's equation, and in Section 3 we state a more general version of the maximum principle without proof.

To derive Pontryagin's maximum principle, we define the *adjoint*, or *costate*, or *auxiliary*, *variable*,

$$p(t) = \partial J^*(x^*(t),t)/\partial x, \tag{1.18}$$

and rewrite (1.15) as the *Hamiltonian*,

$$H\left[x(t),u(t),t,p(t)\right] = f_0(x(t),u(t),t) + p(t) f_1(x(t),u(t),t). \tag{1.19}$$

(1.14') now reads: if $u^*(t)$ is the optimal control and $x^*(t)$ the associated path of the state variable, then there exists a $p(t)$ such that for any t

$$H\left[x^*(t),u^*(t),t,p(t)\right] = \max_{\{u \in U\}} H\left[x^*(t),u,t,p(t)\right] \tag{1.20}$$

Since $p(t)$ equals $\partial J^*/\partial x$, the adjoint variable $p(t)$ is the marginal value of the state variable (if, for example, $x(t)$ is capital, $p(t)$ is the marginal value of

capital) and has the interpretation of the *shadow price* of $x(t)$.

(1.14') also contains information about the adjoint variable. We can rewrite (1.14') as the *Hamilton-Jacobi equation*:

$$-\partial J^*/\partial t = H(x^*, u^*, t, \partial J^*/\partial x). \tag{1.21}$$

If the value function is twice differentiable, the derivative of (1.21) with respect to x can be taken:

$$-\partial^2 J^*/\partial x \partial t = \partial H/\partial x + (\partial H/\partial p)\partial^2 J^*/\partial x^2. \tag{1.22}$$

Differentiating (1.18) with respect to t, however, yields

$$\dot{p} = (\partial^2 J^*/\partial x^2)\dot{x}^* + \partial^2 J^*/\partial t \partial x. \tag{1.23}$$

If we further assume twice continuous differentiability, the second order mixed partial derivatives are equal whatever the order of differentiation: $\partial^2 J^*/\partial x \partial t = \partial^2 J^*/\partial t \partial x$. Since from (1.19) and (1.3) we have

$$\dot{x}^* = (\partial/\partial p)H(x^*, u^*, t, p), \tag{1.24}$$

we can substitute (1.22) and (1.24) into (1.23) to get

$$-\dot{p} = (\partial/\partial x)H(x^*, u^*, t, p). \tag{1.25}$$

Equation (1.25) is often called the *adjoint equation* and the pair, (1.24) and (1.25), is called the *canonical equations of the maximum principle*.

The transversality condition (1.17) gives the value of the adjoint variable at time t_1:

$$p(t_1) = \partial S_0(x^*(t_1), t_1)/\partial x. \tag{1.26}$$

Finally, the time derivative of the Hamiltonian along the optimal path is

$$\frac{dH}{dt} = \frac{\partial H}{\partial x}\dot{x}^* + \frac{\partial H}{\partial p}\dot{p} + \frac{\partial H}{\partial u}\dot{u}^* + \frac{\partial H}{\partial t}.$$

From (1.24) and (1.25), the sum of the first two terms on the RHS is zero. The third term vanishes because either $\partial H/\partial u = 0$ for an interior solution or $\dot{u} = 0$ for a boundary solution.
Thus we have

$$\frac{dH}{dt} = \frac{\partial H}{\partial t} \tag{1.27}$$

except when the control vector has a jump.

The maximum principle approach solves the ordinary differential equations (1.24)

and (1.25) with the boundary conditions $x(t_0) = x^0$ and (1.26). Since boundary conditions are given at two points, i.e., at initial time t_0 and terminal time t_1, this problem is called a *two-point boundary value problem*. The pair of ordinary differential equations are usually easier to solve than the partial differential equation of dynamic programming.

2. An Example: Optimal Growth of Cities

Consider the problem which was formulated in section 3 of Chapter VII: maximize

$$\int_0^{\infty} [U(c(t), P(t)) - u^*]dt \tag{2.1}$$

subject to the differential equation,

$$\dot{k}(t) = f(k(t), P(t)) - \lambda k(t) - c(t), \tag{2.2}$$

and the initial condition,

$$k(0) = k_0, \tag{2.3}$$

where control variables are the per capita consumption of resources, $c(t)$, and the population of a city, $P(t)$; the state variable is the capital stock, $k(t)$; λ is the growth rate of the whole population; and u^* is the utility level at the optimal steady state.

The fact that the terminal time is infinite causes some complications. We first solve the finite-horizon problem of maximizing

$$\int_0^{t_1} [U(c(t), P(t) - u^*]dt + S_0(k(t_1), t_1) \tag{2.4}$$

subject to the same constraints.

The Hamiltonian for this problem is

$$H(k(t), c(t), P(t), t, q(t)) = U(c(t), P(t)) - u^* + q(t)[f(k(t), P(t) - \lambda k(t) - c(t)], \tag{2.5}$$

where $q(t)$ is the adjoint variable associated with the differential equation (2.2). Discussions in the previous section show that $q(t)$ can be interpreted as the marginal value of capital.

According to (1.20), the Hamiltonian must be maximized with respect to the control variables, $c(t)$ and $P(t)$. Assuming an interior solution, we obtain

$$U_c(c(t), P(t) = q(t), \tag{2.5}$$

$$U_p(c(t), P(t) = q(t) f_P(k(t), P(t)), \tag{2.6}$$

which are equations (VII.3.8a) and (VII.3.8b) in Chapter VII.

$q(t)$ satisfies the adjoint equation,

$$-\dot{q}(t) = \partial H / \partial k = q(t)\left[f_k(k(t), P(t)) - \lambda\right], \tag{2.7}$$

which is (VII.3.7).

The transversality condition at $t = t_1$ is

$$q(t_1) = \partial S_0(k(t_1), t_1) / \partial k. \tag{2.8}$$

In the case where the terminal time is infinite, a straightforward application of the transversality condition (1.26) would yield

$$\lim_{t \to \infty} q(t) = 0.$$

It can be shown, however, that this is not the correct transversality condition. As shown in Chapter VII, the optimal path converges to the optimal steady state at which

$$U(c, P) - u^*$$

is maximized subject to the constraint,

$$f(k, P) - \lambda k - c = 0 .$$

Denoting the values of variables at the optimal steady state by asterisks, we can write the transversality condition as

$$\lim_{t \to \infty} q(t)k(t) = q^* k^*, \tag{2.9}$$

where $q^* = U_c(c^*, P^*)$.

3. The Maximum Principle: The Problem of Hestenes and Bolza

In this section the problem in section 1 is generalized in a number of respects. Differences from the problem in section 1 are as follows.

(i) The number of state variables is arbitrary.

(ii) Control parameters are added. Control parameters are choice variables which are restricted to be constant for any t.

(iii) The constraints on the control vector may depend on the state vector, control parameters, and time.

(iv) Isoperimetric constraints, or constraints involving integrals, are added.

(v) The initial time t_0 and the terminal time t, may be chosen by the choice of control parameters.

(vi) The initial state $x(t_0)$ and the terminal state $x(t_1)$ can also be chosen by the choice of control parameters.

The problem to be solved is that of maximizing the objective functional,

$$J = \int_{t_0}^{t_1} f_0(x(t), u(t), b, t)dt + S_0(b), \tag{3.1}$$

subject to the constraints,

$$\dot{x}_i = f_i(x(t), u(t), b, t), \qquad i = 1,2,...n; \tag{3.2a}$$

$$g_j(x(t), u(t), b, t) \geqq 0, \qquad j = 1,2,..m'; \tag{3.2b}$$

$$g_j(x(t), u(t), b, t) = 0, \qquad j = m'+1, m'+2,...m; \tag{3.2c}$$

$$\int_{t_0}^{t_1} h_k(x(t), u(t), b, t)dt + S_k(b) \geq 0, \quad k = 1,2,..., \ell'; \tag{3.2d}$$

$$\int_{t_0}^{t_1} h_k(x(t), u(t), b, t)dt + S_k(b) = 0, \quad k = \ell'+1, \ell'+2,..., \ell; \tag{3.2e}$$

$$t_0 = t_0(b); \tag{3.2f}$$

$$t_1 = t_1(b); \tag{3.2g}$$

$$x_i(t_0) = x_i^0(b), \qquad i = 1,2,...n; \tag{3.2h}$$

$$x_i(t_1) = x_i^1(b), \qquad i = 1,2,...n. \tag{3.2i}$$

$x(t) = (x_1(t), x_2(t),..., x_n(t))$ is the state vector; $u(t) = (u_1(t), u_2(t),..., u_r(t))$ is the control vector; $b = (b_1, b_2,..., b_q)$ is the vector of control parameters; $(x(t), u(t), t)$ lies in a set R_0 in (x, u, t) space; and b lies in an open set B. The maximization is carried out with respect to the control vector and control parameters. S_0, S_k, f_0, f_i, g_j, h_k, x_i^0, x_i^1, t_0, and t_1, are all assumed to be continuously differentiable.

Now, define a set A as the subset of $R_0 \times B$ satisfying

$$g_j(x, u, b, t) \geq 0, \qquad j = 1,2,..m'$$

$$g_j(x, u, b, t) = 0, \qquad j = m'+1, m'+2,..m$$

The set A is called the set of admissible elements.

The constraints are assumed to satisfy the condition that the matrix G, defined as

$$G \equiv \begin{bmatrix} \frac{\partial g_1}{\partial u_1}, \frac{\partial g_1}{\partial u_2}, \cdots, \frac{\partial g_1}{\partial u_r}, g_1, 0, 0, \cdots, 0 \\ \frac{\partial g_2}{\partial u_1}, \frac{\partial g_2}{\partial u_2}, \cdots, \frac{\partial g_2}{\partial u_r}, 0, g_2, 0, \cdots, 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \frac{\partial g_m}{\partial u_1}, \frac{\partial g_m}{\partial u_2}, \cdots, \frac{\partial g_m}{\partial u_r}, 0, 0, 0, \cdots, g_m \end{bmatrix} \tag{3.3}$$

has rank m. This condition is called the *constraint qualification*.

The necessary conditions for the maximization problem can be stated as the following Theorem, which is due to Hestenes (1965).

Theorem: *Suppose the trajectory* $\{(x^*(t), u^*(t), b^*) : t_0 \le t \le t_1\}$ *maximizes* (3.1) *subject to the constraint* (3.2) *among the trajectories whose* $x(t)$ *is continuous*, $u(t)$ *piecewise continuous*, (continuous except possibly for a finite number of discrete jumps), $(x(t), u(t), t) \in R_0$, *and* $b \in B$. *Assume the constraint qualification* (3.3) *holds for any* (x, u, b, t) *in the set of admissible elements*, A. *Then there exist multipliers*;

$$P = (P_0, P_1, ..., P_n),$$

$$\lambda = (\lambda_1, \lambda_2, ..., \lambda_m),$$

$$\mu = (\mu_1, \mu_2, ..., \mu_l),$$

not vanishing simultaneously on $t_0 \le t \le t_1$, *and functions H and L where*

$$H(x(t), u(t), b, t, p(t), \mu)$$
$$= p_0 f_0(x(t), u(t), b, t) + \sum_{i=1}^{n} P_i(t) f_i(x(t), u(t), b, t) + \sum_{k=1}^{\ell} \mu_k h_k(x(t), u(t), b, t)$$

$$L(x(t), u(t), b, t, p(t), \mu, \lambda(t))$$
$$= H(x(t), u(t), b, t, p(t), \mu) + \sum_{j=1}^{m} \lambda_j(t) g_j(x(t), u(t), b, t)$$

such that the following relations hold;

(a) *The multipliers* $p_0, \mu_k, k = 1, 2, ..., \ell$, *are constant*, $p_0 \ge 0$, and $\mu_k \ge 0, k = 1, 2, ..., \ell'$, *with*

$$\mu_k \left\{ \int_{t_0}^{t_1} h_k(x^*(t), u^*(t), b^*, t)dt + S_k(b^*) \right\} = 0, \qquad k = 1,2,...,\ell .$$

(b) *The multipliers* $\lambda_i(t),\ j = 1,2,...,m,$ *are piecewise continuous and are continuous over each interval of continuity of* $u^*(t)$. *Moreover, for each* $j = 1,2,...,m'$, *we have*

$$\lambda_j(t) \geq 0, \quad \lambda_j(t) g_j(x^*(t), u^*(t), b^*, t) = 0.$$

(c) *The multipliers* $p_i(t), i = 1,2,...,n$, *are continuous and have piecewise continuous derivatives. They satisfy the adjoint equations*;

$$-\dot{p}_i(t) = (\partial / \partial x_i) H(x^*(t), u^*(t), b^*, t, p(t), \mu), \qquad i = 1,2,...,n .$$

(d) *The maximum principle expressed in the inequality*

$$H(x^*(t), u^*(t), b^*, t, p(t), \mu) \geq H(x^*(t), u, b^*, t, p(t), \mu)$$

holds for all $[x^*(t), u, b^*, t,] \in A$, *which implies that*

$$(\partial / \partial u) L(x^*(t), u^*(t), b^*, t, p(t), \mu, \lambda(t)) = 0 .$$

(e) *The following transversality condition holds*:

$$\begin{aligned} \int_{t_0}^{t_1} \frac{\partial L^*}{\partial b_j} dt = & -p_0 \frac{\partial S_0}{\partial b_j} - \sum_{k=1}^{\ell} \mu \frac{\partial S_k}{k \partial b_j} \\ & + \left[-L^*(t_1) \frac{\partial t_1}{\partial b_j} + \sum_{i=1}^{n} p_i(t_1) \frac{\partial x_i^1}{\partial b_j} \right] \\ & - \left[-L^*(t_0) \frac{\partial t_0}{\partial b_j} + \sum_{i=1}^{n} p_i(t_0) \frac{\partial x_i}{\partial b_j} \right], \quad j = 1,2,...,q, \end{aligned}$$

where $L^*(t) = L(x^*(t), u^*(t), b^*, t, p(t), \mu, \lambda(t))$.

(f) *The function* $L^*(t)$ *is continuous on* $t_0 \leq t \leq t_1$, *and*

$$(d / dt) L^*(t) = (\partial / \partial t) L(x^*(t), u^*(t), b^*, t, p(t), \mu, \lambda(t))$$

on each interval of continuity of $u^*(t)$.

The reason why these conditions are necessary for the optimum can be understood by considering the following Lagrangian in the integral form:

$$\Lambda = P_0\left[\int_{t_0(b)}^{t_1(b)} f_0(x(t),u(t),b,t)dt + S_0(b)\right]$$
$$+ \int_{t_{0(b)}}^{t_1(b)}\{\sum_{i=1}^{n} p_i(t)[f_i(x(t),u(t),b,t) - \dot{x}_i(t)]$$
$$+ \sum_{j=1}^{m}\lambda_j(t)g_j(x(t),u(t),b,t]dt$$
$$+ \sum_{k=1}^{l}\mu_k\{\int_{t_0(b)}^{t_1(b)} h_k(x(t),u(t),b,t)dt + S_k(b)\}$$
$$+ \sum_{i=1}^{n}\gamma_i^0[x_i(t_0) - x_i^0(b)] + \sum_{\bar{i}=1}^{n}\gamma_i^1[x_{\bar{i}}(t_1) - x_i^1(b)].$$

Observing that integration by parts yields

$$\int_{t_0}^{t_1} P_i(t)\dot{x}_i(t)dt = \int_{t_0}^{t_1}\dot{P}_i(t)\dot{x}_i(t)dt - \int_{t_0}^{t_1}\dot{P}_i(t)x_i(t)dt$$
$$= P_i(t_i)x_i(t_1) - P_i(t_0)x_i(t_0) - \int_{t_0}^{t_1}\dot{P}_i(t)x_i(t)dt,$$

we can rewrite the Lagrangian as

$$\Lambda = \int_{t_0(b)}^{t_1(b)}\left\{L(x(t),u(t),b,t,p(t),\mu,\lambda(t)) + \sum_{i=1}^{n}\dot{P}_i x_i\right\}dt$$
$$- \sum_{i=1}^{n} p_i(t_1)x_i(t_1) + \sum_{i=1}^{n} p_i(t_0)x_i(t_0)$$
$$+ p_0 S_0(b) + \sum_{k=1}^{\ell}\mu_k S_k(b)$$
$$+ \sum_{i=1}^{n}\gamma_i^0\left[x_i(t_0) - x_i^0(b)\right] + \sum_{i=1}^{n}\gamma_i^1\left[x_i(t_1) - x_i^1(b)\right]$$

By analogy to the usual method of Lagrange, this Lagrangian must be maximized, without constraints, with respect to $u(t)$, b, $x(t)$, $x(t_0)$ and $x(t_1)$. Maximization of the Lagrangian with respect to $u(t)$ between t and $t+\Delta t$ is equivalent to maximization of

$$L(x(t),u(t),b,t,p(t),\mu,\lambda(t))\Delta t$$

with respect to $u(t)$. This yields condition (d).

In the same way, maximization with respect to $x(t)$ yields the adjoint equations in (c). Maximization with respect to $x_i(t_0), x_i(t_1)$ and b_j yields

$$\frac{\partial \Lambda}{\partial x_i(t_1)} = -p_i(t_1) + \lambda_i^1 = 0, \qquad i = 1,2,...,n,$$

$$\frac{\partial \Lambda}{\partial x_i(t_0)} = -p_i(t_0) + \gamma_i^0 = 0, \qquad i = 1,2,...,n,$$

$$\frac{\partial \Lambda}{\partial b_j} = p_0 \frac{\partial S_0}{\partial b_j} + \sum_{k=1}^{l} \mu_k \frac{\partial S_k}{\partial b_j} + L(t_1)\frac{\partial t_1}{\partial b_j} - L(t_0)\frac{\partial t_0}{\partial b_j} - \sum_{i=1}^{n} \gamma_i^0 \frac{\partial x_i}{\partial b_j} - \sum_{i=1}^{n} \gamma_i^1 \frac{\partial x_i^1}{\partial b_j} + \int_{t_0}^{t_1} \frac{\partial L}{\partial b_j}\, dt = 0, \qquad j = 1,2,...,q.$$

Condition (e) can be obtained by combining these equations.

Condition (f) is a generalization of (1.27) to allow for time dependent constraints (3.2b,c).

The multiplier p_0 is added to include the so-called *abnormal case* in which $p_0 = 0$. If $p_0 = 0$, the same control is optimal for problems with any objective functionals so long as all the constraints are the same. Thus for abnormal problems the necessary conditions do not involve the objective functional, but are already specified by constraints. This happens, for example, when there is only one control trajectory that satisfies all the constraints. If constraints are

$$x = u(t)^2,$$

$$-1 \le u(t) \le 1, \qquad t_0 \le t \le t_1,$$

$$x(t_0) = 0,$$

$$x(t_1) = 0,$$

then the only possible control trajectory is

$$u(t) = 0, \qquad t_0 \le t \le t_1,$$

and the optimal solution does not depend on the objective functional.

The reason why p_0 is zero in such a case can be seen by going back to the dynamic programming approach in section 1. Since the control cannot be changed, it is also impossible to change the state trajectory. This means that it is prohibitively costly to change the state trajectory: $\partial J^*/\partial x$ in (1.14') and hence $p(t)$ in (1.19) are infinite. Since p_0 was taken to be 1 in section 1, this is equivalent to $p_0 = 0$ with $p_i, \quad i = 1,...,n,$ finite in this section.

In this book, we assume that all the problems are normal, and normalize p_0 to be 1.

The constraint qualification is assumed because the proof of the maximum principle considers perturbation of the control vector $u(t)$ such as

$$\tilde{u}(t) = \begin{cases} v & \text{if } \tau - \varepsilon < t \le \tau \\ u(t) & \text{for other values of } t \in [t_0, t_1] \end{cases}$$

for a small ε, and derives the necessary conditions from the property that at the optimum no perturbation can make the objective functional greater. If the constraint qualification is not satisfied, there exist no nontrivial perturbations that satisfy the constraints (3.2b) and (3.2c). For example, if there are two equality constraints:

$$g_1(u_1, u_2) = 0,$$

$$g_2(u_1, u_2) = 0,$$

which are tangent only at a single point $u^* = (u_1^*, u_2^*)$ as in Figure 2, only one control vector satisfies the constraints and no perturbation is possible.

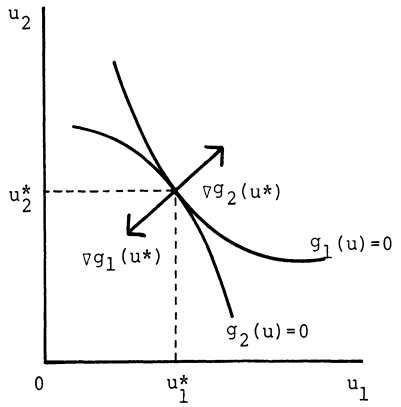

Figure 2. Constraint Qualification

In this case, the *gradient vectors*,

$$\nabla g_1(u^*) = \begin{bmatrix} \partial g_1(u^*)/\partial u_1 \\ \partial g_1(u^*)/\partial u_2 \end{bmatrix},$$

$$\nabla g_2(u^*) = \begin{bmatrix} \partial g_2(u^*)/\partial u_1 \\ \partial g_2(u^*)/\partial u_2 \end{bmatrix},$$

are linearly dependent and the rank of the matrix,

$$G = \begin{bmatrix} \partial g_1/\partial u_1, \partial g_1/\partial u_2, g_1, 0 \\ \partial g_2/\partial u_1, \partial g_2/\partial u_2, 0, g_2 \end{bmatrix} = \begin{bmatrix} \partial g_1/\partial u_1, \partial g_1/\partial u_2, 0, 0 \\ \partial g_2/\partial u_1, \partial g_2/\partial u_2, 0, 0 \end{bmatrix},$$

is less than $m = 2$.

4. Examples: Optimum Cities

Two optimum control problems formulated in Chapter 1 are solved in this section. Consider first the problem of maximizing the Benthamite social welfare function,

$$\int_0^{\bar{x}} u(z(x), h(x))N(x)dx, \tag{4.1}$$

subject to the resource constraint,

$$Pw - \int_0^{\bar{x}} \{[z(x) + t(x)]N(x) + R_a\theta(x)\}dx = 0 \tag{4.2}$$

the population constraint,

$$\int_0^{\bar{x}} N(x)dx - p = 0, \tag{4.3}$$

and the land constraint,

$$\theta(x) = N(x)h(x), \qquad 0 \le x \le \bar{x}. \tag{4.4}$$

Control variables are the consumption of the consumer good, $z(x)$, the consumption of land for housing, $h(x)$, and the population density, $N(x)$. The edge of the city, $\bar{x}$, is a control parameter. There is no state variable in this problem because there is no constraint in the form of a differential equation.

The function H in the previous section now reads

$$H(z(x), h(x), N(x), x, \lambda, \delta, \gamma) = \lambda u(z(x), h(x))N(x) - \delta\{[z(x) + t(x)]N(x) + R_a\theta(x)\} + \gamma N(x)$$

The function L is

$$L(z(x), h(x), N(x), x, \lambda, \delta, \gamma, \mu(x)) = H + \mu(x)[\theta(x) - N(x)h(x)]$$

and the Lagrangian Λ is

$$\Lambda = \int_0^{\bar{x}} L dx.$$

Assuming $\lambda > 0$, we normalize λ. With $\lambda = 1$, condition (d) yields

$$\partial L/\partial z(x) = [\partial u/\partial z - \delta]N(x) = 0$$

$$\partial L/\partial h(x) = [\partial u/\partial h - \mu]N(x) = 0$$

$$\partial L/\partial N(x) = u(x) - \delta[z(x) + t(x)] - \mu(x)h(x) = 0,$$

which corresponds to (I.2.5a), (I.2.5b), and (I.2.5c).

From condition (e), we obtain the transversality condition,

$$L^*(\bar{x}) = u(z(\bar{x}), h(\bar{x}))N(\bar{x}) - \delta\{[z(\bar{x}) + t(\bar{x})]N(\bar{x}) + R_a\theta(\bar{x})\} + \gamma N(\bar{x}) = 0,$$

which corresponds to (I.2.5d).

Condition (f) implies

$$dL^*(x)/dx = \mu(x)\theta'(x).$$

Next, we impose the constraint that households receive equal utility:

$$u = u(z(x), h(x)), \qquad 0 \le x \le \bar{x},$$

and maximize the sum of utilities,

$$\int_0^{\bar{x}} uN(x)dx.$$

Constraints, (4.2), (4.3), and (4.4), remain the same. In this case, u is an additional control parameter. Define

$$H(z(x), h(x), N(x), u, x, \lambda, \delta, \gamma)$$
$$= \lambda uN(x) - \delta\{[z(x) + t(x)]N(x) + R_a\theta(x)\} + \gamma N(x)$$

$$L(z(x), h(x), N(x), u, x, \lambda, \delta, \gamma, \upsilon(x), \mu(x))$$
$$= H + \upsilon(x)[u(z(x), h(x)) - u] + \mu(x)[\theta(x) - N(x)h(x)]$$

$$\Lambda = \int_0^{\bar{x}} L dx.$$

Again, we normalize λ. Condition (d) becomes

$$\partial L/\partial z(x) = -\delta N(x) + \upsilon(x)\partial u/\partial z = 0$$

$$\partial L/\partial h(x) = \upsilon(x)\partial u/\partial h - \mu(x)N(x) = 0$$

$$\partial L/\partial N(x) = u - \delta[z(x) + t(x)] - \mu(x)h(x) = 0,$$

which correspond to (I.2.22a), (I.2.22b), and (I.2.22c), respectively.

Condition (e) yields

$$L^*(\bar{x}) = uN(\bar{x}) - \delta\{[z(\bar{x}) + t(\bar{x})]N(\bar{x}) + R_a\theta(\bar{x})\} = 0$$

$$\int_0^{\bar{x}} N(x)dx - \int_0^{\bar{x}} \upsilon(x)dx = 0,$$

which correspond to (I.2.22d) and (I.2.22e) respectively.

Finally, condition (f) yields

$$dL^*(x)/dx = \mu(x)\theta'(x).$$

NOTES

Discussions in section 1 are greatly influenced by Dixit (1976), Dorfman (1969) and Intriligator (1971). For rigorous proofs of the maximum principle, see, for example, Fleming and Rishel (1975) and Lee and Markus (1967).

The Theorem in section 3 is taken from Hestenes (1965) Hestenes (1966) contains the theorem and its extensions.

REFERENCES

Bellman, R., (1957), *Dynamic Programming*, Princeton University Press, New Jersey.

Dixit, A.K., (1976), *Optimization in Economic Theory*, Oxford University Press, Oxford.

Dorfman, R., (1969), "An Economic Interpretation of Optimal Control Theory, " *American Economic Review* 59, 817-831.

Fleming, W.M. and R.W. Rishel, (1975), *Deterministic and Stochastic Optimal Control*, Springer-Verlag, New York.

Hestenes, M.R., (1965), "On Variational Theory and Optimal Control Theory, " *SIAM Journal of Control* 3, 23-48.

Hestenes, M.R., (1966), *Calculus of Variations and Optimal Control Theory*, Wiley, New York.

Intriligator, M.D., (1971), *Mathematical Optimization and Economic Theory*, Prentice Hall, New Jersey.

Lee, E.B. and L. Marcus, (1967), *Foundations of Optimal Control Theory*, Wiley, New York.