POLISH ACADEMY OF SCIENCES

GEOGRAPHIA POLONICA

42

PWN—POLISH SCIENTIFIC PUBLISHERS

Address of Editorial Board

KRAKOWSKIE PRZEDMIEŚCIE 30
00-927 WARSZAWA
POLAND

Printed in Poland

POLISH ACADEMY OF SCIENCES
INSTITUTE OF GEOGRAPHY
AND SPATIAL ORGANIZATION

GEOGRAPHIA POLONICA

42

PWN — Polish Scientific Publishers • Warszawa 1979

ESSAYS ON URBAN GROWTH

AND STRUCTURE

EDITED BY

GEORGE J. PAPAGEORGIOU

ISBN 83-01-00407-x
ISSN 0016-7282

CONTENTS

PREFACE

Urban research was traditionally split into several branches, each relying upon a specific set of respective sociological, economic, or geographical concepts. Even apparently neighbouring approaches such as those derived from human ecology on the one hand and social physics on the other, remained rather loosely interconnected until very recently. The early 1970s, however, witnessed a growing interaction between individual concepts and theories; findings developed within certain approaches being increasingly used, and reinterpreted within different conceptual frameworks. The application of land rent concepts to a study of population density patterns within cities, and an inclusion of socio-ecological assumption in urban spatial interaction models are but two well-known examples of integration trends.

At the same time, the scope of urban studies has markedly changed. While the prevailing themes during the sixties included city size and functional structure, social ecology, land-use and population density, as well as urban network, and fringe expansion described with a help of innovation-diffusion analogue models, the presently viable topics pertain inter aliae to daily human interaction, inter- and intra-metropolitan migration, the structure of urban systems, environmental perception and interaction. The shift in basic concepts has been perhaps even more pronounced, with an increasing reliance on microeconomic, demographic, systemic and decision-making theories gradually replacing the classical locational paradigms. Evolution of research interest has also been reflected on two other planes. Firstly, urban topics claimed the lion's share of recent work on spatial structure, a trend only partly explained and justified by the changing proportions between urban and non-urban population. Secondly, urban research, in both its theoretical and empirical dimension has become increasingly planning and policy-relevant, a phenomenon exemplified by a move into the domain of transportation, housing, management, or health studies, and by a broad discussion of national urban policies.

The strength of present-day urban research should not overshadow its underdeveloped areas. One such area pertains to ways in which different socio-economic conditions and institutional settings are accounted for in the structure of urban models, for example, in spatial interaction and land-use models. Another problem area which has recently focussed much attention is a poor ability of urban concepts and models to account for apparently divergent trends in urban and population dynamics, as encountered between individual countries and world regions. Processes of metropolitan decline and rapid metropolitan expansion are occurring parallel to each other, yet the tools of urban analysis and basic concepts of urban structure seem to be insufficiently suited to grasp either of the two phenomena. This gap has been attributed mainly to a static nature of the models, while it should also be related to their failure in accounting for basic demographic and socio-economic factors of urban change.

It is argued that development of urban research has been both sustained and following various analytical approaches; the development characterized by some progress in integration, shifts towards new underlying concepts, a growing applicability, but at the same time, by the emergence of new blank spots and areas when confronted with real-world processes. It is hoped that the present collection of articles should contribute to the asset side of the development, by providing it with theoretical and planning-oriented inputs. This would be achieved by the enhancing of discourse between the major schools in urban studies and stimulating more work in the problem areas of urban research.

Piotr Korcell

INTRODUCTION

The idea for this special issue of *Geographia Polonica* was raised in a discussion with Piotr Korcelli during the meetings of the International Geographical Union in Moscow, summer 1976. It was then that a Working Group on Systems Analysis and Mathematical Models of the International Geographical Union was created. The object of this group is to develop international discussion on methods and, especially, on the analysis of substantive issues. Such work remains largely unnoticed because dialogue is typically restricted between specialists. It is hoped that the activities of the group will help to bring to focus such work for more geographers. Moreover, it is hoped that the activities of the group will help to overcome the considerable barriers to knowledge that distance and language impose between researchers in these areas and to provide stimulation and economy of effort. It is of particular significance that the Polish Academy of Science, through *Geographia Polonica*, has graciously provided the means for this first major international activity of the Working Group.

The distinction between pure theoretical and application-oriented research is fundamental. The former feeds the latter and the latter, in turn, provides the eventual *raison d'être* for the former. The first paper, by Wilson and Macgill, addresses itself to application. It provides a model of models by defining the working communalities amongst such diverse fields of enquiry as choice theory, population dynamics and entropy. This significantly flexible framework for application modelling is yet another example of how important the unified realm of systems analysis is for extending human capacity to deal with urban and regional systems.

In the second paper Dziewoński argues persuasively for the usefulness of certain limiting forms generated within Alonso's migration model. These forms may simplify the adequate description of such complex and subtle phenomena.

The remaining papers relate to theory itself. Two fundamental activities of science are the development along established lines of research, and the parallel, incessant striving toward the development of new, superior scientific approaches. Although it is often difficult to classify a scientific work according to this scheme, because it may pertain to both, the ordering of articles adopted here reflects this distinction; the next six articles could be broadly characterized as developments along established lines of research, while the last three articles belong to the second, more speculative activity of science.

The paper by Casetti and Thrall is concerned with the differences arising in the provision and distribution of some urban public goods in ideal competitive and optimal situations. These differences partition otherwise identical households in two subsets, each favouring a particular institutional setting. The institutional preference is generated by space itself, still another instance of Mirrlees' (1972)[1] famous result.

The next two papers deal with another important urban issue, that of urban decay and the creation of slums in market economies. It is of interest to note the extreme differences in approach. Whereas Alao's analysis is neoclassical and micro-geographic, Dendrinos' macro-geographic analysis is based on a new scientific paradigm, the theory of catastrophes, recently developed by the French mathematician René Thom (1975).[2] In spite of their deep differences, both papers complement each other by illuminating different aspects of the complex problem. Alao concludes that the

[1] J. A. Mirrlees, The optimum town, *Swedish Journal of Economics*, 74, 1972, 114–135.
[2] R. Thom, *Structural Stability and Morphogenesis*, Benjamin, New York 1975.

problem of urban decay will persist with poverty, and that partial measures are not sufficient to solve it. Surprisingly, within his model, there is no *prima facie* evidence that in all cases government intervention is called for. On the other hand, Dendrinos identifies a model of catastrophe which displays many attributes akin to those intuitively belonging to slum phenomena, including an important distinction between stable and unstable slums. Clearly, development of this last idea is relevant to a theory of planning.

The two urban decay papers share with the subsequent three as a common feature the explicit treatment of time. This is certainly one of the most significant recent developments in spatial theory. Von Rabenau examines the growth path of a small urban economy in the presence of agglomeration economies or diseconomies. In the case of agglomeration economies, he finds that a minimum urban threshold size exists below which the local economy will decay and above which it will grow. Intuitively, since centres of different order are associated with different technologies and, therefore, different agglomeration economies, they must also be associated with different threshold sizes of the type that von Rabenau discusses. His macro-geographic approach is again to be contrasted with the micro-geographic approach of Anas and Fujita. In a careful discussion, Anas points out the difference between a truly dynamic spatial analysis and the comparative statics approach, traditionally used to infer long-run equilibrium behaviour. Anas demonstrates that the two are equivalent only under some special circumstances, and he proceeds to define one such case. The nature of the difficulty is to be found in that comparative static analysis refers to a particular environment. As the city unfolds, the environment changes because of the urban change itself. Thus, although comparative statics may apply to each instance of the system's history, the method cannot be used as a substitute to urban dynamics. How different are the results in the case of a truly dynamic analysis can be seen in the paper by Fujita. This is another step in Fujita's long-range programme toward the development of an optimal planning theory of land-use. One should note the considerable variety of theoretical maps and the rich flexibility of circumstances that this promising, significant contribution can generate.

Dynamics is also the concern of Domański. He conceptualizes the main components of the immensely complex structures and processes that spatial dynamics should possess. Present-day formal theory is far from that stage of development. One may even wonder if such an all-encompassing theory will ever be realized. Even if it does not, our faith in theory should not diminish. Indeed, if we could construct consistently realistic images of the world, there would be no need for theory. The last two papers are more specific in their speculative attitude. Fisch recognizes explicitly that spatial choice behaviour depends both upon preferences and the environment. Whereas neoclassical economies has traditionally emphasized preferences, geography has traditionally recognized the environment as a determinant of choice behaviour. A feeling of the formidable difficulties to be encountered in the synthesis of the two main determinants is clear in Fisch's paper. However, efforts toward this difficult direction cannot be avoided if a more realistic theory of spatial behaviour is to be sought. Finally, Curry also ponders on individual behaviour from a neoclassical and from a novel, potential-theoretic point of view. This is the deterministic part of an ambitious work that aims at the very foundations of economic geography.

G. J. Papageorgiou
McMaster University

A SYSTEMS ANALYTICAL FRAMEWORK FOR COMPREHENSIVE URBAN AND REGIONAL MODEL BUILDING

A. G. Wilson and S. M. Macgill

School of Geography, University of Leeds, Leeds, England

1. INTRODUCTION

It is argued in this paper that it is useful to look at urban and regional systems in a relatively abstract way. The aim is to make broad but useful generalisations about urban and regional modelling. Many particular comprehensive models can then be seen as special cases which can be generated from the framework presented. The approach also enables us to note similarities between models which at first sight seem to be very different in structure.

The main objective of the paper is the development of a framework for comprehensive urban and regional modelling which represents the *minimum basis* for any such model. By adding specific assumptions in various respects, many specific models can be derived.

A particular systems theoretic approach is used in the development of the framework. First a systematic state description is built up in section 2. Then, a variety of methods are introduced in turn: accounting, in section 3, which enables us to keep track of system components; process-activity modelling in section 4, with particular reference to the identification of basic model mechanisms. A range of methods for model building are reviewed in section 5. Examples of their application in combination are presented in section 6 and some relationships between models explained in section 7. Section 8 contains a summary of the progress achieved by this argument in the tasks of developing a comprehensive model building kit. Various points which arise in relation to specific models are explored in a related paper (Macgill and Wilson, 1977).

2. REPRESENTATION OF THE SYSTEM STATE

It is important to strike a balance between the potential benefits of a very high level of abstraction, and so to generate frameworks which may be applicable to many systems, and one which takes advantages of particular features of cities and regions. In any case, the first task is to identify the main components of the system of interest. These can be taken as people, organisations (sometimes called sectors), commodities and structures. The *activities* of people and organisations can also be defined, formally, as 'components'. Such activities take place within structures, such as buildings, or use other kinds of structures, such as machines. It is also important to identify organisational and institutional structures. Activities are often *processes* which transform commodities in various ways.

An essential feature of urban and regional analysis is the spatial dimension. Thus, the main state variables will be counts of system components at locations (*stocks*) and *flows* between locations. Within each broad category of components, types will be identified. Thus, state variables will involve labels concerned with location, the broad group of components involved, and type within that broad group. For example, x_i^{pr} may be the number of people (component group p), age r (subgroup) at location i. Often, the broad group will be identified, less abstractly, by the name of the variable; x_i^{pr} defined above might then become p_i^r, reducing the number of indexes by one. Flow variables would often be distinguished by two location indices — T_{ij}, a flow from i to j — more may be needed if a list of locations which constitute a *route* had to be specified.

TABLE 1. Examples of urban system state variables

A_j	Total land area in zone j
A_j^u	Unusable land in zone j
A_j^B	Land used by the basic sector in zone j
A_j^R	Retail land use in zone j
A_j^H	Land available for housing in zone j
E_j	Total employment in zone j
E_j^B	Basic employment in zone j
E_j^{rg}	Retail employment producing good g in zone j
T_{ij}	Number of residents of zone i working in zone j
W_i^{res}	Residential attractiveness of zone i
c_{ij}	Travel cost between zone i and zone j
β^{res}	Residence-work travel parameter
P_i	Population of zone i
P_i^B	Basic population of zone i
f_j	Inverse activity rate in zone j
Z_1	Set of zones infringing residential density constraints
Z_i^H	Maximum residential density in zone i
Z_2	Set of zones not infringing residential density constraints
S_{ij}^{HRg}	Flow of good g from zone j to residents of zone i
e_i^{Hg}	Expenditure per capita on good g by residents of zone i
W_j^{HRg}	Attractiveness of zone j for residents seeking good g
β^{HRg}	Travel parameter of residence to centre for good g
S_{ij}^{WRg}	Flow of good g from zone j to workers in zone i
e_i^{Wg}	Per capita expenditure on good g by workers in zone i
W_j^{WRg}	Attractiveness of zone j for workers seeking good g
β^{WRg}	Travel parameter of work to centre for good g
Z_3^g	Set of zones infringing minimum size constraints for good g
Z_4^g	Set of zones not infringing minimum size constraints for good g
$Z^{g\,\min}$	Minimum employment, good g

It is also useful to recall that a particular label may represent a *list*: for example, r may represent $(r_1, r_2, r_3, \ldots)$ or whatever. In other words, if the component type is characterised by several indices at some level of resolution, such a list can be formally represented by a single index.

We can summarise as follows. There will be a variable for each major component group with indices representing type within that group and location (together with a time label); other variables will represent population or organisation activities at locations at particular times (or in particular time periods); and other variables will represent flows. Since the flow variables usually involve different components or activities at each 'end', then such variables may be quite complicated. For example, x_{ij}^{mnk} may be the flow of commodity k from industry m (an 'organisation') in zone i to industry n in zone j. Flow variables can be more complicated still, since a component (say a person) at one location may interact with several other activities. Thus $T_{ij_1j_2j_3}^{n}\ldots$ may specify the destination choices $j_1, j_2, j_3, \ldots$ of a type n person living in zone i for work, shop, children's school and so on. Such arrays cause problems by their high dimensionality and other methods have to be used in practice if all such information has to be retained simultaneously (as distinct, say, from defining separate arrays $T_{ij_1}^n, T_{ij_2}^n, \ldots$) — see Wilson and Pownall (1976) and Macgill (1977a). (Note that when stock variables are disaggregated, as here, they can also play the role of flow variables and the stock-flow distinction becomes blurred).

The principles on which suitable state variables can be defined should now be clear. Examples are presented in Table 1, taken from Wilson (1977a). The level of abstraction chosen will be clear from this table. If a higher level of abstraction is required, this could be achieved by defining all system components to be 'commodities' or 'organisations' and all activities as processes carried out by organisations. This sort of approach is used by Barras and Broadbent (1974) and Broadbent (1973).

3. ACCOUNTING

Flow variables involve interaction between locations or activities or both. Accounting, at its simplest, consists of tracing the destination of all system components leaving particular origin states and, conversely, the origins of all components arriving at particular final states (usually from one time to another). Detailed discussions of these principles are given elsewhere (for example, Stone, 1967, 1970; Rees and Wilson, 1977). Accounting equations within models arise because such 'origin' or 'destination' totals are often known independently of the flow variables. For example, the journey-to-work array can be related to the residential location of the workforce and the spatial distribution of jobs. These variables in turn can be related to the housing stock and the location of economic activity. Thus, two lots of accounting equations, say of the form

$$\sum_j T_{ij} = P_i, \tag{1}$$

$$\sum_i T_{ij} = E_j \tag{2}$$

can be related to population and economic activity:

$$P_i = P_i(H_i, \ldots), \tag{3}$$

$$E_j = E_j(X_j^1, X_j^2, \ldots), \tag{4}$$

(where H_i is housing stock in i, X_j^k is the amount of commodity k produced in zone j). This means that when separate submodels are produced for housing (H_i) and industrial location ($X_j^{\cdot}$) these submodels will be linked to transport flow variables through *accounting* equations of the form (1) and (2). One common but ill-understood feature of many models can usefully be noted at this point. Such location submodels may themselves be functions of flow-sensitive variables such as accessibilities. For example, P_i may be a function of 'access to employment', a_i say, and a_i is itself a function of variables which are closely related to the existing transport pattern, $\{T_{ij}\}$. At first sight, this may appear to create a circular argument, but it is more likely to be a representation of a feedback process which is a feature of many urban systems. In the example cited, the representation of the feedback is particularly simple and an appropriate research task is to seek ways of representing more complicated lag structures in models of this type. In building comprehensive models of cities and regions, interdependence between subsystems is particularly important, and it can now be seen that accounting equations play a very important role in this respect.

Some models are so dependent on assumptions which are closely connected to accounting equations that they can be called *account-based models.* These usually involve assumptions about average rates. A discussion of this role of accounting will be postponed until section 5, however. The next step in the argument is to review the kinds of accounts which are appropriate to cities and regions.

Accounts can arise in the following ways:

(i) 'ageing' of system components: this involves the definition of states by age (and possibly other characteristics) and change occurring due to the basic processes of birth, death and migration,

(ii) inter-organisational transactions — or, organisational activities which involve interaction. The interactions can be intersectoral or spatial or both,

(iii) the transformation of system components by organisational activity (such as productive processes),

(iv) 'balance' accounts: if a commodity is conserved, then flows into and out of a zone must balance. A more complicated constraint of this type arises in balancing flows at the nodes of a network carrying interactions between zones,

(v) these different kinds of accounts can be combined in different ways into hybrids or composites.

Accounts of the first type include arrays of the form $\{K^{pq}\}$ where p and q are demographic states (and recall that, here and below, such labels may themselves be 'lists'); the second is illustrated by economic flows $\{Z^{mn}\}$ from sector m to sector n, or the shopping model array $\{S_{ij}\}$ which is an interaction between people at their residences and shops; the third involves the specification of the process of transformation and the flow of a commodity through it. Thus the array $\{X^{mnk}\}$ specifies the inputs of k into n from m (and the outputs, conversely, of m to n) which, for a set of commodities (or goods), k, implies a description of the input-output structure of some production processes (cf. Cripps, Macgill and Wilson, 1974; Macgill, 1975a). Composite sets of accounts can easily be formed in many of these cases, for example by adding the spatial interaction representations to form $\{K_{ij}^{pn}\}$, $\{Z_{ij}^{mn}\}$ or $\{x_{ij}^{mnk}\}$. These arrays involve the intermeshing of two different kinds of accounts into more complicated composite arrays. (And, indeed, $\{x^{mnk}\}$ is already 'composite': it represents inter-industry flows and specifies a set of productive processes).

Such accounting arrays ensure consistency and that system components are not 'lost'. The different kinds of accounts reflect the different kinds of processes involved — a point which will be taken further in the next section.

The equations generated by different accounts have different characteristics and play different roles in models. Because of the different natures of the underlying processes of change, demographic rates are best constructed by dividing account elements by origin-state totals (row sums), because the existing population is the driving force; in the economic transactions case, destination-state totals (column sums) are usually used because the level of output of the receiving sector is the main driving force. In the case of spatial interaction accounts, rates are not usually of interest — rather the row and column totals. When rates are used in model building, these distinctions lead to fundamentally different models. Change in account-based models is likely to be based on changes in row-sum rates for people, column-sum rates for organisations, and row and column absolute totals for spatial interaction.

Finally, it should be recalled that occasionally balancing constraint equations are important: for example, the possibility of materials balance at a location for certain non-transformed commodities; or constraints arising from flows sharing links of a transport network and balancing at each node.

4. PROCESSES AND MECHANISMS

Most actual processes take place at a micro scale within organisations — in households, factories or whatever. Some, particularly certain governmental controlling or planning processes, relate directly to meso or macro scales. Micro phenomena, however, can be aggregated to the meso scale, which is our main level of interest in this paper. A principal objective of meso-scale modelling therefore is to identify the phenomena (through the kinds of system state and accounting representations discussed in sections 2 and 3) at that scale and to connect to the processes driving them at the micro scale. They should also be effectively connected to processes which are usually represented at the macro scale, such as demographic change in the population as a whole. The identification and articulation of processes and mechanisms in this manner is sometimes taken to be the *theoretical basis* of the model — as distinct from 'methods and techniques' associated with modelling. In this section, therefore, we emphasise the crucial importance of this step (though for our own taste, 'theoretical basis' has wider connotations, to include the methods used).

The aggregation problem is a very difficult one. Micro-scale behaviour takes place in a meso and macro-scale environment which is largely determined by the aggregate of micro-scale behaviour. The difficulties of modelling the two feedbacks explicitly are clear (Wilson, 1975). The essential modelling task in the first instance, therefore, is to be as explicit as possible at the micro scale about the processes involved and the representation of the processes at higher scales. Sometimes, a process will be represented directly, say by having variables determined endogenously through some optimisation process. In other cases, it is a matter of identifying assumptions which have to be made about exogenous variables in the model and the underlying processes associated with them. Modelling is always likely to be additionally complicated because several processes can operate simultaneously in one particular subsystem — for example household decision processes on the one hand, and government planning on the other in relation to residential location.

These points will be taken up in the next section on methods, where we discuss particular methods for the representation of particular processes — especially optimisation processes involving (possibly nested) mathematical programmes — together with the explicit investigation of exogenous variables in particular submodels (since the assumptions about such variables *drive* the models). Also, the careful consideration of exogenous variables in submodels facilitates the investigation of interdependence between submodels, since exogenous variables are often common between submodels (which would lead to some correlations between the behaviour of such submodels — and dangers of the 'ecological fallacy'), or some variables which are exogenous to one submodel are endogenous to others.

5. METHODS

The range of methods available for urban and regional modelling is described and illustrated elsewhere (Wilson, 1977b).
The headings used are:

(i) entropy maximising,
(ii) account-based average rate methods,
(iii) optimisation and mathematical programming,
(iv) network analysis (graph theory),
(v) dynamical systems theory.

We should also, of course, recognise that many ad hoc methods, which are not so neatly classified, have been and are used. The reasons for choosing particular methods in particular cases were related to the 'Weaver-type' of the system — whether the system is one of disorganised complexity (or can be taken as approximately that) or of organised complexity. With occasional exceptions, the first three methods are applicable to systems of the first type. Entropy maximising is distinguished from other mathematical programming methods as being very closely related to disorganised complexity, although most programming methods offer a way of solving the micro and meso aggregation problem when the underlying processes can be represented as maximisation or minimisation subject to constraints. The term 'optimisation' is used in this context in its mathematical sense, though sometimes it applies in a normative sense also. Account-based average rate methods also rely on disorganised complexity assumptions though, as argued in section 3 above, accounting equations are often fundamental to a wider range of models and methods.

Organised complexity, by definition, involves more interdependence between subsystems or system components. No general methods are available for solving all the problems which turn up, though some of the methods of network analysis and dynamical systems theory help. Graph theory in particular emphasises the identification of the relationship between system components.

The ranges of application of the individual methods are mostly well known, as are particular examples associated with each. The entropy maximising method is particularly appropriate for estimating the numbers of components in particular states in systems with large numbers of components in all. It has been applied to a wide variety of systems of which the journey-to-work, as an interaction model, and the flow to, and distribution of, shopping activity are perhaps the best known examples of interaction and location models, respectively (Wilson, 1970). The method has also been widely applied to help fill information gaps (Macgill, 1975b).

Account-based models are most useful for systems where the underlying processes can be adequately described in terms of rates (and hence this is a different kind of disorganised-complexity argument). The best known examples are population models, where birth, death and migration processes are described directly in terms of rates and economic models, where the rates in an input-output model form a representation of the technological structure of the economy.

Models involving optimisation also arise directly from a consideration of the processes involved. Such models may involve only a small number of variables and be essentially simple, but more often in an urban and regional context they involve large numbers of system components. The mathematical programming representation of the processes involved then usually rely on disorganised-complexity arguments and provide the basis for solving the aggregation problem in this case. Methods range from linear programming (e.g. Herbert and Stevens, 1960) through nonlinear models (cf. Wilson, 1976a for a review) to probability methods such as random utility theory (Williams, 1977).

Much of the difficulty of urban modelling in particular arises from the underlying transport network structure. Network theory plays a direct part in the resolution of such difficulties (see, for example, Evans, 1976, for a description of a joint distribution-assignment model). The various methods developed for assigning flows to networks, and taking account of network congestion, provide ways of taking into account some of the basic nonlinear effects in urban modelling. Network theory in its more abstract sense, perhaps then better known as graph theory, also provides a basis for a formal examination of interdependence within systems (for example, using the digraph methods as in Roberts, 1976). This also turns out to have close connections to setting up differential equation models and elementary dynamics (Keys, 1977; Macgill, 1977b). Such models can sometimes be interpreted as average rate models. More complex aspects of structure can be sought using the methods of Atkin (1974) which also perhaps, at least broadly, fall under the heading of graph theory.

Finally, dynamical systems theory offers a method for studying types of change which cannot be represented in elementary dynamic models: these involve rapid or discontinuous change in the neighbourhood of certain critical regions of the parameter space of the models. These methods are most powerful in urban and regional modelling when applied in combination with other methods, as we shall see below.

6. METHODS IN COMBINATION

The preceding paragraphs above provide a brief survey of available methods and a prelude to the two main points to be made here: first, the methods can often be applied in combination; and secondly, there are strong relationships, even equivalences, between models which result from combinations of methods being applied in different ways which are not immediately obvious at first sight. We discuss each point in turn below, beginning with some examples of the methods applied in combination.

(i) The first example can be taken either as methods applied in combination or as a comment on a relationship between two apparently very different methods. It has been established that many mathematical, and especially linear, programming models can be seen as special cases of entropy maximising models (Evans, 1977; Wilson and Senior, 1974): the programming models arise from entropy maximising models as certain parameters in the latter tend to infinity. The entropy maximising models can then be seen as suboptimum (but still 'most probable' in the suboptimum situation) versions of the programming models. It therefore becomes possible for an analyst who builds a programming model based on some optimising process to transform it to an entropy maximising model which represents a suboptimal version of the same process (say involving an imperfect market instead of a perfect one), or vice versa.

(ii) Entropy maximising models are, of course, derived from nonlinear mathematical programmes. This means that the entire theory of programming is available for application if required. It has been useful for interpretation and for calibration purposes, for example, to formulate the *dual* of entropy maximising models (Wilson and Senior, 1974; Champernowne, Williams and Coelho, 1976).

(iii) There are a number of applications of the entropy concept within account-based modelling. One arises because information relating to subsets of the account elements is often missing. Such gaps can often be filled using entropy maximising methods. Applications range from the estimation of disaggregated accounts from more aggregate ones, as with $\{T_{ij}^{kn}\}$ from $\{T_{ij}^{*}\}$ and $P(k/n)$ (where T_{ij}^{k} is work trips from i to j by mode k by persons of type n and $P(k/n)$ is the aggregate probability of using mode k for given n), as in Southworth (1977), to more complicated examples of balancing factors in spatial population analysis (Rees and Wilson, 1977). Balancing factor methods are also used to adjust economic accounts — the well know 'RAS' method being one example (Stone, 1967).

A second application uses the Kullback (1959) form of entropy function which makes it possible to develop input-output models by adjustment of past accounts subject of the appropriate constraints (Macgill, 1978).

Finally, entropy maximising submodels sometimes play a direct role in account-based models, especially in relation to spatial interaction components (Wilson, 1970; Cripps, Macgill and Wilson, 1974; Macgill, 1975a).

(iv) Sometimes, an investigation of the processes driving some system-of-interest may reveal two connected optimisation processes. For example, the pattern of development and usage of shopping centres is determined by producers (governmental or commercial) trying to distribute shops to centres to maximise their own profits or some collective goals while users are each trying to optimise their own utility. Such a model involves nested mathematical programmes, one 'inside' the other. It turns out that this can be converted into a single overall programme with the appropriate properties (Coelho and Wilson, 1976, 1977; Coelho, Williams and Wilson, 1978). This principle of nested mathematical programmes may be extended to cases with more tiers in the hierarchy.

(v) Network theory is nearly always used in combination with other methods. Flows generated from interaction models (and associated location models) are loaded onto networks and costs are modified in the light of network congestion. This is a standard procedure — so-called capacity restraint assignment — in transport modelling but has not been applied as systematically as it ought to have been in other branches of urban modelling. We have already noted that graph theory is applied in conjunction with methods involving differential equations. And we note that sufficient progress has now been made with q-analytical investigations of urban structure for workers in that field to propose some integration with, for example, entropy maximising modelling (Atkin, 1977).

(vi) A number of topics are now mentioned which involve combining dynamical systems theory with various other methods. The first involves noting that accounting equations lead directly to difference or differential equations describing change. This means that the methods involving such equations can be applied to some account-based models, such as the investigation of whether solutions are stable in the long run. (See Cordey-Hayes, 1972, for a presentation of such methods which relate to kinetic theory and Rees and Wilson, 1977, Appendix 6, for applications in spatial population analysis.)

(vii) There is a very close relationship between mathematical programming and dynamical systems theory. If time is treated as 'just another state variable' then the methods of mathematical programming can often be applied directly to the total set of variables. Otherwise, with mathematical programming models, the constraint equations define a feasible region in state space and the optimising mechanism a point in that space for each time value. Behaviour is therefore represented by a curve in state space. The methods of catastrophe theory, for example, may then be relevant to the investigation of discrete system changes within certain critical regions of state space. Since mathematical programming models are usually based on strong equilibrium assumptions and have 'comparative static dynamics', this comment shows that such models can in principle have quite complicated dynamic behaviour (see Wilson, 1976b, Poston and Wilson, 1977).

(viii) As we noted in another context ((iii) above), account-based models are sometimes inapplicable because of the large number of gaps in the accounting elements, and we can add, in this context, sometimes because of the sheer size of the arrays. When the problem arises in a dynamic model, a common solution is to reduce the size of the array by using a pool. For example, instead of modelling N^2 terms $\{K^{ij}\}$, it may be appropriate to model N terms, K^{ii} (the stayers) and $2N$ terms $F^i = \sum_{j \neq i} K^{ij}$, $G^j = \sum_{i \neq j} K^{ij}$ (the movers) as flows from i into a pool (F^i) and flows into j from the pool (G^j). This kind of more aggregate accounting underpins much of the so-called 'industrial dynamics' style of modelling (Forrester, 1968).

(ix) Dynamical systems theory can often be applied, in practice, to only a small number of variables. Otherwise the mathematical problems associated with systems of organised complexity become dominant and intractable. Such methods are therefore used most powerfully in conjunction with others which predict the bulk of the variables. For example, an entropy maximisnig or programming model may be used to predict a large set of state variables subject to a small number of exogenous variables being given. It is then often appropriate to apply dynamical systems methods to that small number of exogenous variables. (For an example, involving an explicit model of central place theory, see Wilson, 1977a.)

7. SOME RELATIONSHIPS BETWEEN MODELS

In sections 5 and 6 we have presented methods which can be used, individually and in combination, for building system models. We can now generate an interesting related question by turning the issue around: do our explorations to date throw any light on the relationships between different existing models (usually at the meso scale) of the same system? We offer a number of preliminary comments which may stimulate more research.

(i) First, of course, models may differ because they have different underlying theoretical assumptions about the processes and mechanisms involved. It is always of interest to try to isolate such differences so that, in any particular case, a clear decision can be taken as to which underlying assumptions are preferable. The converse effect is also interesting: often, the same model (at the meso scale) can be derived from many different assumptions. A very obvious example is the gravity model of spatial interaction. We list some of the alternative derivations in Table 2. This effect may be called (following Thrift, 1977) 'model equifinality' with respect to assumptions. Even more frequently, of course, different assumptions generate similar models, which may be called 'model homeofinality'.

Subsequent comments mostly relate to alternative *methods*, though it is useful

TABLE 2. Alternative derivations of the gravity model of spatial interaction

Label	Characteristics	Selected references
Newtonian	By analogy with gravitational force between two masses	
Probabilistic	Based wholly on probabilistic assumptions as to human behaviour	Harris (1964)
Statistical mechanical (maximum entropy)	Most probable of all possible distributions satisfying origin, destination and interaction cost constraints	Wilson (1967) Wilson (1970)
Information theoretic (maximum entropy)	Maximally unbiased with respect to origin, destination and cost information	Wilson (1970) Batty and March (1976)
Micro-economic behavioural (constant utility)	Consistent with economic theory of consumer behaviour, with travel as a consumer good	Niedercorn and Bechdolt (1969) Golob and Beckmann (1970)
General share model	Sequential subdivisions from origin totals to destinations, models, routes	Manheim (1973)
A 'generalised' distribution model (macro--behavioural)	Assumptions on aggregate human behaviour to establish sources of variation in data	Cesario (1973)
Spatial discounting	Travellers discount potential interaction opportunities in terms of perceived distances	Smith (1975) Smith (1976)
Random utility (micro--economic behavioural)	Utility and/or cost alternatives regarded as random variables	Williams (1977)
Cost efficiency (macro--behavioural)	Trip patterns with lower total cost are always more likely to occur	Smith (1977)
Entropy constrained transportation problem	Classic linear programming transportation problem with entropy function as an additional constraint	Coelho and Wilson (1977) Erlander (1977)

to bear in mind the concepts of model equifinality and homeofinality, this time with respect to methods rather than assumptions (and there will obviously be cases where both *methods* and assumptions differ!).

(ii) As a first point on similarities between apparently dissimilar models, we re-iterate the first point of 'methods in combination', but in a slightly different way: to each mathematical programming model, there is an associated suboptimum entropy maximising model, and vice versa.

(iii) There are two well-known distinctive forms of account-based model. The first is exemplified by the population model

$$\boldsymbol{w}(t+T) = \boldsymbol{G}\boldsymbol{w}(t), \tag{5}$$

where $\boldsymbol{w}(t)$ is a vector of population at time t and $\boldsymbol{G}$ a 'change' operator — which is actually a transposed matrix of rates. The second is illustrated by the well-known input-output model

$$\boldsymbol{x} = (\boldsymbol{I}-\boldsymbol{a})^{-1}\boldsymbol{y}, \tag{6}$$

where $\boldsymbol{x}$ is a vector of total products, $\boldsymbol{y}$ a vector of final demands and $\boldsymbol{a}$ a vector of column based rates; $\boldsymbol{I}$ is the unit matrix. Since the model involves the inversion of an often large matrix, the interdependencies of such a model are much more complicated than its population equivalent. This complexity arises partly because one sector, final demand, has been picked out as the driving force of the model (as distinct from the whole initial population, $w(t)$); and partly because the process modelled involves a sequence of increasingly higher order effects as spelled out in equation (7) below in another context.

We should note that the population model can be cast into (6)-form, by picking out 'births' to describe it (Stone, 1970). Also, the two kinds of assumptions can be combined (that is a mixture of row-and-column-based rates are used — Macgill, 1977b) and digraph methods have to be used to handle the resulting complexity.

(iv) Equation (6) can be expanded to give

$$\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{a}\boldsymbol{y} + \boldsymbol{a}^2\boldsymbol{y} + \boldsymbol{a}^3\boldsymbol{y} + \ldots \tag{7}$$

This is equivalent to an iterative scheme of the form

$$\begin{aligned} \boldsymbol{x}^{(0)} &= \boldsymbol{y}, \\ \boldsymbol{x}^{(n+1)} &= \boldsymbol{y} + \boldsymbol{a}\boldsymbol{x}^{(n)}, \qquad n \geqslant 0. \end{aligned} \tag{8}$$

Some urban models are presented as sets of equations which can only be solved by an iterative scheme of this type. It is then easy to see that such a model can also be cast in the form of equation (6). Such an equivalence may be of considerable interest for two reasons: first, one form may be computationally more convenient than the other in a particular case; secondly, when a model is cast in the new form, immediate theoretical improvements may suggest themselves. An example of such a model is the Lowry (1964) model (Macgill, 1977c). In particular, when the model, which is usually presented in iterative form, is presented in equation (6) form as an input-output model, improvements can be made — for example to incorporate a full set of inter-industry relations*.

(v) It is interesting to examine the preceding two points on aspects of account-based modelling in the context of a third kind of such modelling, mentioned briefly above and based on mathematical programming. The argument is conducted in terms of two examples. First, note that the Lowry model can be formulated as a mathematical programming model (Coelho and Williams, 1977; Williams and Coelho, 1977). In appropriate circumstances, the three formulations of the model (iterative-algebraic, input-output and mathematical programming) are equivalent, and in other cases only similar. Differences arise in relation to the way land-use capacity and service minimum-size constraints are handled in the three cases.

The second example relates to the family of rectangular input-output models. As is usual with input-output multiplier models; these are based on rate coefficients derived from base period inter-industry transactions tables. In their original versions they took the form of matrix inverse models, triggered in the usual way by final demand elements, but rather different from the classic input-output models (cf. equation (6)) due to the need to accommodate the assumption that sectors may

* Note that the better known matrix version of the Lowry model, due to Garin, takes a matrix inverse form but is not an input-output model in the usual sense.

produce any number of products (see Gigantes, 1970 and Armstrong, 1975 for descriptions of these models). The need for an exactly determined system of equations for these matrix inverse models restricted the variety of forms they could take, but more recent work has shown that the rate coefficients central to them can in fact be used to generate constraint equations (that is, under-determined rather than exactly determined systems of equations) within mathematical programming frameworks, in particular entropy maximising frameworks using the Kullback entropy function (Macgill, 1977c, 1977d). These lead to an extended family of rectangular input-output models sharing some obvious similarities with the original models, but exact equivalence holding only when the rate coefficient equations are identical in the matrix and mathematical programming representations.

(vi) The next point relates to alternative formulations of mathematical programming models. It is sometimes necessary to seek the mathematically most convenient form. For example, with the shopping model embedding problem mentioned earlier, if spatial interaction model equations are taken as constraints, then since these are numerous and nonlinear, the problem appears intractable. However, there is an alternative and exactly equivalent formulation which can be solved by standard methods (see previous references on this point). Thus, mathematical programming models which appear to be very different may be equivalent.

(vii) Another general point about different-looking programming models of a particular system is, in effect, a more general case of both (i) and (vi) above: by the use of appropriate Lagrangian multipliers, constraints can be moved in and out of the objective function and formulations can be made to look more different than they are. Many such alternative formulations can be made equivalent provided the multipliers take particular values to ensure this. If this last condition is not met, there will only be a broad equivalence of form. For example, if a term $\lambda f(\boldsymbol{x})$ which appears in an objective function is taken out and a constraint $f(\boldsymbol{x}) = y$ added, then the two formulations will be equivalent if λ is constrained to have the value corresponding to y, but not otherwise.

(viii) An important result has recently been derived (Evans, 1976) showing in effect that certain iterative capacity restraint assignment procedures are equivalent to a particular mathematical programming formulation which encompasses both interaction and network variables. This is a result which has important practical consequences (Boyce, 1977; Florian and Nguyen, 1976).

(ix) Certain signed and weighted digraph methods of analysis, although looking completely different, are equivalent to models formulated with difference equations. This is so whenever the signs or weights on the digraph links represent causal relationships between the nodes which they join, specifically when they represent the change in the forward node that will result from a unit change in the backward node (or something equivalent to this) (see Roberts and Brown, 1975). The digraph methods then provide an accessible way of deriving difference equations for the system of interest by methodically picking up the separate interdependencies between each pair of system components.

(x) Interesting formal equivalences are harder to come by in relation to aspects of dynamical systems theory. This is at least in part because of inexperience with the techniques. One example may suffice to show that alertness in this respect may be worthwhile, however. Consider the shopping model example again and the problem of predicting the distribution of shopping centre sizes $W = (W_j)$. Let $P(W)$ be the profit associated with a particular arrangement. Then a programming model would take the form $\text{Max}\, P(W)$ subject to a set of constraints. A differential equation model may take the form

$$\dot{W} = \varepsilon P(W)$$

for some constant, ε. In particular circumstances the equilibrium points of such a dynamical system will be the same as those generated by the mathematical programme (Harris and Wilson, 1978). So two different formulations are equivalent in an important respect, although the differential equation form is a 'stronger' model because it allows transient states to be analysed after a disturbance.

8. TOWARDS A COMPREHENSIVE MODEL BUILDING KIT

We have argued, at least by implication, that it is not appropriate to recommend a single comprehensive model as the best available. A rich variety of models can be built with the methods sketched in this paper and we have also seen that subsets of this variety often have more in common than may appear at first sight. What we have sought to identify, in effect, are the rules and procedures for building comprehensive urban and regional models and this constitutes the basis of a model building kit. The emphasis is on making the rules and procedures as explicit as possible.

The argument can be summarised as follows:

(i) Section 2 above: identify the system of interest, the major subsystems and the state variables. This involves defining appropriate sectoral, spatial and temporal levels of resolution for the task in hand. A particular representation of variables was recommended. This is a powerful one which allows many theoretical questions to be tackled without many of the usual restraints (such as the uniform plain assumptions of traditional central place theory).

(ii) Section 3: establish the appropriate accounting foundations. This will provide many model equations and will ensure the internal consistency of the model.

(iii) Section 4: identify the processes and mechanisms of the various parts of the model and make explicit theoretical hypotheses about them to build into the model. This will often involve grappling with a major aggregation problem when the hypotheses relate to micro-scale behaviour while the main model state variables are defined at a meso-scale.

(iv) Section 5: choose appropriate methods for model building.

(v) Section 6: note the richness of choice of methods available and the variety of ways in which they can be used in combination.

(vi) Section 7: different models may have a greater degree of similarity than may be evident at first sight.

This approach is in contrast to those of many authors who seek to identify a unique methodology for building a comprehensive model, but who are often less explicit about the model design decisions implied by the points (i)–(vi) above. The strategy generated by the argument of this paper involves major problems, however, which are faced by anyone attempting to build a comprehensive model. There is a major decision of style to be faced at the outset: whether or not to attempt to link the 'best available' partial, or subsystem, models. The arguments in favour of an affirmative answer are clear, but there is one big argument against — the resources involved in building each of the best partial models are huge, and so those involved in building comprehensive models in this style are probably unattainable at the present time. This suggests a research strategy which involves building simpler comprehensive models, but where the main purpose is to seek to represent and reproduce systemic effects which would otherwise be lost. (There is always the danger, of course, that some of the most interesting effects will only be seen in a more detailed comprehensive model, but that is a question for research over the longer term). A complementary avenue of research involves building new partial models, possibly in the hope of seeking methods which are both theoretically and empirically effective, but which also need less data for testing and projection purposes. It is essentially the latter strategy which has been followed in transport modelling in recent years and there is scope in other sectors. (For example, the development of spatial input-output models will only be possible if the estimation of such models can be achieved with relatively modest surveys.)

REFERENCES

Armstrong, A. G., 1975, Technology Assumptions in the Construction of U. K. Input-output Tables, in R. I. G. Allen and W. F. Gossling (eds.), *Estimating and Projecting Input-output Coefficients*, Input-Output Publishing Company, London.

Atkin, R. H., 1974, *Mathematical Structure in Human Affairs*, Heinemann, London.

Atkin, R. H., 1977, *Q-analysis: Theory and Practice*, Research Report X, Department of Mathematics, University of Essex.

Barras, R. and T. A. Broadbent, 1974, The development of an Activity-commodity Representation of Urban Systems as a Potential Framework for Evaluation, in E. L. Cripps (ed.), *Space-time Concepts in Urban and Regional Models*, Pion, London.

Batty, M. and L. March, 1976, The Method of Residues in Urban Modelling, *Environment and Planning*, A, 8, 189–214.

Boyce, D. E., 1977, *Equilibrium Solutions to Combined Urban Residential Location, Modal Choice, and Trip Assignment Models*, Discussion Paper 98, Regional Science Research Institute, Philadelphia.

Broadbent, T. A., 1973, Activity Analysis of Spatial-allocation Models, *Environment and Planning*, A, 5, 673–691.

Cesario, F. J., 1973, A Generalised Trip Distribution Model, *Journal of Regional Science*, 13, 233–247.

Champernowne, A. F., H. C. W. L. Williams and J. D. Coelho, 1976, Some Comments on Urban Travel Demand Analysis, Model Calibration and the Economic Evaluation of Alternative Plans, *Journal of Transport Economics and Policy*, 10, 267–285.

Coelho, J. D. and H. C. W. L. Williams, 1977, *On the Analysis and Evaluation of Land Use-transportation Plans II: A Mathematical Programming Approach to Activity Allocation and Evaluation*, Working Paper 202, School of Geography, University of Leeds, to be published, *Papers, Regional Science Association*.

Coelho, J. D., H. C. W. L. Williams and A. G.. Wilson, 1978, Entropy Maximising Submodels within Overall Mathematical Programming Frameworks: A Correction, *Geographical Analysis*, 10, 195–201.

Coelho, J. D. and A. G. Wilson, 1976, The Optimum Location and Size of Shopping Centres, *Regional Studies*, 10, 413–421.

Coelho, J. D. and A. G. Wilson, 1977, Some Equivalence Theorems to Integrate Entropy Maximising Submodels within Overall Mathematical Programming Frameworks, *Geographical Analysis*, 9, 160–173.

Cordey-Hayes, M., 1972, Dynamic Frameworks for Spatial Models, Socio-Economic Planning Sciences, 6, 365–385.

Cripps, E. L., S. M. Macgill and A. G. Wilson, 1974, Energy and Materials Flows in the Urban Space Economy, *Transportation Research*, 8, 293–305.

Erlander, S., 1977, *Entropy in Linear Programming — an Approach to Planning*, Linkoping Institute of Technology, Report LiTH–MAT–R–77–3, Linkoping, Sweden.

Evans, S. P., 1973, A Relationship Between the Gravity Model for Trip Distribution and the Transportation Problem of Linear Programming, *Transportation Research*, 7, 39–61.

Evans, S. P., 1976, Derivation and Analysis of Some Models for Combining Trip Distribution and Assignment, *Transportation Research*, 10, 37–57.

Florian, M. and S. Nguyen, 1976, Recent Experience with Equilibrium Methods for the Study of a Congested Urban Area, in M. Florian (ed.), *Traffic Equilibrium Models*, Springer Verlag Heidelberg.

Forrester, J. W., 1968, *Urban Dynamics*, M. I. T. Press, Cambridge, Mass.

Gigantes, T., 1970, The Representation of Technology in Input-output Systems, in A. Brody and A. P. Carter (eds.), *Contributions to Input-output Analysis*, North-Holland Publishing Co., Amsterdam.

Golob, T. F. and M. J. Beckmann, 1970, A Utility Model for Traffic Forecasting, *Transportation Science*, 5, 79–89.

Harris, B., 1964, A Note on the Probability of Interaction at a Distance, *Journal of Regional Science*, 5, 31–35.

Harris, B. and A. G. Wilson, 1978, *Notes on Shopping Model Dynamics*, mimeo.

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

Keys, P., 1977, *A Relationship Between Diagraphs and Differential Equations with an Application to an Urban System*, Working Paper 194, School of Geography, University of Leeds.

Kullback, S., 1959, *Information Theory and Statistics*, John Wiley, New York.

Lowry, I. S., 1964, *A Model of Metropolis*, RM–4035–RC, Rand Corporation, Santa Monica.

Macgill, S. M., 1975a, *Urban Energy Flows within a General Systems Framework: Towards the Design of New Model Structures*, Working Paper 123, School of Geography, University of Leeds.

Macgill, S. M., 1975b, *Balancing Factor Methods in Urban and Regional Analysis*, Working Paper 124, School of Geography, University of Leeds.

Macgill, S. M., 1977a, *Multiple Classifications in Socio-economic Systems*, Working Paper 212, School of Geography, University of Leeds.

Macgill, S. M., 1977b, Simple Hybrid Input-output Models: A Graphical Approach, to be published, *Environment and Planning*, A.

Macgill, S. M., 1977c, The Lowry Model as an Input-output Model and its Extension to Incorporate Full Inter-activity Relations, *Regional Studies* 11, 337–354.

Macgill, S. M., 1977d, *Alternative Rectangular Input-output Models*, Working Paper 182, School of Geography, University of Leeds.

Macgill, S. M., 1978, Rectangular Input-output Tables, Multiplier Analysis and Entropy Maximising Principles: A New Methodology, *Regional Science and Urban Economics* 8, 355–370.

Macgill, S. M. and A. G. Wilson, 1978, Equivalences and Similarities between Some Alternative Urban and Regional Models, *Sistemi Urbani* 1.

Manheim, M. L., 1973, Practical Implications of Some Fundamental Properties of Travel Demand Models, *Highway Research Record*, 422, 21–38.

Neidercorn, J. H. and B.V. Bechdolt, 1969, An Economic Derivation of the Gravity Law of Spatial Interaction, *Journal of Regional Science*, 9, 273–282.

Poston, T. and A. G. Wilson, 1977, Facility Size Versus Distance Travelled: Urban Services and the Fold Catastrophe, *Environment and Planning*, A, 9, 681–686.

Rees, P. H. and A. G. Wilson, 1977, *Spatial Population Analysis*, Edward Arnold, London.

Roberts, F. S., 1976, *Discrete Mathematical Models*, Prentice-Hall, Englewood Cliffs, N. J.

Roberts, F. S. and T. A. Brown, 1973, Signed Digraphs and the Energy Crisis, *American Mathematical Monthly*, 82, 577–594.

Smith, T. E., 1975, An Axiomatic Theory of Spatial Discounting Behaviour, *Papers, Regional Science Association*, 35, 31–44.

Smith, T. E., 1976, A Spatial Discounting Theory of Interaction Preferences, *Environment and Planning*, A, 8, 897–915.

Smith, T. E., 1977, *A Cost Efficiency Principle of Spatial Interaction*, paper presented at the 17th European Congress, Regional Science Association, Krakow.

Southworth, F., 1977, *The Calibration of Disaggregated Trip Distribution and Modal Split Models: Some Comparisons of Macro and Micro-analytic Approaches*, Ph. D. thesis, University of Leeds.

Stone, R., 1967, *Mathematics in the Social Sciences*, Chapman and Hall, London.

Stone, R., 1970, *Mathematical Models of the Economy*, Chapman and Hall, London.

Thrift, N. J., 1977 *Private Communication.*

Williams, H. C. W. L., 1977, On the Formation of Travel Demand Models and Economic Measures of User Benefit, *Environment and Planning*, A, 9, 285–344.

Williams, H. C. W. L., and J. D. Coelho, 1977, *On the Analysis and Evaluaton of Land Use-transportation Plans I: Activity Location Models Based on Ratinnal Choice Behaviour*, mimeo, School of Geography, University of Leeds.

Wilson, A. G., 1967, A Statistical Theory of Spatial Distribution Models, *Transportation Research*, 1, 253–269.

Wilson, A. G., 1970, *Entropy in Urban and Regional Modelling*, Pion, London.

Wilson, A. G., 1975, Learning and Control Mechanisms for Urban Modelling, in E.L. Cripps (ed.), *Regional Science: New Concepts and Old Problems*, Pion, London.

Wilson, A. G., 1976a, *Nonlinear and Dynamic Models in Geography: Towards a Research Agenda*, Working Paper 160, School of Geography, University of Leeds.

Wilson, A. G., 1976b, Catastrophe Theory and Urban Modelling: An Application to Modal Choice, *Environment and Planning*, A, 8, 351–356.

Wilson, A. G., 1977a, *Spatial Interaction and Settlement Structure: Towards an Explicit Central Place Theory*, Working Paper 200, School of Geography, University of Leeds.

Wilson, A. G., 1977b, Recent Developments in Urban and Regional Modelling: Towards an Articulation of Systems Theoretical Foundations, *Proceedings*, vol. 1, Giornate di Lavoro, A. I. R. O. Parma, Italy.

Wilson, A. G. and C. E. Pownall, 1976, A New Representation of the Urban System for Modelling and for the Study of Micro-level Interdependence, *Area*, 8, 246–254.

Wilson, A. G. and M. L. Senior, 1974, Some Relationships Between Entropy Maximising Models, Mathematical Programming Models and their Duals, *Journal of Regional Science*, 14, 207–215.

GEOGRAPHIA POLONICA 42, 1979

RÉFLEXIONS SUR LA THÉORIE DES MIGRATIONS DE WILLIAM ALONSO

KAZIMIERZ DZIEWOŃSKI

Institut de Géographie et de l'Aménagement du Territoire, Académie Polonaise des Sciences, Varsovie, Pologne

William Alonso, économiste et géographe, ancien professeur de l'Université de Californie à Berkeley, actuellement professeur et directeur du Centre de la Recherche Démographique de l'Université de Harvard a présenté récemment une proposition, fort intéressante, de la théorie générale des mouvements migratoires*. C'est une théorie entièrement formalisée, possèdant à la fois une construction remarquable dans sa logique et dans sa simplicité. Elle englobe, sous une forme homogène et unie, l'ensemble des modèles et des théories partiels élaborés jusqu'à présent.

Cette théorie s'enferme dans cinq équations décrivant les mouvememts, qui dans le contexte des mouvements migratoires signifient successivement: le nombre de personnes émigrants, le nombre de personnes immigrants, les flux de migrants et enfin les valuers de deux coefficients qui décrivent les caractéristiques — les relations, l'un entre un territoire (une localité) appartenant à la classe *i* et envoyant les migrants aux autres territoires (localités) et l'autre — entre un territoire (une localité) appartenant à la classe *j* et recevant les migrants de tous les autres territoires (localités). Ce sont des équations suivantes:

(1) $\sum_{i} M_{ij} = M_{ix} = v_i D_i^{\alpha}$ émission des migrants

(2) $\sum_{j} M_{ij} = M_{xj} = w_j C_j^{\beta}$ attraction des migrants

(3) $M_{ij} = v_i w_j t_{ij} D_i^{\alpha-1} C_j^{\beta-1}$ flux de migrants

(4) $D_i = \sum w_j t_{ij} C_j^{\xi-1}$ variable d'un système d'émigration

(5) $C_j = \sum v_i t_{ij} D_i^{\alpha-1}$ variable d'un système d'immigration

où:

M_{ix} — nombre total de personnes partant des territoires (localités) appartenant à la classe *i*,

* W. Alonso, A theory of movements. Rapport présenté à International Institute of Applied Systems Analysis, Décembre 1976, Laxenburg, Autriche).

M_{xj} — nombre total de personnes arrivant aux territoires (localités) appartenant à la classe j,

v_i — fonction décrivant les caractéristiques des territoires (localités) appartenant à la classe i, ou/et de leur population,

w_j — fonction décrivant les caractéristiques des territoires (localités) appartenant à la classe j, ou/et de leur population,

D_i — fonction décrivant une relation entre les territoires (localités) appartenant à la classe i et le reste du système, par l'unité de v_i (autrement dit, indice de la tendance migratoire centripète),

C_j — fonction décrivant une relation entre les territoires (localités) appartenant à la classe j et le reste du système, par l'unité de w_j (autrement dit, indice de la tendance migratoire centrifuge),

D_i et C_j sont des variables de système,

α, β — exposants — indices d'élasticité de tout le système des mouvements migratoires,

t_{ij} — relation particulière entre i et j, par exemple la facilité de transport.

Il faut souligner que le nombre d'équations présentées ci-dessus est excessif. En effet, pour définir le modèle il suffirait retenir les trois premières équations, ou bien les trois équations quelconques. Si l'on retient les trois équations quelconques du système, les autres peuvent en être déduites. Il existe ainsi plusieurs variantes de la construction et de la justification de cet ensemble d'équations. Cependant, la forme élargie présentée par l'auteur nous donne la possibilité de saisir toutes les approches possibles et tous les aspects de la problèmatique migratoire, les plus intéressants pour les chercheurs.

Alonso a démontré que tous les modèles de migration construits jusqu'à présent ne sont que des cas particuliers de son approche très généralisée. Puisque les paramètres — les indices d'élasticité sont les seuls éléments qui différencient la formalisation mathématique, les différences entre les modèles partiels doivent s'exprimer par la valeur de ces indices. En particulier, les différences se montrent lorsque les indices d'élasticité prennent la valeur de zéro ou de l'unité.

Si l'on considère toutes les combinaisons possibles formées par deux indices α et β dont les valeurs sont le zéro ou l'unité, on obtient quatre types des modèles. Ces quatre types, pris déjà en considération par A. G. Wilson, sont les suivants: 1) modèles de l'émission (*push models*) ayant $\alpha = 0$ et $\beta = 1$; 2) modèles d'attraction (*pull models*) ou $\alpha = 1$ et $\beta = 0$; 3) modèles de gravité élastiques ayant $\alpha = 1$ et $\beta = 1$; et 4) modèles inélastiques où $\alpha = 0$ et $\beta = 0$.

Ce ne sont que certains modèles, dits modèles d'interaction spatiale (formulés par A. G. Wilson) qui possèdent un (et seulement un) des indices dont la valeur est différente de zéro et de l'unité. Alonso mentionne aussi des situations théoriques dans lesquelles on pourrait peut-être attribuer des valeurs négatives ou bien des valeurs supérieures à l'unité aux indices. Néanmoins, ce ne sont que des cas particuliers et, de plus, il est douteux qu'ils existent en réalité pendant des périodes plus longues.

Il faut souligner qu'en majorité des cas le fait que les indices prennent la valeur de zéro n'était pas du tout, ou n'était que partiellement pris en conscience. Les auteurs des modèles ne se sont pas rendu compte du fait qu'en attributant à un des indices la valeur égale à zéro ils éliminaient certaines expressions de leur modèles. En effet, il ne faut pas oublier qu'un indice égal à zéro, lorsqu'on l'utilise comme un exposant, donne à toute expression exponentielle la valeur égale à l'unité.

La théorie des mouvements migratoires de W. Alonso contient des approches et des propositions intéressantes. Par exemple, elle offre une possibilité d'expliquer pourquoi nous avons des difficultés de définir certaines valeurs dans certains cas de l'application pratique des modèles. La formulation mathématique, telle quelle est envisagée dans le cadre de la théorie d'Alonso, montre qu'il est impossible de les identifier de façon univoque.

Dans ce rapport nous ne considérons que des situations dans lesquelles les indices d'élasticité prennent les valeurs égales à zéro ou bien à l'unité. Il faut ajouter que des modèles dans lesquels les deux indices prennent simultanément des valeurs intermédiaires dans l'intervalle de zéro à l'unité n'ont pas encore été construits, quoique ce cas reflète, selon Alonso, la situation plus proche à la réalité.

Nous supposons qu'à partir des remarques et des difficultés mentionnés plus haut on pourrait formuler une hypothèse d'une conception théorique, dans laquelle les indices d'élasticité ne prennent que des valeurs extrêmes de l'intervalle de zéro à l'unité. Nous obtiendrions ainsi une structure qui pourrait être nommée, par analogie à la structure des quanta en physique, la structure des quanta des mouvements migratoires. Mais pour atteindre ce but il faudrait d'abord confronter les quatre modèles de base aux systèmes migratoires réels, caractéristiques pour les différents pays et les différentes régions.

Envisageons tout d'abord des modèles de type gravitaire, élastiques et inélastiques, c'est-à-dire les modèles ayant les deux indices simultanément égales soit à l'unité soit à zéro. Nous proposons d'appeler ces modèles les modèles *d'équilibre* entre les deux types de mouvements migratoires, conformément à l'hypothèse soujacente que la demande de la main d'oeuvre (ou des migrants en général) est égale à l'offre de celle-ci. Étant donné une élasticité totale et l'équilibre au niveau de la région, les migrations peuvent s'effectuer soit entre des territoires soit entre des localités situés à l'intérieur de la région. Les différences locales sont couvertes par des flux migratoires provenant des territoires (des localités) ayant les surplus de la main d'oeuvre aux territoires (localités) souffrant de la pénurie de la main d'oeuvre. En réalité, ces mouvements ont toujours lieu à cause d'une différentiation de l'espace socio-économique du point de vue de la structure d'âge, de sexe et de la structure professionnelle. Si l'on considère le système inélastique, dans lequel tous les mouvements sont définis à l'avance, l'équilibre doit s'effectuer également au niveau local, à l'intérieur de tout territoire, toute localité. Ce n'est pas un hasard si les modèles de migration dont les indices sont inélastiques ont été formulés et appliqués avant tout dans le cadre des mouvements quotidiens de la population, c'est-à-dire des mouvements sans changement de domicile à l'intérieur de la ville ou bien de la région.

En effet, on observe facilement qu'en réalité les situations d'équilibre sont temporelles, souvent entièrement aléatoires. D'habitude, il existe un déséquilibre, fréquemment très net, entre l'offre et la demande de la main d'oeuvre. Selon des opinions prononcées par de nombreux théoriciens et praticiens, un faible surplus de la demande de la main d'oeuvre sur l'offre est nécessaire surtout pour le fonctionnement de l'économie capitaliste et, en particulier, celle de libre marché. Le problème de l'équilibre "dynamique", temporelle et aléatoire, devient plus facile à comprendre si l'on prend en considération que le processus d'égalisation de la répartition de la main d'oeuvre à l'aide de migration prend un certain temps et que la quantité de la main d'oeuvre, ainsi que le nombre de postes de travail sont soumis aux fluctuations permanentes, liées à l'adolescence des jeunes, à la retraite des vieux et à la création, développement et modernisation des établissements. Un élément supplémentaire, tel que le niveau d'activité professionnelle dont les fluctuations dépendent du niveau de vie et de la politique de l'emploi, renforce encore le caractère temporel de l'équilibre. Ainsi, les systèmes migratoires ayant une certaine élasticité se caractérisent par la présence des territoires qui possèdent soit un surplus soit un déficit de la main d'oeuvre, même s'il existe un équilibre au niveau de toute économie nationale. Les indices d'élasticité ont des valeurs égales à l'unité et à zéro en cas d'élasticité d'émission et d'inélasticité d'attraction, ou les valeurs de zéro et de l'unité en cas inverse.

En conclusion, les plus fréquents sont les systèmes migratoires extrêmes ayant une certaine stabilité et durabilité, c'est-à-dire les sytèmes ayant soit le surplus de la main d'oeuvre, soit le déficit. Le premier cas est plus fréquent que le second. Étant donné l'équilibre relatif au niveau national, ces systèmes ont un caractère régional ou local; en cas de déséquilibre à l'echelle nationale, ils ont un caractère national. C'est bien le cas des pays exportant ou important la main d'oeuvre.

Nous supposons qu'on peut admettre à la base des constatations précédentes que dans un pays à l'économie integrée à l'echelle nationale, et pour une période definie, domine un seul système migratoire auquel correspond un et un seul modèle, c'est-à-dire soit le modèle de surplus de la main d'oeuvre, soit le modèle de déficit. Il devient maintenant intéressant de savoir comment s'opère la transition d'un système à l'autre, comment faut-il et comment peut-on la modifier? En se référant aux modèles d'Alonso et à ses constatations il devient possible de répondre, ne serait-ce que partiellement, à ces questions.

Le passage d'un système de surplus de la main d'oeuvre à un système de son déficit peut être considéré, sur le plan théorique, comme un passage de l'indice, d'élasticité d'une valeur égale à zéro à une valeur égale à l'unité. Ce passage ne s'effectue que par le système de l'équilibre instable, défini par le modèle gravitaire élastique. Ce passage une fois effectué, l'indice d'elasticité de l'émission atteint d'abord la valeur de l'unité et ensuite se stabilise. Par contre, l'indice d'élasticité de l'attraction décroit jusqu'à zéro. Inversement, si l'on considère le passage du système de déficit au système de surplus, l'indice d'élasticité d'émission, après avoir dépassé le moment de l'équilibre, diminue jusqu'à zéro.

Il faut souligner qu'en cas de processus spontané de transition ce passage doit s'effectuer par l'intermédiaire d'un système élastique de l'équilibre (les deux indices étant égales à l'unité), c'est-à-dire par le système à l'interieur duquel peuvent exister des différences locales. Mais lorsqu'on considère le processus dirigé, ce passage ne s'effectue que par le sytème defini à l'aide d'un modèle gravitaire inélastique, c'est-à-dire par le système n'ayant pas de différences locales importantes. Autrement dit, il s'agit d'un sytème sans migrations spontanées permanentes.

Retournons maintenant, encore une fois, au problème de l'interprétation et celui de signification des valeurs des indices d'élasticité. Que signifie-t-elle la notion d'élasticité ou celle d'inélasticité dans le contexte des flux migratoires? Nous croyons qu'il est possible de les interpréter en termes de satisfaction (ou de dissatisfaction) de la demande de l'emploi (pour ceux qui cherchent l'emploi il s'agit de trouver des postes vacants, pour ceux qui les offrent il s'agit de trouver des travailleurs).

Abstraction faite d'une mobilité de la main d'oeuvre relativement insignificante, la possibilité de l'emploi existe, ou elle n'existe pas. De plus, supposant que la structure de l'offre de la main d'oeuvre correspond à celle de la demande, cette possibilité est univoque. En effet, nous avons soit une situation de surplus de la main d'oeuvre, soit son déficit. Il est donc logiquement correct d'accorder aux indices d'élasticité la valeur de l'unité (lorsque la possibilité de l'emploi existe) ou bien la valeur de zéro (en cas où cette possibilité n'existe pas). Les indices d'élasticité ne prennent pas des valeurs intermédiaires, car dans des situations temporelles il n'y a pas de conditions favorables à la stabilité des valeurs intermédiaires.

Finalement il reste à répondre à la question suivante: comment se forment les flux migratoires lorsque les migrants ne cherchent plus de l'emploi mais s'intéressent plutôt à l'amélioration du cadre de vie, notamment du point de vue de logement et du milieu? Nous supposons qu'il est possible de prendre ces cas en considération en agissant sur les variables de systeme, D_i et C_j, car les conditions de vie influencent, sans doute, les rapports entre l'offre de l'emploi et la demande. Une autre possibilité d'interprétation de ces cas s'offre à partir de la notion de l'utilité.

GEOGRAPHIA POLONICA 42, 1979

TAX SCHEDULES IN THE IDEAL CITY: EQUILIBRIUM VERSUS OPTIMALITY

Emilio Casetti

Department of Geography, The Ohio State University, Columbus, Ohio, USA

Grant Thrall

Department of Geography, McMaster University, Hamilton, Ontario, Canada

Many contributions to the literature on idealized urban centers are concerned with the spatial implications of the households' preferences for uncrowded residential sites and low commuting costs (Beckmann, 1973; Wingo, 1961; Alonso, 1964; Muth, 1969; Casetti, 1970a, b). These and other agglomerative and disagglomerative forces operate within idealized urban settings where only the interaction between households determines the distribution of population and land values. Models of this type generally involve the following. Households strive to maximize their utility subject to the condition that the cash outlays for various expenses, including those for commuting and for residential land do not exceed their income. Places of work are concentrated at one or more points and are surrounded by residential areas. Households are either identical in preferences and income or are differentiated into income classes (Beckman, 1969; Casetti, 1970a, 1972, 1974; Pines, 1975). Utility functions incorporate both the households' preference for larger parcels of residential land and a positive or negative preference for proximity to nodal points where jobs and services are concentrated (Casetti and Papageorgiou, 1971; Papageorgiou, 1971, 1973; Solow, 1973). Spatial equilibria are assumed to arise from the interactions among households, owners of land and producers of housing (Muth, 1969; Casetti, 1970b).

Equilibrium patterns are often contrasted to spatial patterns which correspond to the maximum of an aggregate welfare function (Mills and Deferranti, 1971; Mirrlees, 1972; Casetti, 1973; Riley, 1973; Dixit, 1973; Oron, Pines, and Sheshinski, 1973). One common element in these models has been the implicit assumption that equilibrium land values and population densities are unconstrained by pre-existing features and situations; they often do not allow for the possibility that spatial structures produced by earlier houshold preferences and transportation technologies may to some degree prevent subsequent preferences and technological environments from shaping density and land value surfaces. This is equivalent to assuming that long run equilibrium is attained instantaneously. However, mechanisms and factors associated with short run versus long run equilibria has been emphasized in some recent literature on housing capital (Muth, 1976; Evans, 1975; Anas, 1976).

A second common aspect of this class of models has been in the failure to include either public goods or taxes among the factors affecting the spatial patterns of population densities and land values. However, exceptions to this trend are emerging (Muth, 1969, p. 104; Barr, 1972; Thrall, 1975; Thrall and Casetti, 1977; Fisch, 1976) sometimes in connection with the investigation of externalities (Mills and DeFerranti, 1971; Hochman and Pines, 1971; Solow and Vickrey, 1971; Solow, 1972, 1973).

In this paper 'rational' households which maximize their utility subject to an income constraint are placed within a context where spatial patterns of population densities and land values are exogeneously given rather than endogeneously determined. Consequently the analysis focuses upon a short run time horizon over which land values do not change. Also, the model used incorporates an idealized local government which collects revenue and provides a single idealized public good. The consumption of this public good is associated with that of land in a manner reminiscent of Tiebout's formulation (Tiebout, 1956). The public good consumed by a household is defined as the product of the number of hours worked by public employees per unit of residential land times the amount of residential land utilized by the household. Police and fire protection, and in some respect utilities, may be thought of as real world analogues. The rationale for characterizing the public economy in this manner is in its well defined spatial qualities. A similar public good has been introduced into an urban model by Thrall (1975), and its relation with various types of local public goods found in the literature is discussed in Thrall-Casetti (1977).

THE MODEL

Assume the following. An ideal city occupies a circular area of radius h in an homogeneous plane. All jobs are concentrated at a central location which will be referred to as the CBD. All urban land is residential land. The city is inhabited by a fixed number P of one person ideal households which are identical in preferences and income. An indifference map

$$u = u[z, q, gq] \tag{1}$$

specifies the combinations of a composite good z, quantity of residential land q, and quantity of a locally provided public good gq, that would yield to a household the same level of satisfaction of its preferences. The quantity gq of public good consumed by a household is the product of hours of service by public employees per unit of land g, times the quantity of land consumed by a household q. Assume that u is a well behaved function of its arguments and that

$$u_z > 0, \tag{2}$$

$$u_q > 0, \tag{3}$$

$$u_g > 0, \tag{4}$$

$$u_{zz} < 0, \tag{5}$$

$$u_{qq} < 0, \tag{6}$$

$$u_{gg} < 0, \tag{7}$$

where all subscripted functions here and hereafter stand for partial derivatives of the functions with respect to the variables indicated in the subscripts. The households are confronted by a budget constraint that restricts a household's expenditures for commuting, composite good, land, and for the public good to equal its income; namely,

$$y = pz + r[s]q + t[s]gq + k[s] \tag{8}$$

where y is the households' income, and p, $r[s]$ and $k[s]$ are respectively the price of the composite good, land value, and commuting costs at distance s from the CBD, all exogenously given; $t[s]$ is a schedule of charges per unit of public good per unit of land. The nature and role of t will be discussed later in this section.

The households are 'rational' so that at any distance from the CBD, given values of p, r, and t, they will choose a mix of composite good, quantity of residential land, and of public good, such that the satisfaction of their preferences is maximized and their budget constraint is not violated.

The local government of the ideal city sets a schedule $t[s]$ of charges per unit of public good per unit of land for all the urban area. Given a schedule of $t[s]$ all the households accurately reveal how much public good they demand, and the local government will provide exactly the amount of public good that the households demand. In other words, the government chooses the $t[s]$ schedule, but it is bound to provide all the public good demanded as a result of such a schedule. Also, given a schedule of charges $t[s]$, the quantity $g[s]$ of public good per unit of land that a household at s demands on the basis of $t[s]$, the quantity $q[s]$ of residential land demanded, and the tax due by a household tgq become defined. To sum up, on the one hand t has the same role as any price in a consumer behavior model, but then the assumption is also introduced that the household pays tgq dollars to the 'local government'.

The local government is confronted by costs for the provision of the public good demanded, and the schedule of charges yields revenue to the government. The condition is imposed that the aggregate revenue must equal the aggregate cost of the public good. Consequently, in this model the local government is only allowed to select schedules of $t[s]$ that generate as much revenue as required to pay for the amount of public good demanded by the citizens on the basis of it. This balanced budget condition has also been employed in Thrall and Casetti (1977) in a somewhat different context.

In a capsule, this model is characterized by (1) an exogenously given land value function, (2) ideal households, (3) a local government that sets a schedule of charges for the local public good and then provides the amount of public good demanded on the basis of such a schedule, and (4) by the condition that the aggregate tax revenues should equal the aggregate costs of the public good.

Two $t[s]$ schedules will be investigated. The first one insures the existence of a spatial equilibrium by rendering the optimum utility level attainable by the households identical throughout the urban area (Casetti, 1971). The second schedule is designed to maximize an aggregate welfare function.

The mathematical formulations defining spatial equilibium and optimum welfare tax rate functions will be discussed first in general terms, then spatial equilibrium and optimal tax rate schedules will be derived for a specific utility function. Finally, a numerical illustration will be provided.

BACKGROUND MATHEMATICAL FORMULATIONS

A household at distance s from the CBD is confronted by given schedules of $t[s]$, $k[s]$, $r[s]$ and p. Denote by a bar superscript the optimum values of variables. Hence:

$$u[\bar{z}, \bar{q}, \overline{gq}] = \text{maximum}, \tag{9}$$

$$p\bar{z}+r[s]\bar{q}+t[s]\overline{gq}+k[s] = y. \tag{10}$$

The optimum values $\bar{z}$, $\bar{q}$, and $\overline{gq}$ are demand functions specifying the quantity of composite good, quantity of land of public good that would maximize a household's utility and are consistent with its budget constraint.

Assume that $u[z, q, gq]$ is so defined that $\bar{z}$, $\bar{q}$, $\overline{gq}$ exist and that these quantities demanded are nonconstant functions of their given prices only. Namely

$$\bar{z}_p,\ \bar{q}_r,\ (\overline{gq})_t \neq 0,$$

while

$$\bar{z}_r,\ \bar{z}_t,\ \bar{q}_p,\ \bar{q}_t,\ (\overline{gq})_p,\ (\overline{gq})_r = 0,$$

where, for example, $(gq)_i$ indicates the partial derivative of gq with respect to the variable i. Since $\bar{q}[r, s]$ is the quantity of residential land demanded by a household, and all urban land is residential land, the population density at s, denoted by $D[s]$ is

$$D[s] = (\bar{q}[r, s])^{-1}. \tag{11}$$

The population within a radius of s from the CBD, denoted by $P[s]$, is

$$P[s] = \int_0^s (2\pi s/\bar{q})\,ds \tag{12}$$

and the total population within the urban area, P, is

$$P[h] = \int_0^h (2\pi s/\bar{q})\,ds. \tag{13}$$

Since $\bar{q}$ is only a function of r, and not of p and t, by specifying $r[s]$ the variables $\bar{q}$, $D[s]$, and $P[s]$ are determined. This in turn implies that the constraint that the ideal city be inhabited by P households need not be considered in connection with the determination of optimal and spatial equilibrium tax schedules. This constraint will be satisfied by appropriate parameters of $r[s]$. If instead the assumption that $\bar{q}_t = 0$ is not made, condition [12] becomes one of the constraints entering into the determination of the optimal and spatial equilibrium $t[s]$.

Since $g[t, s]$ is the quantity of public good per unit of land demanded, the aggregate public good APG is

$$APG = \int_0^h 2\pi s g \, ds. \tag{14}$$

Let the cost of the public good be made of fixed and of variable components, CF and CV respectively. Then the cost of the aggregate public good $C[APG]$ is

$$C[APG] = CF + CV \int_0^h 2\pi s g \, ds. \tag{15}$$

The aggregate public revenue APR, generated by a schedule of $t[s]$ charges is

$$APR = \int_0^h 2\pi s t g \, ds. \tag{16}$$

Therefore, the condition that aggregate cost should equal aggregate revenue can be written

$$CF + CV \int_0^h 2\pi s g \, ds = \int_0^h 2\pi s t g \, ds$$

or

$$\int_0^h 2\pi s g (t - CV) \, ds = CF. \tag{17}$$

In other words, the condition that the aggregate cost of the public good should equal the aggregate public revenue is in the form of an integral constraint.

Define two distinct schedules of $t[s]$. The first one $\bar{t}[s]$ insures that the system is in a state of spatial equilibrium. The second schedule, $\dot{t}[s]$, maximizes an aggregate welfare function. It will be shown that $\bar{t}[s]$ creates a condition of spatial equity by insuring that all households will attain the same level of optimum utility, regardless of location, while $\dot{t}[s]$ maximizes the aggregate welfare and may or may not produce inequalities in the level of welfare of the households in the system.

SPATIAL EQUILIBRIUM $t[s]$

The derivation of the spatial equilibrium $\bar{t}[s]$ will be considered first. A rational household residing at distance s from the CBD and confronted by a given p, $r[s]$, $k[s]$, and $t[s]$, will demand quantities of composite good, of residential land, and of public good, identified by the demand functions z, q, and gq. The households will relocate or change consumption mixes only if by so doing they can attain a higher utility level. Call $\bar{u}$ the optimal utility level that a household may attain by choosing the 'best' mix of z, q, and gq; namely

$$\bar{u} = u[z, q, gq]. \tag{18}$$

A state of spatial equilibrium exists if no household is motivated to relocate or to change its consumption mix. This will be the case if the optimal utility level is identical throughout the city. Therefore, a spatial equilibrium condition is easily imposed by setting the optimal utility of a household equal to a spatially invariant constant u:

$$\bar{u} = \bar{u}[z, q, gq] = u = \text{constant}. \tag{19}$$

Since $r[s]$ is exogenous, $\bar{u}$ can be written as a function of t and s:

$$\bar{u} = \bar{u}[t, s].$$

Then call $\bar{t}[s]$ the function $t[s]$ such that

$$\bar{u}[t[s], s] = u = \text{constant} \tag{20}$$

for $0 \leqslant s \leqslant h$. Namely, the function $\bar{t}[s]$ is implicity defined by equation (20). The equations defining a schedule of $\bar{t}$ which generates a state of spatial equilibrium in the system and satisfies the condition that $C[APG] = APR$ are

$$\bar{u}[\bar{t}, s] = \bar{u},$$

$$\int_0^h 2\pi sg(\bar{t}-CV)\,ds = CF.$$

OPTIMAL $\hat{t}[s]$

Assume that the local government, interpreting ethical norms of the community, choses to define a welfare function consisting of the sum of the optimal utility levels of all the ideal households in the system. The locational welfare of the households in the annulus with boundaries at distance s and $(s+ds)$ from the CBD is equal to the number of the households in it times their optimal utility levels; namely, $(2\pi s\bar{u}/q)\,ds$. Consequently, the aggregate welfare function I is given by the integral

$$\int_0^h (2\pi s\bar{u}/q)\,ds. \tag{21}$$

The optimum problem under consideration involves determining a function $\hat{t}[s]$ defined over the interval $0 \leqslant s \leqslant h$ that maximizes the welfare integral I and satisfies the integral constraint

$$\int_0^h 2\pi sg(t-CV)\,ds = CF. \tag{22}$$

This maximum problem is a degenerate form of the isoperimetric problem in the calculus of variation of the type discussed in Casetti (1973). Define F, G, and H as

$$F = 2\pi s\bar{u}/q, \tag{23}$$

$$G = 2\pi sg(t-CV), \tag{24}$$

$$H = F-\mu G, \tag{25}$$

where μ is a Lagrangian multiplier. It can be shown (Casetti, 1973, p. 363 ff.) that in our problem an extremum function $\bar{t}[s]$, if it exists, must satisfy the necessary condition

$$H_t = F_t - \mu G_t = 0, \tag{26}$$

and that if

$$H_{tt} = F_{tt} - \mu G_{tt} < 0 \tag{27}$$

the extreme function identifies the required maximum.

THE INSTITUTIONAL SETTINGS OF $\bar{t}$ AND $\mathring{t}$

Let us now investigate what difference it makes to which households whether equilibrium or optimum schedules of charges for local public goods are adopted. Specifically we will identify which type of charges yield the highest level of satisfaction to the households at which location in the ideal city.

Since $z = z[s], q = q[s]$, and $g = g[s, t]$ rewrite $u = u[z, q, gq]$ in the form

$$\bar{u} = u[s, t].$$

It could be easily shown that $u_t < 0$.

$\bar{u}[s]$ and $\mathring{u}[s]$ are the optimum utility level attained by a household located at s respectively when the equilibrium and the optimum schedule of charges for the local public good are effective. Namely

$$\mathring{u}[s] = u[s, \mathring{t}[s]],$$

$$\bar{u}[s] = u[s, \bar{t}[s]].$$

For which values of s is $\mathring{u}$ greater than, equal to, or smaller than $\bar{u}$? Clearly, $\bar{u} = \mathring{u}$ only for those values of s for which $\bar{t} = \mathring{t}$. Also, since $u_t < 0$, $\mathring{u} > \bar{u}$ for those values of s for which $\mathring{t}(s) < \bar{t}(s)$. If $\mathring{t}(s)$ is not identically equal to $\bar{t}(s)$ throughout the urban area, and if $\mathring{t}(s)$ and $\bar{t}(s)$ are *smooth* functions of s, two alternatives are possible: (1) either $\mathring{u} > \bar{u}$ or $\mathring{u} < \bar{u}$ for $0 \leqslant s \leqslant h$ or (2) the interval 0 to h can be divided into subintervals with boundaries $h_0, h_1, \ldots, h_n$ where $h_i < h_{i+1}$ and $h_n < h$, such that $\mathring{u}(h_i) = \bar{u}[h_i]$ for all i' s, and either $\mathring{u}[s] > \bar{u}[s]$ or $\mathring{u}[s] < \bar{u}[s]$ for $h_i \leqslant s \leqslant h_{i+1}$. The households residing at distances s from the CBD for which $\mathring{t}[s] > \bar{t}[s]$ will attain a higher level of optimum utility if the spatial equilibrium rather than the optimum schedule of charges for the local public good is adopted, and *vice versa*. Rational households such as the ones considered in this context, will prefer equilibrium or optimal schedules of charges, respectively, if the former or the latter yield to them higher levels of optimal utility. Consequently if $\mathring{t}[s] > \bar{t}[s]$ for $0 \leqslant s \leqslant h$ all households will prefer the spatial equilibrium schedule of charges for the local public good. If $\mathring{t}[s] < \bar{t}[s]$ for $0 \leqslant s \leqslant h$ all households will have a preference for a maximum welfare $\mathring{t}$. However, if the $\mathring{t}[s]$ and $\bar{t}[s]$ curves intersect at values of s within the urban area, the preferences of the households will be determined by the distance band within which they reside. The nonobvious implication of this reasoning is that even rational households who are identical in tastes and income, may be induced by their position within the system to differ in their attitude toward welfare oriented planning.

This analysis raises issues of major 'philosophical' interest. The spatial equilibrium schedule of $\bar{t}[s]$ corresponds to an institutional setting in which the individual households pursue independently their 'interests' and preferences, while the government provides them with public good that they, in some sense, 'demand'. Instead the optimum schedule of $\dot{t}[s]$ corresponds to a setting in which the decisions are made by a central planner concerned with the maximization of an aggregate welfare. Therefore, the preference for spatial equilibrium or optimum schedules, in this context, is a preference for one of these two institutional settings.

Preferences for settings emphasizing the pursuit of individual interests or the collective welfare are ordinarily related to differences in income and social status as in the analyses of the causes of municipal fragmentation (Tiebout, 1956; Oates, 1969; Barr, 1973). It is indeed intriguing that opposing preferences for the two institutional arrangements may arise even in an abstract system characterized by rational and identical households, purely as the result of the location of the households' residences. It is also intriguing that within this ideal system, a setting could arise in which there is an incentive to form municipalities, namely institutions which levy taxes and provide public goods, extending over bands about the CBD, rather than over the entire urban area.

AN EXAMPLE

Equilibrium and optimum t schedules will now be derived for specific functions of u and r. Let

$$u = z^a q^b (gq)^c \tag{28}$$

and assume that $0 < a, b, c < 1$ so that restrictions (2) through (7) are satisfied. It can be easily shown that the values z, q and g that maximize (28) subject to the budget constraint (8) are

$$\bar{z} = ay^*/pv, \tag{29}$$

$$\bar{q} = by^*/rv, \tag{30}$$

$$\bar{g} = cr/bt, \tag{31}$$

where

$$y^* = y - ks,$$

$$v = a+b+c,$$

provided that the following inequalities hold

$$b+c < 1, \quad a+c < 1.$$

The inequalities arise out of the second order conditions for a constrained maximum of u. Equations (29) and (30) are the demand functions for the composite good and residential land, respectively. The quantity of public good demanded by a household $\overline{gq}$ is easily obtained as the product of (30) and (31):

$$\overline{gq} = cy^*/tv. \tag{32}$$

Since a large number of empirical studies have established that negative exponential functions often yield good fits to density distance data, let us specify $r[s]$ as a function that in conjunction with (11) and (30) yields a density relation of the negative exponential type. Namely, assume that

$$r[s] = (b/v)y^* e^{m-ns}, \tag{33}$$

so that

$$D[s] = e^{m-ns}. \tag{34}$$

Given the $D[s]$ function specified in (34) it becomes apparent that the integral constraint (13), stipulating that the aggregate population within radius h from the CBD must equal a given value of P, is satisfied for appropriate values of the parameters m and n of the rent function. That is, given n an m exists, or given m an n exists such that the integral (13) equals any arbitrary P. Consequently, the integral constraint may be disregarded in the problem of determining the spatial equilibrium and the optimum $t[s]$ functions. The spatial equilibrium problem may then be reformulated as follows: given equations (29), (30), (32) and (33), determine a function $\bar{t}[s]$ such that

$$z^a q^b (\bar{g}q)^c = \bar{u} \tag{35}$$

and

$$\int_0^h 2\pi s (y^*/v) e^{m-ns}(t - CV)\, ds = CF. \tag{36}$$

The function $\bar{t}[s]$ obtained from (35) and (33) is

$$\bar{\bar{t}}(s) = c\big((a/p)^a (y^*/v)^{a+c} (1/\bar{u}) e^{bns-bm}\big)^{1/c}. \tag{37}$$

It can be shown that for any CV, $CF > 0$ there is a $\bar{u} > 0$ such that (36) is satisfied. $\bar{t}[s]$ can be either an increasing or decreasing function of s. The spatial equilibrium demand for the local public good, using (32) and (37), is

$$\bar{\bar{g}}q = (a/p)^{-a/c} (y^*/v)^{-a/c} (\bar{u})^{1/c} (e^{mb-nbs})^{1/c}. \tag{38}$$

The demand for the public good (38) is a decreasing function of s.

The optimum schedule $\mathring{t}[s]$, if it exists, maximizes the aggregate welfare function (21) and satisfies the integral constraint (17). If a function $t[s]$ exists which satisfies equation (26) and inequality (27) for all values of s this function is the required optimum schedule $\mathring{t}[s]$. $\mathring{t}[s]$ is derived as follows. Using (23), (30), and (35), we have that

$$F = 2\pi s (a/p)^a (b/r)^{(b-1)} (c/t)^c (y^*/v)^{(v-1)} \tag{39}$$

or

$$F = Lt^{-c}, \tag{40}$$

where

$$L = 2\pi s (a/p)^a (b/r)^{(b-1)} (y^*/v)^{(v-1)} (c)^c. \tag{41}$$

Then, from (40)

$$F_t = L(-c)t^{-(c+1)} \tag{42}$$

and

$$F_{tt} = c(c+1)L^{-(c+2)}. \tag{43}$$

Similarly, from (24), (31), and (35),

$$G = 2\pi s(c/b)r-(CV)2\pi s(c/b)rt^{-1}. \tag{44}$$

Then from (44)

$$G_t = Mt^{-2} \tag{45}$$

and

$$G_{tt} = -2MT^{-3}, \tag{46}$$

where

$$M = (CV)2\pi s(c/b)r. \tag{47}$$

By substituting (42) and (45) into (26) the necessary conditions can be easily shown to be

$$(F_t-\mu G_t) = -cLt^{-(c+1)}-Mt^{-2} = 0 \tag{48}$$

and by substituting both (43) and (46) into (27)

$$(F_{tt}-\mu G_{tt}) = c(c+1)Lt^{-(c+2)}+\mu 2Mt^{-3} < 0. \tag{49}$$

Solving (48) for

$$\mu = (-cLt^{1-c})/M. \tag{50}$$

Substituting (50) into (49) and simplifying, we have

$$(F_{tt}-\mu G_{tt}) = cLt^{-(c+2)}(c-1). \tag{51}$$

Since

$$cLt^{-(c+2)} > 0, \quad (c-1) < 0,$$

it follows that

$$F_{tt}-\mu G_{tt} < 0,$$

which implies that the $t[s]$ which solves (48), if it exists, is the sought after $\dot{t}[s]$.

Solve (48) for $t[s]$ and the following is obtained:

$$\dot{t} = (\mu M/cL)^{1/(1-c)}. \tag{52}$$

Equation (50) shows that μ is a negative constant since

$$M, L, c > 0, \quad t > 0.$$

Using (41) and (47), (52) can be rewritten as

$$\dot{t}[s] = \left(-\mu N(y-ks)^{1-a-b}e^{bm-bns}\right)^{1/(1-c)}, \tag{53}$$

where

$$N = (CV)b^{-b}a^{-a}p^{a}v^{v-1}c^{-c}. \tag{54}$$

To determine whether some or all the ideal households in the model would prefer spatial equilibrium or optimum welfare schedules of charges if the utility function is specified as in equation (28), we procede as follows. The schedules $\bar{t}(s)$ and $\dot{t}(s)$ from (37) and (53) are introduced into the optimum utility function $\bar{u}[s, t]$, to obtain respectively $\bar{u}$ and $\hat{u}$. These are then compared to determine which of the two is consistently larger than the other for all s's, or alternatively whether over some interval(s) only one of the two is larger than the other.

It does not seem that further insights would be gained by pursuing the analysis at this level of generality since alternative sets of values of the model's parameters would produce different spatial patterns of the equilibrium and welfare optimum utility functions. Consequently, only an exploration of the differences of $\bar{u}$ and $\hat{u}$ based on a numerical specification of the model's parameters was attempted. This exploration is described in the section that follows. A similar approach was used in Thrall and Casetti (1978).

AN ILLUSTRATION

The approach that will be followed in this illustration consists in placing identical "typical households" into an idealized Columbus, Ohio, in which an idealized local public good is provided. The implementation of the approach requires (a) reasonable numerical values of a, b, and c for a typical household, (b) an average transportation cost per mile k, and (c) numerical values for y, P, h, m, n, CF and CV corresponding to the idealized metropolitan area of Columbus.

Essentially, the approach shifts the analysis to an ideal city with some specific traits of a real world entity, and allows using the abstract model discussed in the previous sections of this paper as a tool to extract the implications of selected aspects of reality.

The determination of the parameters a, b, and c of $\bar{u} = \bar{z}^a \bar{q}^b (g\bar{q})^c$ will be considered first. From equations (29), (30) and (31) we have

$$p\bar{z} = ay^*/v, \tag{55}$$

$$\bar{r}\bar{q} = by^*/v, \tag{56}$$

$$tg\bar{q} = cy^*/v \tag{57}$$

and substituting (55), (56) and (57) into the budget constraint we have

$$y^* = ay^*/v + by^*/v + cy^*/v. \tag{58}$$

Dividing (58) by y^* the following is obtained

$$a/v + b/v + c/v = 1. \tag{59}$$

Equations (55), (56) and (57) imply

$$p\bar{z}/y^* = a/v, \tag{60}$$

$$\bar{r}\bar{q}/y^* = b/v, \tag{61}$$

$$tg\bar{q}/y^* = c/v, \tag{62}$$

where the left sides of equations (60), (61), and (62) are operationalized as the percent of total income net of federal and state taxes and mean commuting expenses, spent in 1973 respectively on composite good, residential land and total local taxes, by an average United States four person urban household. The relevant data are given in Table 1. From a gross annual income of \$12626 we subtracted personal income taxes, social security and disability taxes, along with the transportation expenses as given in Table 1. The disposable net annual income obtained is \$9358.00; this remaining quantity is then split between the three items z, $\bar{q}$, and $\bar{gq}$ considered in this example.

Forty percent of the expenditure for housing (\$1163 on a total of \$2908) was imputed to land expenditure. Then $pq/y^* = \$1163/\$9358 = 0.1243$. tgq is assumed to be \$540, obtained by assuming \$300.00 in property taxes, \$120.00 in local sales taxes and \$120.00 in city income tax. Therefore, $t\bar{gq}/y^* = \$540/\$9358 = 0.0577$. Using equations (59) through (62) we have $b/v = 0.8180$ which is the proportion of y^* devoted to the composite good. One of the second order conditions involved

TABLE 1. Annual budgets for 4-person urban families, autum 1973*

Food	\$ 3,183
Housing	2,908
Transportation	1,014
Clothing	995
Personal Care	275
Medical Care	664
Other Consumption	722
Total Family Consumption	9,761
Other Items	611
Personal Income Taxes	1,607
Social Security Taxes and Disability	647
TOTAL BUDGET	\$ 12,626

* *Source*: Bureau of Labor Statistics, as quoted in *Information Please Almanac*, 1975, p. 80. Data based on estimates of costs for goods and services, rather than actual expenditures, as projected for an ideal "intermediate budget" household.

TABLE 2. Columbus public expenditures

Public good	Expenditures in Columbus per year (in thousands)
Highways	\$ 9,784
Police	15,057
Fire	11,237
Sewerage	9,306
Sanitation (not including sewerage)	5,198
Parks and Recreation	6,767
TOTAL	\$ 57,349

Source: United States Bureau of the Census, *City Governmen t Finances in 1971–72*, series GF72-NO. 4, U.S. Government Printing Office, Washington, D.C., 1973; p. 45.

in the household's constrained utility maximization is that $0 < v < 1$. Arbitrarily specifying $v = 0.9$ yields $a = 0.7362$, $b = 0.1119$, and $c = 0.0519$. The n and m parameters of the density distance function used are those estimated by Newling (1966) for Columbus, Ohio:

$$D[s] = \exp(9.2103 - 0.19s) \tag{63}$$

where $D[s]$ is persons per square mile at a distance of s from the CBD, (in miles). If we assume that the urbanized area ranges between the CBD and to where the population density is 2000 persons per square mile, the radius h of the idealized Columbus, Ohio, specifies to 8.4 miles.

Commuting costs per mile are calculated assuming one round trip between a household's dwelling and its place of work in the CBD for each of 260 working days per year, at the rate of $0.15 per mile, which yields a cost k of $78.00 per mile per year.

The cost of the aggregate public good $C[APG]$ was assumed to consist of a fixed cost CF, in dollars, plus a variable cost, obtained by multiplying a rate CV in dollars per hour, times the aggregate public good in hours. The values of CF and CV are calculated using data on police expenditures for Columbus, Ohio, as quoted in the *Municipal Yearbook*, 1975. The heroic assumptions and methodology for parameterizing the local public economy follow.

(a) The ratio B of fixed cost of police protection FCP to total expenditures on police protection TCP in Columbus, Ohio, was assumed to be representative of all fixed cost total cost ratios for the range of public goods depicted within this essay. In 1973, according to the *Municipal Yearbook* (1975)

$$B = FCP/TCP = \$720E3/\$20.931E6 = 0.0344. \tag{64}$$

(b) The fixed costs portion CF of the total public expenditure for Columbus was obtained by multiplying B from (64) times the total public expenditure obtained from *City Government Finances in 1971–72*, p. 45:

$$CF = TCA \cdot B = (0.0344)(\$57.349E6) = 1.95E6. \tag{65}$$

(c) Cost CV for Columbus in dollars per hour worked by public employees, is assumed to equal the cost per hour worked by the employees engaged in police protection within Columbus. CV is calculated dividing the total variable cost for police protection by the product of the total uniformed personel devoted to police protection service TPP times the number of duty per hour per year per employee in this service TDP, namely

$$CV = (TCP - FCP)/(TPP \cdot TDP). \tag{66}$$

Using 1973 data from the *Municipal Yearbook* (1975) we have

$$CV = (2.09E7 - 7.2E5)/(1105)(2080) = 8.79.$$

The value of TDP was calculated by multiplying 40 hours per week times 42 weeks. The total variable cost TVC, for Columbus is then the product of the cost per hour, CV, times the aggregate public good APG in hours worked by public employees.

TABLE 3. Numerical values of $\bar{t}$, $\dot{t}$, $\bar{u}$, and $\dot{u}$ for the idealized Columbus, Ohio

Distance from the CBD (in miles)	$\bar{t}$	$\dot{t}$	$\bar{t}-\dot{t}$	$\dot{u}$	$\bar{u}-\dot{u}$
0.0	1.97	9.83	−7.86	317.71	27.64
0.5	2.27	9.72	−7.45	320.25	25.10
1.0	2.61	9.60	−6.99	322.80	22.55
1.5	3.01	9.49	−6.49	325.37	19.98
2.0	3.47	9.38	−5.91	327.96	17.39
2.5	3.99	9.27	−5.28	330.56	14.79
3.0	4.59	9.16	−4.57	333.17	12.17
3.5	5.27	9.05	−3.77	335.81	9.59
4.0	6.06	8.94	−2.88	338.46	6.89
4.5	6.97	8.84	−1.87	341.22	4.22
5.0	8.01	8.73	−.72	343.80	1.54
5.5	9.20	8.63	.57	346.50	−1.15
6.0	10.56	8.53	2.04	349.21	−3.87
6.5	12.13	8.42	3.70	351.94	−6.59
7.0	13.92	8.32	5.60	354.69	−9.34
7.5	15.97	8.23	7.75	357.45	−12.10
8.0	18.32	8.13	10.19	360.22	−14.88
8.5	21.01	8.03	12.98	363.02	−17.67

The numerical values of a, b, c, y, k, and those of CF and CV for a "typical Columbus household" obtained using the procedures discussed above were entered into the relevant equations in order to derive numerical values for equilibrium and optimal t's as well as equilibrium and optimum utility levels. The results obtained which are shown in Table 3 indicate the following. The equilibrium schedule of charges $\bar{t}$ is smaller than the optimal schedule $\dot{t}$ between the CBD and a band 5 miles distance from the CBD. $\bar{t}$ and $\dot{t}$ become equal between 5 and 5.5 miles from the CBD. Beyond 5.5 miles the optimal charge is lower than the equilibrium one. $\bar{u}$ lies above $\dot{u}$ whenever $\bar{t} < \dot{t}$. Consequently the population within the inner annulus with boundary somewhere between 5 and 5.5 miles from the CBD will prefer spatial equilibrium schedules of charges, whereas households located in the outer annulus, beyond this boundary, will prefer an optimal schedule of charges.

CONCLUSIONS

This paper incorporates a local government which levies taxes and provides a public good into a mathematical land use model. Two distinct schedules of public good charges were evaluated: one creates a condition of spatial equity by insuring that all households attain the same level of optimum utility regardless of location, while the other maximizes the aggregate welfare and may or may not produce inequalities in the level of welfare of the households in the system. If the optimum utility levels attained by the household at spatial equilibrium are larger than the ones generated by the optimum charges over some distance bands about the central business district, even rational households who are identical in tastes and income wil differ with respect to their attitude toward welfare oriented planning and this provides a spatial rationale for conflicts of interest.

REFERENCES

Alonso, W., 1964, *Location and Land Use — Toward a General Theory of Land Rent*, Cambridge Harvard University Press.

Anas, A., 1976, A Short Run Model of Urban Growth, in G. J. Papageorgiou (ed.), *Mathematical Land Use Theory*, Lexington Books.

Barr, J. L., 1972, City Size, Land Rent and the Supply of Public Goods, *Regional and Urban Economics*, 2.

Barr, J. L., 1973, Tiebout Models of Community Structure, *Papers and Proceedings of the Regiona Science Association*, 30.

Beckmann, M. J., 1973, Equilibrium Models of Residential Land Use, *Regional and Urban Economics*, 3.

Beckmann, M. J., 1974, Spatial Equilibrium in the Housing Market, *Journal of Urban Economics*, 1

Beckmann, M. J., 1969, On the Distribution of Urban Rent and Resodential Density, *Journal of Economic Theory*, 1.

Casetti, E., 1970a, *Spatial Equilibrium Distribution of 'Rich' and 'Poor' Households in an Idealized Urban Setting*, Discussion Paper 13, Department of Geography, The Ohio State University.

Casetti, E., 1970b, On the Derivation of Spatial Equilibrium Urban Land Value Functions, in D. C. Sweet (ed.), *Models of Urban Structure*, Chapter 8, Lexington, Books.

Casetti, E., 1971, Equilibrium Land Values and Population Density in an Urban Setting, *Economic Geography*, 47.

Casetti, E., 1972, Spatial Equilibrium in an Ideal Urban Setting with Pareto Distributed Incomes, in E. Cripps (ed.) *Proceedings of the Annals of Regional Science*, London, Pion.

Casetti, E., 1973, Urban Land Value Functions: Equilibrium vs. Optimality, *Economic Geography*, 49.

Casetti, E., 1974, Optimum Urban Population Densities, in M. Yeates (ed.) *Proceedings of the International Geographical Union, Commission on Quantitative Geography*, Montreal, McGill--Queen's University Press.

Casetti, E., and G. J. Papageorgiou, 1971, A Spatial Model of Urban Structure, *Canadian Geographer*, 15.

City Government Finances in 1971–72, United States Government Printing Office, Washington, D. C., 1973.

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

Evans, A. W., 1975, Rent and Housing in the Theory of Urban Growth, *Journal of Regional Science*, 15, 2.

Fisch, O., 1976, Spatial Equilibrium with Local Public Goods: Urban Land Rent, Optimal City Size and the Tiebout Hypothesis, in G. J. Papageorgiu (ed.), *Mathematical Land Use Theory*, Lexington Books.

Hochman, O., and D. Pines, 1971, Competitive Equilibrium of Transportation and Housing in the Residential Ring of an Urban Area, *Environment and Planning*, 3.

Mills, E. S., and D. M. de Ferranti, 1971, Market Choice and Optimum City Size, *American Ecnomic Review, Papers and Proceedings*, 61.

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

Muth, R. F., 1969, *Cities and Housing, The Spatial Pattern of Urban Residential Land Use*, University of Chicago Press. Chicago.

Muth, R. F., 1976, A Vintage Model with Housing Production, in G. J. Papageorgiou (ed.), *Mathematical Land Use Theory*, Lexington Books.

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

Oates, W. E., 1969, Taxes and the Tiebout Hypothesis, *Journal of Political Economy*, 77.

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

Papageorgiou, G. J., 1971, The Population Density and Rent Distribution Model with a Multicentre Framework, *Environment and Planning*, 3.

Papageorgiou, G. J., 1973, The Impact of the Environment Upon the Spatial Distribution of Population and Land Values, *Economic Geography*, 49.

Pines, D., 1975, On the Spatial Distribution of Household According to Income, *Economic Geography*, 51.

Riley, J. G., 1973, Gammaville: An Optimum Town, *Journal of Economic Theory*, 6.

Solow, R. M., 1973, On Equilibrium Models of Urban Locations, in J. M. Parkin (ed.), *Essays in Modern Economics*, London, Longmans.

Solow, R. M., 1972, Congestion, Density and the Use of Land in Transportation, *The Swedish Journal of Economics*, 74.

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

Thrall, G. I., 1975, *A Geographic Theory of a Public Economy in Spatial Equilibrium*, Unpublished dissertation in the Department of Geography, The Ohio State University.

Thrall, G. I., and E. Casetti, 1978, Local Public Goods and Spatial Equilibrium in an Ideal Urban Centre, *Canadian Geographer*, 22.

Tiebout C., 1956, A Pure Theory of Local Expenditures, *The Journal of Political Economy*, 64.

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

SPATIO-ECONOMIC ANALYSIS OF THE URBAN SLUM: A PERSPECTIVE

NURUDEEN ALAO

Department of Geography, University of Lagos, Lagos, Nigeria

1. INTRODUCTION

A cursory reading of the literature is sufficient to confirm that considerable progress has been made in the last ten years or so in the theoretical analysis of the equilibrium structure of urban land uses. Perhaps the richest of such studies have been those that have exploited the Thünian-Alonso framework (viz. Muth, 1969; Beckmann, 1974; Solow, 1973; Mills, 1972) and have employed the neoclassical theoretical style and its vocabulary to establish valuable results about urban locational rent, urban housing and population density and the general problem of possible divergence between equilibrium and efficient urban land uses.

The commonest forms of assumptions that have formed the basis of the theoretical approaches described above have included the following: (i) a single urban centre to which all residents travel, (ii) uniform taste and income, (iii) housing quality uniform in all respects except that of areal density. It is often this variation in residential density with location rent which is employed to equalize utility and effect equilibrium. To be sure some measure of qualitative differences have been introduced into the areal structure of housing in these models through the incorporation of public goods and amenities characterized by differential spatial accessibility (see Wright, 1977; Polinsky, 1976; and Alao, 1977). These innovations, however, cannot cope with the problem of the areal differences in the quality of housing observable in the city, and are therefore unable to throw light on the problem of the emergence, and elimination of the slum. Why not?

There are three major reasons which may account for this inability. First, housing is regarded as a homegenous commodity or service within the competitive structure encompassed by these neoclassical models. Thus for such models we either have to regard all housing service as of high quality or as of low quality. There can be no mixture. Secondly, the static nature of such models precludes any focus on slums which strictly require a dynamic stucture for analysis. Thirdly, there is a real sense in which the economics of slum emergence and slum persistence is closely tied with the economics of externalities which are excluded from most of the models referred to above.

The slum has, however, been a persistent feature of the urban structure. It has constituted a major problem zone in the urban structure and seems to have defied many policy measures directed at its elimination. It is therefore obvious that a need exists for the construction of general models from which theoretical propositions can be derived about the behaviour of slums under various policy measures.

At present we are still far from realizing such models. This paper is intended to provide some simple frameworks from which some pertinent policy questions can be answered concerning the slum. Accordingly, in the next section we briefly highlight the results achieved by three existing economic approaches which I refer to as game-theoretic approach, competitive theoretic approach and macro-economic income-based approach. In the third section we establish models which clarify many of the issues raised in the second section.

2. CRITICAL REVIEW OF EXISTING APPROACHES

a. THE GAME-THEORETIC APPROACH

There is a sense in which the study of the external determinants of property qualities and property values can be said to have been initiated by Alfred Marshall. Thus in Book Five, chapter eleven of his *Principles*, p. 369, Marshall said,

> But the general rule holds that the amount and character of the building put upon each plot of land is, in the main (subject to the local building bylaws), that from which the most profitable results are anticipated with little or no reference to its reaction on the situation value of its neighbourhood. In other words, the site value of the plot is governed by causes which are mostly beyond the control of him who determines what buildings shall be put on it.

Marshall's discussions was far from systematic and his comments were largely incidental. It was, however, Davis and Whinston (1961) who first provided an economic framework based on the game-theoretic notion of the Prisoner's Dilemma from which testable propositions or hypotheses about slum structures could be deduced, analysed and further developed. Exploiting the game-theoretic framework alluded to above, Davis and Whinston maintain that the pay-off matrix confronting owners of slums in a given neighbourhood is such that would (if individual action is the rule) make poor maintenance the optimum decision. The persistence of the slum or blight is predicated upon interdependencies implied by the Prisoner's Dilemma and upon the prevalence of individual action. But then, it is clear from the system presented by Davis and Whinston that collective action would socially be most profitable. Consequently, it would pay the parties concerned to agree among themselves to invest in the improvement of their property so that benefits can accrue to all of them. Alternatively one of them might purchase all the properties and undertake the improvement and reap the benefits. Thus Davis and Whinston explain that in so far as the Prisoner's Dilemma is useful as a framework for understanding the slum it leads one to the assertion that the real explanation for the prevalence of the slum must be sought in the factors which militate against coordinated or collective action on the part of property owners. On this point, Davis and Whinston observe that "while it might be easy for the two property owners in our simple example to commu--nicate and coordinate their decision this would not appear to be the case as the num ber of individuals increased. If any single owner were to decide not to invest while all other owners decided to redevelop then the former would stand to gain by such action. The mere presence of many owners would seem to make coordination more difficult and thus make our assumption more realistic". It is this situation that justifies the use of the power of eminent domain by governement to effect urban renewal.

There are two important outcomes of this game-theoretic analysis. The first is the definition of an urban blight or urban slum as a state of affairs in which (a) strictly individual action will not result in redevelopment, (b) coordination of action would result in redevelopment, and (c) the sum of benefits from renewal would exceed the sum of costs. The second outcome is that it is possible to establish a theoretical argument for the need for governmental intervention to effect urban renewal.

The several shortcomings of this analysis fall into two broad categories: theoretical and empirical. Theoretically, although Whinston–Davis analysis indicates that slum equilibrium is a stable equilibrium and represents a misallocation of resources, it is impossible through this analysis to establish how much investment should be put into urban renewal, i.e. what quality of the environment is optimal. Furthermore the analysis is theoretically incomplete in that (a) it totally excludes the possibility of multiple equilibria one of which may occur at a higher level of housing and environmental quality and which may at that level be stable and so be attainable through decentralized decision making, and (b) it is silent on the transaction costs involved in the assembly of plots in reaching the conclusion that collective decision-making would result in better resource allocation and hence increased welfare.

On the empirical front, Muth (1969) and Mills (1972) have pointed up three shortcomings of the Davis–Whinston analysis:

1. If the Davis–Whinston thesis is valid, then public urban renewal projects should be profitable. However, most of these projects lose money at least in the United States and in many parts of Europe.

2. Using the Davis–Whinston theory we must come to the conclusions that slums need not be concentrated in any particular zone in the urban area and furthermore, where individual decision-making prevails and urban population grows, the proportion of poor-quality housing in the city must increase. Empirical evidence contradict these conclusions because the overall quality of urban housing has been on the increase, and because slums tend to be locationally concentrated rather than dispersed.

We must then conclude that the Whinston–Davis theory has failed to capture the major processes that are responsible for the creation and perpetuation of the slum.

b. COMPETITIVE-THEORETIC ANALYSIS

Olsen (1969) provided a closer study of the competitive processes involved in the formation of the slum. In his scheme he evolved a sharper characterization of the filtering process by which housing quality is adjusted and was able to offer reasons why urban renewal projects undertaken by governments have so often been ineffective and unprofitable.

Olsen made seven major assumptions which are a suitable translation of the usual conditions required for the validity of competitive market framework as follows:

1. The object of sale in the market is housing service of which the buyers and sellers (who will hereafter be referred to as actors) are numerous.

2. The quantity of the service which each seller commands or which each buyer purchases is small compared to the total available in the market.

3. Each actor in the market seeks his own self interest and there is no collusion·

4. Entry into and exit from the market are completely free.

5. Perfect knowledge of prevalent market conditions is possessed by every actor; each buyer employs that knowledge to maximize his utility and each seller to maximize his profit.

6. No artificial restrictions are imposed on demands for or supplies of housing services; there are no rent controls.

7. Housing service is a homogeneous commodity.

While it is possible to attack these assumptions on the basis of their realism, we shall not pursue that line. We shall instead examine the analytical results based on these assumptions and concentrate on the validity of the prescription offered by Olsen.

The deductions are mainly long-run situations and for their validity we must bear in mind the point that the unit price of housing service must be independent of the size of package in which the service is offered. That price is coincident with the long-run minimum average cost. A slum dwelling unit may be defined as one which provides lower than an arbitrary predetermined level of service per unit time. Correspondingly a slum area is one in which slum units predominate.

The market's goal is to reach that state in which the prevalent price per unit of housing service is the long-run minimum average cost. What is the process embodied in this behaviour? Simply put, it is the filtering process. A housing unit is said to have filtered if the quality (and *ipso facto* the quantity) of service yielded by it per unit of time has changed. It is said to have filtered up or down according to the change being positive or negative.

Suppose that for any particular housing service unit, the existing demand (in the sub-market) exceeds the existing supply. This could happen, for example, in the case of slum if an area experiences immigration of workers in search of jobs. Such immigration, other things being equal, would shift outwards the demand for slum--type housing. As the supply is fixed in the short run, the housing price in this sub-market would rise and suppliers in such a market would begin to enjoy excess profits. Landlords would allow their properties to filter down to slum level so that on the one hand they increase slum housing supply and on the other, decrease supply of higher quality housing. Thus through the filtering process excess profit in the slum market is graudally competed away, slum housing price falls, while the price of higher quality housing rises. Equilibrium is reached only when the long-run minimum average cost per unit prevails.

The process briefly enunciated above enables one to provide definite answers to two important questions in the housing market. One of these questions is: is it possible for slum housing to be more expensive in relation to the service it performs than higher quality housing? The second question is, can urban renewal be effected by slum clearance scheme, i.e. is it possible permanently to eliminate the slum from an urban system by mere bulldozing of the existing slum? In both cases the answer is in the negative.

The scheme described above leads to the conclusion that it is theoretically impossible to uphold the assertion of Muth that slum housing is expensive in relation to the service it renders. For if that were so, landlords in that sub-market would be enjoying excess profit and other landlords would allow their properties to filter down to the slum level, thereby increasing the supply of slums, depressing their high price and consequently competing away the excess profit.

We can also demonstrate that physical clearance of slums cannot provide the desired lasting solution to the existence of slums. For suppose as an example that the central slum in an urban area is obliterated. Slum supply is thereby curtailed. Since this exercise does not affect the incomes of slum occupiers, it does not shift the demand for slum housing. The equilibrium in the slum sub-market is, however, disturbed by the change in suply, so, slum-housing price goes up, and slum owners enjoy excess profit. Consequently, other home owners find it profitable to allow their properties to filter down to make up for the shortfall in supply. This filtering down causes decrease in the supply of better quality housing whose price must consequently begin to rise. The increase in price in the high quality housing market leads to increase in construction of high quality houses. Hence a new equilibrium will have to emerge in which new slums are created to replace the destroyed ones and in which the demand for and supply of higher quality housing are brought into equality. Where the new slums would be created will depend upon the economics of the urban land use of the particular city, viz: the relative spatial disposition of work-place, amenities, shopping centres, etc. The slum clearance exercise may also incur monetary losses for the following reasons. The process of plot assemblage, clearance and renewal often takes a long time (up to ten years) and factor costs especially of labour may rise considerably. So that on the cost side the exercise may incur much more than was anticipated at the planning stage. On the revenue side the general flattenning of the rent gradient may reduce the rent per acre that may be realizable.

One of the most important conclusions to which the competitive analysis leads is that housing quality is a direct function of income and that slum housing is directly attributable to low income or poverty. Hence to ensure the consumption of high quality housing service it is necessary to subsidize the incomes of the poor. Olsen has suggested that:

probably the most efficient method of subsidizing housing of low income families is to allow these families to buy certificates which they could use to pay the rent or make mortgage payments up to an amount equal to the face value of the certificate. The low income family would purchase this certificate for an amount less than the face value. These certificates would be redeemed by the government from sellers of the housing service. It would be illegal to exchange these certificates for other than housing service. (Olsen, 1969, p. 620).

He has thus shown one possible avenue of exploiting the conclusion for further analysis. Namely, he has provided an operational means of effecting income subsidy and has paved the way for the possibility of empirically ascertaining the influence of income on housing.

c. MACRO-ECONOMIC ANALYSIS

The alternative use that one can make of the conclusion about income is to explore the theoretical implications of having income play a central role in determining the quality of housing. Baumol in a couple of papers has provided excellent insights along this line. Baumol (1972) and Oates (1971) have asserted that slum growth is but an example of the phenomenon of cumulative deterioration in a system.

The dynamics of environmental quality based on the Baumol's principles may be outlined as follows. Let A be an area located within a larger region Ω such that free flow of people takes place between A and Ω-A in response to changes in environmental quality in region A. Denote by $q(t)$ the environmental quality (which is here synonymous with the quality of housing service) in A at time t; by $y(t)$ the income per capita in A at time t. We can now write four definitional equations:

$$y(t) = \varphi(q(t)), \quad \varphi' = \frac{d\varphi}{dq} > 0, \tag{1}$$

$$q(t+1) = \psi(y(t)), \quad \psi' = \frac{d\psi}{dy} > 0, \tag{2}$$

$$y(t) = \varphi_0\psi(y(t-1)), \tag{3}$$

$$\frac{dy(t)}{dy(t-1)} = \psi' \cdot \frac{d\psi}{dy(t-1)} > 0. \tag{4}$$

These equations embody important dynamic processes which must be clearly described. Equation (1) is validated by the fact that any increase in environmental quality in A attracts high income individuals into A instantly and hence causes the per capita income in A to rise. Conversely, environmental quality decline exerts heavy pressure on high income individuals to move out of A causing a decline in income and via equation (2) further decline in environmental quality. Equation (2), however, indicates that changes in quality reflect a clear lag. Combination of (1) and (2) yields (3) which establishes a dynamic path of per capita income in region A. Since it is difficult to construct a precise picture for such a path we shall substitute simpler affine relations for equations (1) and (2). Thus we have

$$y(t) = a+bq(t), \quad b > 0, \tag{5}$$

$$q(t+1) = c+hy(t), \quad h > 0. \tag{6}$$

We can combine equations (5) and (6) into a single first order difference equation (in environmental quality) whose solution is

$$q(t) = q_e + (q_0 - q_e)m^t, \tag{7}$$

where:

q_0 = initial housing quality,

q_e = equilibrium housing quality,

$m = bh > 0$.

Equation (7) describes the time path of improvement (degradation) in housing quality in A. If $m < 1$, the equilibrium quality q_e is a stable one. In such a case any initial disturbance of the system (through, for example, government assistance in renovation) will be met by countervailing forces which ensure that the original equilibrium q_e is re-established. The sticky problem is that occasioned when q_e is a slum equilibrium. When such is the case huge expenditures in slum clearance which raises initial average quality q_0 but which does not affect the parameters a and b of equations (5) and (6) cannot provide lasting amelioration of the environment.

On the other hand, if $m > 1$, a valid strategy for escaping from the poverty trap would be that involving an initial big push, e.g. large once-and-for-all investment in renovation.

The truth is probably that a mixture of the two strategies would be demanded in reality. If that is the case we must be able to understand the factors which determine the parameters a, b, c, and h. These in turn determine the nature and magnitude of the equilibrium quality of housing.

It is ultimately through a disaggregated approach that we can hope to grasp the nature of the determinants of those parameters. Consequently, in the next section I shall put forward mathematically simple models which are designed to throw light on the determinants of the housing quality in a dynamic sense. These models are related to a class recently developed by Schall (1976) in three major respects:

(i) They are disaggregative.

(ii) They emphasize the interdependent nature of the decisions which ultimately create slums.

(iii) They permit the exploration of the implications of slum abatement or removal via individual decision-making, cooperative (but nongovernmental) decision process and state solution.

They however differ from Schall's analysis because they are mathematically simpler, spell out all essential assumptions and in addition contradict some of the major results obtained by Schall.

3. ELEMENTARY DECISION-THEORETIC MODELS OF SLUMS

We assume the existence of two properties on two locationally contiguous areas Ω_1, and Ω_2, each property owned by a separate landlord. The quality of each property denoted by q_i, $(i = 1, 2)$, is some summary measure of the level of service derivable from each property. The level of q_i depends upon the level of current investment on the upkeep of property i and the level of q_j $(j \neq i)$. Given q_j, the owner of property i, adjusts q_i so as to maximize his net returns π_i. Thus the gross returns f_i is defined as $f_i = f_i(q_i q_j)$, $i, j = 1, 2, i \neq j$, and is assumed to be twice continuously differentiable. In addition we assume

$$\partial\pi_i/\partial q_j > 0, \quad i \neq j, \; i = 1, 2,$$
$$\partial^2\pi_i/\partial q_i^2 < 0, \quad i = 1, 2, \tag{*}$$
$$\partial^2\pi_i/\partial q_i \partial q_j > 0, \quad i, j = 1, 2, \; i \neq j.$$

The two property owners face the following net returns functions in their contiguous neighbourhoods:

$$\pi_1 = f_1(q_1 q_2) - C_1(q_1), \tag{8}$$

$$\pi_2 = f_2(q_1 q_2) - C_2(q_2). \tag{9}$$

We use equation (8) to indicate that owner 1 will assume that owner 2 will not change the quality of his property. On the basis of this assumption owner 1 will determine the appropriate quality level for his property which will enable him maximize his net returns. $C_i(q_i)$ with $C_i'(q_i) > 0$ is the cost function. The same interpretation (*mutatis mutandis*) goes for equation (9).

The first order conditions for the maximization of (8) and (9) are

$$\partial\pi_1/\partial q_1 = \partial f_1/\partial q_1(q_1 q_2) - \partial C_1/\partial q_1 = 0, \tag{10}$$

$$\partial\pi_2/\partial q_2 = \partial f_2/\partial q_2(q_1 q_2) - \partial C_2/\partial q_2 = 0. \tag{11}$$

Examine equation (10) in conjunction with the assumptions embodied in (*). These assumptions guarantee that the implicit function theorem holds in respect of q_1 and q_2 of equations (10) and (11). So that if q_1^* and q_2^* are the optimal solutions of (10) and (11), then there exist functions h_1 and h_2 continuous throughout an appropriate domain, so that

$$q_1^* = h_1(q_2), \tag{12}$$

and

$$q_2^* = h_2(q_1). \tag{13}$$

Differentiating equation (10) we obtain

$$\frac{\partial^2\pi_1}{\partial q_1^2}\left(\frac{dq_1}{dq_2}\right) + \frac{\partial^2\pi_1}{\partial q_1 \partial q_2} = 0. \tag{14}$$

Hence,

$$\frac{dq_1^*}{dq_2} = \frac{-\partial^2\pi_1}{\partial q_1 \partial q_2} \Bigg/ \frac{\partial^2\pi_1}{\partial q_1^2}. \tag{15}$$

Thus

$$\frac{dq_1^*}{dq_2} > 0.$$

We can similarly obtain the result

$$\frac{dq_2^*}{dq_1} > 0.$$

The continuous functions h_1 and h_2 represent the decision curves for owners 1 and 2, respectively. Each of these functions has a positive slope, and given that quality levels are determined by individual decision, h_1 and h_2 contain all the information needed to map out the path leading to a slum equilibrium or to a higher quality equilibrium. Figure 1 shows one possible disposition of the curves, h_1 and h_2. The pair of points Z_1 and Z_2, where $Z_1 = (q_1^{s_1}, q_2^{s_1})$ and $Z_2 = (q_1^{s_2}, q_2^{s_2})$ are equilibrium points in the sense that $h_1(q_1^{s_1}) = h_2(q_2^{s_1})$ and $h_1(q_1^{s_2}) = h_2(q_2^{s_2})$. The point defines a stable (slum-quality) equilibrium whereas the point $(q_1^{s_1}, q_2^{s_2})$ is an unstable higher quality equilibrium.

The former is the more likely of the two to prevail if competitive (atomistic) decision-making is the rule. The arrows in the figure indicate the path of reaction to any disturbance of the equilibrium. The differences in the stability property of the two equilibria have important implications for any policy designed to improve the quality of housing in the area. If the neighbourhood were in an initial slum equilibrium state in which the properties were in quality states $q_1^{s_1}$, $q_2^{s_1}$ respectively, any of the following steps may be taken to improve the situation. (a) The owners can be persuaded individually to make investments towards improving their property. (b) The government may induce the individuals to effect renewal on their property and set up machinery to enforce a minimum quality code. Clearly step (b) is likely to be the more effective policy in attempting to raise the quality beyond the level indicated by Z_1 because the stability of Z_1 indicates that unconstrained individual behaviour would always eventually return the property condition to a state of slum condition. One must emphasize that there is nothing in the argument proffered so far to indicate that the effective step (b) will be profitable in the sense of yielding net social return.

On the other hand, if the initial equilibrium state is Z_2, then any inducement which yields the increase in quality level will ensure that subsequent individual actions would lead cumulatively to improvement in neighbourhood conditions. Hence no governmental control need be exercised to enforce any quality code, and hence Z_2 is less costly socially to improve than Z_1.

So far, we have sought to analyze only results associated with purely individual uncoordinated action. What are the decision curves associated with coordinated action or associated with a government decision to assemble plots and invest in property improvement. For simplicity in this respect we just assume that costs are just the sum of individual costs and returns are similarly the sum of individual returns. The goal in this case is to maximize $\pi = \pi_1 + \pi_2$ where π_1 and π_2 satisfy the marginal conditions (*) stated earlier. The first order conditions are now,

$$\frac{\partial \pi}{\partial q_1} = 0 \Rightarrow \frac{\partial \pi_1}{\partial q_1} + \frac{\partial \pi_2}{\partial q_1} = 0 \Rightarrow \frac{\partial f_1}{\partial q_1} - C_1' + \frac{\partial f_2}{\partial q_1} = 0, \tag{16}$$

$$\frac{\partial \pi}{\partial q_2} = 0 \Rightarrow \frac{\partial \pi_1}{\partial q_2} + \frac{\partial \pi_2}{\partial q_2} = 0 \Rightarrow \frac{\partial f_1}{\partial q_2} - C_2' + \frac{\partial f_2}{\partial q_2} = 0. \tag{17}$$

The validity of the implicit function theorem here implies the existence of functions H_1, H_2 such that $q_1^G = H_1(q_2)$ and $q_2^G = H_2(q_1)$, where the superscript G indicates that the equilibrium values are those associated with coordinated or government action. What is the relationship between H_1 and h_1 on the one hand, and H_2 and h_2 on the other? To obtain the critical relationship, observe equation (16). It shows that

$$\frac{\partial \pi_1}{\partial q_1} = -\frac{\partial \pi_2}{\partial q_1}, \tag{18}$$

Equation (18) together with the assumption $\frac{\partial \pi_i}{\partial q_j} > 0$, $i \neq j$, implies

$$\frac{\partial \pi_1}{\partial q_1}(q_1^G, \bar{q}_2) < 0. \tag{19}$$

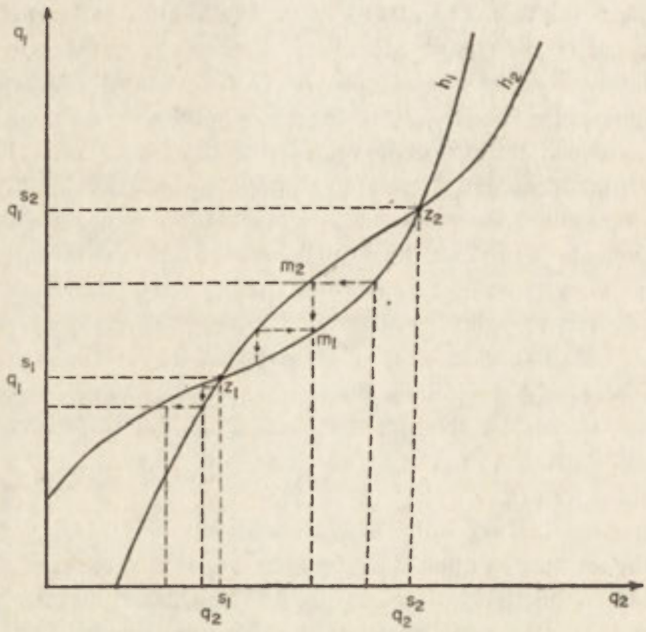


Fig. 1. Housing quality decision curves

Thus given any $\bar{q}_2$, the optimal coordinated solution q_1^G occurs at a point where the individual profit is declining. Recalling that q_1^* is the optimal solution in the individual case, we must have $q_1^G > q_1^*$. Hence for every q_2, it must be true that

$$H_1(q_2) > h_1(q_2). \tag{20}$$

By the same reasoning, we must have,

$$H_2(q_1) > h_2(q_1). \tag{21}$$

However, many possibilities exist for the values of q_1^G and q_2^G which satisfy the equilibrium,

$$H_1(q_2^G) = H_2(q_1^G).$$

One such is that these equilibrium values lie inside the closed curve $Z_1 M_1 Z_2 M_2$ in Fig. 1.

We have another possible set of situations in Fig. 2 which displays curves h_1, h_2, H_1 and H_2 with single equilibrium in each case. In Fig. 2, Z_h and Z_H are respectively the equilibrium decentralized and equilibrium cooperative solution. In this figure the cooperative equilibrium produces universally higher quality housing than individual decentralized equilibrium. Actually we can express a sharper conclusion, namely that given single equilibrium situation, Z_H provides higher quality on all plots than Z_h whenever Z_h (and hence Z_H) is a stable equilibrium.

We can now highlight the major differences between the results obtained here and those established in Schall (1976), and subsequently indicate lines of generalization for the present model.

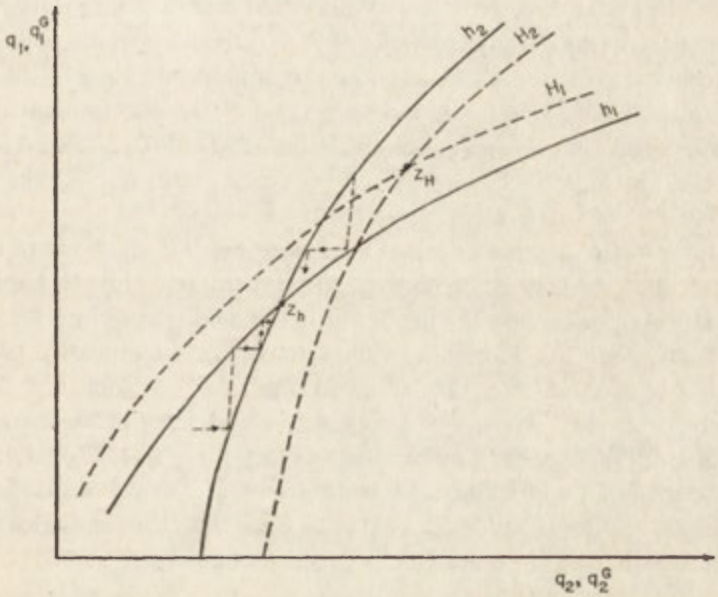


Fig. 2. Alternative solution involving cooperative decision

Whereas in Schall's model, in the state of competitive equilibrium every real property must have the same quality, here it will be accidental if that situation holds; indeed in competitive equilibrium, the quality levels of each real property may be different. Secondly in Schall's model coordinated equilibrium solution always results in higher housing quality in respect of every plot of land compared with competitive equilibrium. In the present model coordinated action will in general have differential impact on the quality of housing on different plots and such quality will not always be higher. It is therefore a model such as that developed in this paper which brings out the need for a third type of solution — which may be called state solution — in which the critetion of equality in housing standard is set as the objective of the system. Whereas the "state solution" has no visible role to play in Schall's scheme, its role is clear in our model.

The generalization of the models treated in this section for a case in which we have

n plots ($n > 0$) can be achieved in one of two ways. First the revenue function f_i can be defined as $f_i(q_1, q_2, ..., q_i, ..., q_n)$ while the cost function is merely $C_i(q_i)$. This definition will yield n first order conditions, for which situation we obtain a complex picture and general conclusions are then hard to draw. Alternatively we can define each revenue function f_i to be $f_i(q_i\varphi_i)$ where $\varphi_i = \varphi_i(q_1, q_2, ..., q_{i-1}, q_{i+1}, ..., q_n)$. For example, for Schall's case $\varphi_i = \frac{1}{n-1}\sum_{j\neq i} q_i$ and for that case the analysis can be treated graphically.

4. CONCLUSIONS

This paper has been directed at a very restricted aspect of urban renewal problem — the problem of the spatio-economics of slum formation. Sound policy prescriptions can hardly be proffered in the absence of firm understanding of the processes which sustain the slum. This paper has then (a) collated and brought into sharp focus the more significant theoretical results in the area and (b) obtained some independent results which are in disagreement with existing ones notably those derived by Schall.

A fair number of policy prescriptions have been made in the body of the paper. However, the following may be highlighted. (i) It is probably idle to hope to effect large-scale improvement in housing quality or to eliminate the slum before (income) poverty is eliminated. (ii) Slum formation is basically a cumulative process that cannot readily be terminated by attending to only one symptom, e.g. destruction of slum houses. Detailed studies of the empirical correlates of the parameters of equations (5) and (6) are called for, because it is they that can really change the dynamics of the system. (iii) Finally, there is no *prima facie* evidence that in all cases government intervention is called for. We have seen that there are cases where decentralized decision may be more effective and less costly.

REFERENCES

Alao, N., 1977, *The Relevance of Equilibrium Models in Urban Land Use Analysis and Policy*, Paper presented at 1977 Conference of the Nigerian Economic Society.

Alao, N., 1977, Housing Location Processes and Policy Intricacies, *Nigerian Geographical Journal*, 20.

Alonso, N., 1964, *Location and Land Use*, Cambridge, Harvard Univ. Press.

Asmus, K. H. and H. J. Iglarsh, 1975, Dynamic Model of Private Incentives to Housing Maintenance: Comment, *Southern Economic Jornal*, 42, 323–333.

Bailey, M. J., 1959, Note on the Economics of Residential Zoning and Urban Renewal, *Land Economics*, 35, 288–292.

Baumol, W. J., 1972, The Dynamics of Urban Problems and Its Policy Implications, in M. Preston and P. Corry (eds.), *Essays in Honour of Lord Robbins*, Lond Weiderfeld and Nicolson, 380–393.

Beckmann, M. J., 1974, Spatial Equilibrium in the Housing Market, *Journal of Urban Economics*, 1, 99–107.

Breger, G. E., 1967, The Concept and Causes of Urban Blight, *Land Economics*, 43, 369–376.

Davis, O. A. and A. B. Whinston, 1961, The Economics of Urban Renewal, *Law and Contemporary Problems*, 26, 105–117.

Dildine, L. L. and F. A. Massey, 1974, Dynamic Model of Private Incentives to Housing Maintenance, *Southern Economic Journal*, 40, 631–639.

Lowry, I. S., 1960, Filtering and Housing Standards: A Conceptual Analysis, *Land Economics* 36, 362–370.

Marshall, A., 1959, *Principles of Economics*, Bk. V, 8th edition, London, Macmillan.

Mills, E. S., 1972, *Studies in the Structure of the Urban Economy*, Baltimore, Resources for the Future.

Mills, E. S., 1972, *Urban Economics*, Glenview, Ill., Scott, Foresman & Co.

Moorhouse, J. C., 1972, Optimal Housing Maintenance Under Rent Control, *Southern Economic Journal*, 39, 93–106.

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

Oates, W. E., E. P. Howrey and W. J. Baumol, 1971, The Analysis of Public Policy in Dynamic Urban Models, *Journal of Political Economy*, 79, 142–153.

Olsen, E. O., 1969, A Competitive Theory of the Housing Market, *American Economic Review* 59, 612–621.

Polinsky, A. M. and S. Shavel, 1976, Amenities and Property Values in a Model of an Urban Area, *Journal of Public Economics*, 5. 119–129.

Rothenberg, J., 1973, Urban Renewal Programs, in R. Dorfman (ed.), *Measuring Benefits of Government Investments*, Washington D. C., The Bookings Institute, 292–341.

Schali, L. D., 1976, Urban Renewal Policy and Economic Efficiency, *American Economic Review*, 66, 612–628.

Solow, R. M., 1973, On Equilibrium Models of Urban Location, in M. Parkin (ed.), *Essays in Modern Economics*, New York, Barnes and Noble, 2–16.

Wright, C., 1977, Financing Public Goods and Residential Location, *Urban Studies*, 14, 51–58.

GEOGRAPHIA POLONICA 42, 1979

SLUMS IN CAPITALIST URBAN SETTINGS: SOME INSIGHTS FROM CATASTROPHE THEORY

Dimitrios S. Dendrinos*

Graduate Program in Urban Planning, The University of Kansas, Lawrence, Kansas, USA

1. SLUMS AND THE EXISTING ECONOMIC THEORIES OF URBAN FORM

A consistent pattern has been observed in urban settings in capitalist nations: the incidence of poverty, substandard housing, and the spatial concentration of a high portion of poor into spatially clustered substandard dwelling units (Mills, 1972a, p. 167)[1]. Such concentrations have been referred to as "slums". The terms "poverty" and "slums", however, do not have a widely accepted specific definition. Mills (1972a) views them as a matter of degree (p. 165). There have been many attempts made to study slums, poverty and slum formation, and there is a number of theories (economic and otherwise) available, generated at various time periods and responsive to the perceptions of the current social agenda [2]. In Muth (1969) the existing economic theories of slum formation are analyzed. According to Muth these theories are based on the proposition that a decline in the demand for housing occurred in certain neighborhoods at a given time period that caused rapid deterioration of the quality level of the dwellings; closely related theories that emphasize the factors inhibiting investment in housing are also discussed. Muth finds these theories deficient (p. 125) and he proposes an alternative theory of slum formation and their spatial concentration based on a theoretical model of residential location.

This model adheres to the neoclassical microeconomic paradigm; it is found in the work of Alonso (1964) and Mills (1972b) and it is based on the Von Thünen construct of agricultural rent. According to this theory slum formation is basically the result of different bid-rent gradients by households with different income levels, and it is due to the flattening of rent-distance functions. There are certain issues of concern associated with this theory. First, a failure to detect such a theoretical cons-

* The paper was written during the summer of 1976.

[1] The paper focuses on the capitalist urban settings without attempting to examine the cases of non-capitalist settings.

[2] The reader is referred to E. S. Mills (1972a), R. Muth (1969), J. Kain (1969), J. Wilson (1966), W. Wheaton et al. (1966), O. Davis and A. Whinston (1961) for various approaches to the issue. In Mills (1972a) the issue of poverty is discussed. He indicates the existence of two distinct theoretical approaches to poverty: the "economic" one, according to which poverty is viewed as lack of access to income and assets that permit an adequate standard of living; and a "sociological-anthropological" one, according to which poverty is a way of life, that has even created its own culture (p. 139).

truct at the micro-scale would result to elimination of the validity of the specific rationale with respect to the slum incidence. In fact it is rather rare to observe all the individual agents in an urban economy operating under the specifications of the Alonso–Mills–Muth construct. Muth attempts to empirically test the model using data from the Chicago area for 1950 and 1960. The application of multiple regression analysis, as reported in Chapter 10 in Muth (1969), tends to support the hypothesis that income and prices are correlated with the slum incidence. However, regression is not the best tool for identifying cause-effect relationships; further, the independent variables and the form of the regression equation (linear, logarithmic, double logarithmic) that provide a good fit for Chicago may or may not work well for other urban settings[3].

Second, the disaggregative nature of this model implies that the urban form is the result of interrelationships among the economic agents operating in the urban setting. This is only one way, however, of approaching a macro-scale issue (urban form), i.e., from the micro-scale, and it may not be the best way. It could be argued that it amounts to the same thing as trying to derive the national unemployment rate by examining the interrelationships among all the economic agents in the nation. The difficulty of the large number of agents and interactions is overcome in the Alonso–Mills–Muth model by assuming that all households are identical in preferences.

It is here suggested that slums is a macro urban phenomenon. It exhibits a regularity beyond any specific conditions found in a specific urban area. The pattern of slum dwelling has been systematic as to rule out any claim that slums are random events or peculiarities of specific urban settings. The slum incidence has been systematic over time; over different levels of development of a nation's economy; over different nations with different levels of economic development. It has been consistently present in urban areas of nations with different degrees of ethnic composition; over different city sizes; over different cities with a varying degree of fiscal viability. It is, and it has been, a dominant factor of urban form. It follows into the category of phenomena that Amson (1975) defines as "too universal, too regularly spread throughout the world and distributed through too many epochs to be entirely haphazard" (p. 177). It is a phenomenon that requires a macro urban theory of spatial regularities. Yet no researcher has attempted to derive such a theory. It may have been because the necessary theoretical structure was lacking, or because the analytical tools were not available.

It is also here suggested that the study of slums trascends the confines of any specific disciplinary paradigm, be it economic, social or other. As Mills (1972a) noted: "economic poverty is only one cause of misery, and other causes certainly ought to be studied" (p. 140).

---

[3] For example, a critical assumption in the Muth model, in order to derive a negative exponential price of housing-distance function, is that the real income compensated price elasticity of housing demand be equal to −1, (p. 72). This may be true for Chicago in 1960, but there is no guarantee that it may be true for all metropolitan areas, during different time periods.

2. BEHAVIORAL DEDUCTION AND REDUCTIONISM

In spite of the risk of overwhelming the topic of slums by a general reference to the behavioral theories, the following discussion is necessary in order to understand the inherent weaknesses in the available theories of slums. The Von Thunen based Alonso–Mills–Muth model is a behavioral construct in the neoclassical microeconomic tradition. This and its derivative models have tried to replicate the urban macro scale from the bottom-up: they have tried to replicate the behavior of the macro unit (the form of the urban setting) from the behavior of its micro units (households, firms, etc.). Such an approach may have been more appropriate to phenomena that have been shown to be intrinsic to a part of an urban area and not so much to phenomena that have been shown to exhibit a regularity which goes beyond the interrelationships observed among agents operating within the context of a specific urban setting.

This defect limits the behavioral theories' capacity to deal with observed urban macro phenomena since it fails to guarantee the reasonableness of their results. This argument needs elaboration in view of the fact that the counter paradigm (the structuralist construct to be presented later) builds upon this weakness. The generalized pattern involved in behavioral theory construction and testing in urban economics is as follows:

— from a set of plausible assumptions regarding the real world, a researcher selects a subset. Such a subset consists of a number of parameters and variables that are considered by the researcher to be pertinent to the analysis, as well as a number of relationships among these variables and parameters (functions). Such a formulation is thought of as a "reasonable" abstraction of reality; R. Solow refers to such abstractions as "parables".

— Next, a set of deductive statements (theorems and/or corrolaries) are proven, through mathematical transformations from the original set of assumptions.

— The above step results in the derivation of a final set of (not further meaningfully deductible) relationships among the initial variables and parameters. Such a final set relates each of the original variables to the given parameters (assuming that the original subset of assumptions imply a well defined problem). This is referred to as the "solution" to the original problem under investigation or, the derivation of the functions that describe the phenomenon the researcher intended to examine.

The terms "variable" and "parameter" are used here to classify variables into two categories: endogenously derived and exogenously prespecified ones. The analysis of the slum incidence, to the extent that it adheres to the microeconomic paradigm, closely follows the above scheme. The first step consists of identifying the micro behavior of individuals and firms (in terms of their utility function and their resource constraints), as well as the objective function of a (possible) public sector and its constraints. This step is the setting of some behavioral assumptions that regulate the disaggregative performance of the system. Then, following a number of mathematical transformations (presented as a simulation of the actual market or quasi-market behavior), a final set of relationships is derived. This final set of functions is claimed to contain the ranges of the behavioral variables that can produce the actual configuration of the urban setting (together with a large number of other possible configurations). These configurations are obtained when the parameters are assigned different values. The final set of functions is tautological to the original one since it has been derived in a purely deductive way.

The process outlined is formally a mapping of a given set of assumptions (an abstraction of reality) to a specific functional form capable of reproducing a set of possible urban configurations depending on the specific values of the parameters included in the final functions. This is a one-to-one correspondence if the solution is unique, or a one-to-many correspondence if the solution is not unique. Such a mapping is a one directional mapping from the abstraction space towards the solution space.

There is a calibration (feedback) process involved following these steps, and the solution set (comprised of all possible urban configurations) is further reduced through a selection procedure involving the values of the given parameters. This procedure involves mathematical transformations by which a specific set of values is assigned to the parameters. The selection procedure is based on the criterion of reasonableness (i.e., how close the values of the variables come to the actual ones for a given set of parameter values). The selected set of parameter values produces values of the variables closer to the actual ones than any other set of parameter values. If the closest configuration contained in the final set is not close enough to the observed values of the behavioral variables describing the urban configuration, then the search for a new point in the abstraction space starts again, and the above process repeats itself.

There are a number of discomforting issues involved with such a process. In the general case the final set of functions may be impossible to obtain in closed form, given the present level of development of the mathematical techniques used. Then, the final set of functions containing the urban configuration may not be "clear enough" to compare with the observed system performance. In such a case, it may be possible to obtain the solution from computer performed approximations only if the search costs allow it. This constraint considerably limits the researchers' capacity to focus and continue searching.

The parameter values resulting from the calibration process may be unobtainable in closed form, or may not be unique. This is a case of degeneracy (infinite number of solutions), or a case of multiple solutions. Such a one-to-many correspondence is not very useful for research purposes. It presents problems is search during the calibration phase, and it does not allow for validation of the original hypothesis. However, the most discomforting element is the direction of the correspondence (mapping). Since the direction is from the assumptions towards the final solution there is no guarantee that the resulted urban configuration will be close enough to the observed values. The modus operandi of researchers using the neoclassical paradigm (as applied to urban settings) has been a constant effort to implement this one-to-one correspondence, hoping that the results will fare well.

A second aspect related to the existing behavioral theories applicable to urban structure has been the fact that such theories have been compartmentalized, reductionist in Amson's (1975) term, since they tend to adhere to paradigms belonging to one discipline, in this case economics. The reasons for such reductionism are outside the scope of this paper, and will not be elaborated upon. The following portion of the paper attempts to set a macro urban theory which has as a starting point the urban macro structure and it is more holistic than the current analyses.

3. TOWARDS A MACRO URBAN THEORY; URBAN MORPHOLOGY AND SLUMS

The urban setting is the result of spatial agglomeration of people and capital stock at different locations (areal units within urban settings), at different time periods. The state (behavior) of the population and the capital stock at each areal unit is defined over the values of a number of specific aggregate behavioral variables *m*. The values of these *m* variables describe the behavior of the areal unit. The state of an areal unit is determined by a number of specific independent control variables *n*, that operate in the context of a capitalist urban economy. An areal unit is a point in the $R^{m \times n}$ space at a given time period. A sequence of time periods of an areal unit defines an intertemporal path; the sequence of all points representing the areal units of a specific urban area defines an areal path, identifying the urban setting in the $R^{m \times n}$ space at a given time period.

The set of all intertemporal paths of all areal paths define the urban morphology. Thus, an urban morphology type is a specific manifold representing a specific function relating the *m* behavioral variables to the *n* control variables applicable to capitalist urban settings. This paper is a study in urban morphology; it is here suggested that the more one knows about the type of the urban morphology, the better one understands the nature of the urban settings.

An intrinsic element of the capitalist urban morphology is the element of slums. Before entering into the discussion of slums, the behavior and control variables operating in the capitalist urban morphology will be presented.

In the literature stemming from the Alonso–Mills–Muth model various variables were identified as descriptors of the state (behavior) of an urban area and its areal units (in the standard model the areal units are suburban rings at different distances from the center). Such variables associated with the supply of the capital stock were: the price of housing, the quantity demanded and supplied, the quality level of the stock under market conditions, etc.; on the population side variables descibring the bevavior of the households residing in the dwellings available were: the utility level (in the case of a closed urban area) of the residents, expressed as a cardinal index, the number of individuals residing in the areal unit, etc. Control variables (parameters) encountered were: the wage rate, the total urban population (in the case of a closed urban area), the opportunity cost of the land, the market rate of interest, the transportation rate, etc. The rest of the literature already cited in part 1, draws from these and other variables like: racial discrimination, occupancy type, capital construction and upgrading costs, vintage of the dwellings, capital depreciation rate, education level of occupants, etc. The role and importance of these variables (or other variables that are not part of existing analyses) in the study of the urban morphology in capitalist settings is not clear entirely. Some of the above variables are obviously interrelated and some are independent. The cause and the effect relationship is not well established.

A documentation of the urban morphology, the function among the generic variables of urban morphology, is a matter of speculation since the data necessary for the establishment of the urban morphology along various variables are lacking. This paper puts forward a proposition for the generation of such a morphology and thus it needs empirical verification. This is a starting point which, as in the case of all theories, needs further refinement.

Within a market context the single most important variable determinant of housing demand is income. Income, due to its non-uniform distribution in capitalist settings, has been considered by many analysts to be an important factor of urban structure. Little needs to be added on this variable here. On the supply side, the willingness to invest by the private sector depends on the internal rate of return of the investment at a particular location, at a given time period; such rate of return is computed based on foresight regarding local demand and supply conditions in the future.

A third control variable is associated with the existence of a public sector and its implicit preference function on the supply of public goods and/or public housing; this variable is the social rate of discount and its values need not be the same during a given time period over all urban areas or over all areal units of an urban area. Finally, a fourth variable determinant of urban form and associated with the zoning policies of an urban government is the population to capital stock density, at a specific location at a given time period.

These four control variables, it is proposed here, are the key determinants of behavior of urban settings and of their areal units in a capitalist mode. The behavior of these units can be recorded along the following two dimensions: the quality level of the housing stock, and the (cardinal) utility level enjoyed by the residents in an areal unit. The capitalist urban morphology is a specific functional type relating these six variables. The urban morphology type to be generated by the manifold of the function relating these six variables has to satisfy the following criteria of consistency with empirical facts: first, it must reproduce specific type of discontinuous behavior observed in the behavior space of the morphology. A small move along an intertemporal path (i.e., small time changes in certain urban areal units) must produce sudden discontinuous change in their behavior: these are the urban renewal, or the transitional neighborhood phenomena observed in certain urban areas in capitalist settings. In the same manner, a small move along an areal path (i.e., going from an areal unit of a given urban area to a neighboring one, in the same time period) may result to sudden changes in the observed quality level of the housing stock.

Second, the urban morphology generated by the function relating these six variables must be able to reproduce the slum incidence. Slums are viewed here as a mode of living in capitalist urban settings competing with other nonslum modes for urban space, and the functional form of the urban morphology has to be able to reproduce areas of conflict of modes.

The paper now turns ot the search for the functional type among the six variables identified earlier, that would reproduce these discontinuous cases and the conflict of modes areas in the unfolding of the morphology, satisfying thus the reasonableness of the results criterion, by starting from the empirical phenomena recorded in capitalist settings.

4. A STRUCTURAL ANALYSIS

An attempt to study the form of systems exhibiting discontinuous behavior has been the work by Thom (1975) on catastrophe theory which can be used as the basis for a structural analysis of capitalist urban settings and their slums[4]. Two basic features of this structuralist theory are of interest here. One, it asserts that if a system needs only a given (small) number of variables and (independent) control parameters to describe it, then there is only a given (small) number of ways that the system's behavior can be described; and two, it proposes that there is a given potential function corresponding to every possible morphology type that regulates the system's behavior.

The above two points imply that, first the structuralist mode is a mapping from a morphology type towards a potential function (i.e., a correspondence from an observed set of urban configurations towards a meaningful theoretical abstraction). This point makes the theory relevant for the understanding of phenomena of macro scale that exhibit a regularity transcending local conditions. Second, the functional form of the potential function is unique[5]. Each potential function corresponding to a morphology type can generate a large number of specific urban settings in a one-to-many correspondence. The theory outlined is a multiple-to-one mapping from a set of urban configurations (defining the urban morphology) towards their potential function. This correspondence makes the theory *ex ante* pertinent to the study of urban settings.

The additional element that makes catastrophe theory pertinent to the study of slums is its intrinsic ability to deal with discontinuities in the behavior of a system. The behavioral theories, due mostly to their reliance on calculus, are capable of handling only continuous behavior. Thom suggests that a basic characteristic of form is its discontinuousness. Catastrophe theory defines these discontinuities as structural singularities. Such singularities occur in the parts of the morphology where the value of a behavioral variable suddenly vanishes (i.e., it obtains imaginary values), or where there is a sudden transition in a conflict domain among possible states (values) that the behavioral variables may obtain depending on the dynamic path of the system. Discontinuities thus are sudden transitions from observable

[4] Catastrophe theory is young in its development and it does not have yet many disciples in the field of urban planning. This is true also because of the radical departure (discontinuity?) that is involved in passing from a given mode of scientific inquiry based on the behavioral deduction-reduction to the structural one.

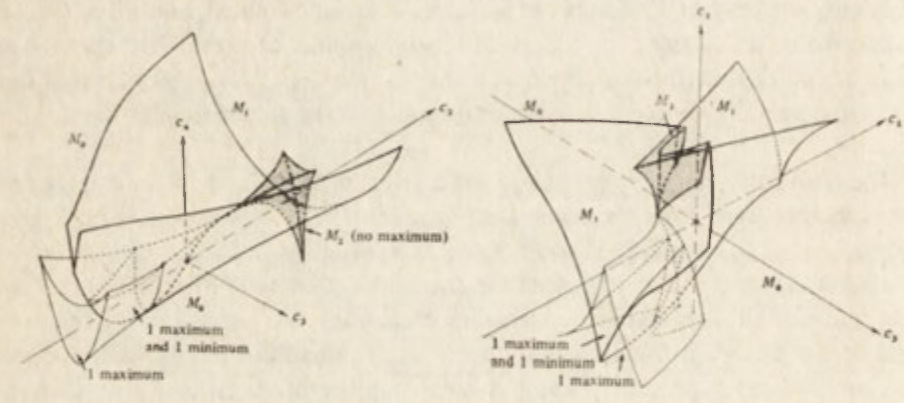
[5] This needs qualification: one of Thom's results is that the unfolding of the structural singularity of a specific catastrophe type does not basically change if higher order terms of the behavioral variables are introduced into the potential function. What really matters is the number of independent control variables involved in the analysis.

behavior to non-observable, or from one state (mode) to another. The formal steps involved in using catastrophe theory as a tool of urban morphology are quite a departure from the steps identified earlier in the case of a behavioral theory:

— first, the pertinent (small) number of independent parameters, and the number of variables of urban morphology are identified, as well as the structural singularity of the system is established;

— then, based on the number of parameters and variables, a catastrophe type (morphology) is identified from the set of elementary catastrophes. The specific unfolding of the catastrophe is matched against the observed urban morphology;

— the potential function of the catastrophe identified is then arrived at and interpreted.



(a) (b)

Fig. 1. The urban morphology as a mushroom catastrophe. Some sections of the bifurcation set (a): $C_1 = 0$; (b): λC_4 positive. M_0 is the region of non-observed behavior. M_1 is the region of one local regime. M_2 is the conflict set.
Source: J. Amson

In the above scheme the existing urban configuration is the starting point in a search for the fittest theory. There are questions involved with such an approach as well: how can one be sure that all pertinent parameters were selected? what is the suitable transformation of the original parameters and variables selected, since catastrophe theory operates with topological transformations? how is the matching between the observed urban morphology and the theoretical construct to occur? The data necessary to answer these questions are lacking and the techniques appropriate for calibration have not yet been well developed.

In section 3 the critical control and behavior variables of a capitalist urban morphology were identified, four control variables and two behavior ones. According to Thom (1975) a system defined by four control variables and two behavior ones and exhibiting discontinuities in the behavior space can only behave as described by the mushroom (parabolic umbilic) catastrophe, shown in Figure 1. The parts (a) and (b) correspond to the sections of the bifurcation set (i.e., the projection of the unfolding of the urban morphology onto the control space) in the $(c_1 c_2 c_3)$ and $(c_2 c_3 c_4)$ dimensions. It is suggested that if the urban morphology of capitalist settings were to be recorded along the six dimensions mentioned, the bifurcation set would resemble these two sections of the mushroom catastrophe[6]. This is shown below.

The mushroom catastrophe has been studied by Woodcock and Poston (1974) and Godwing (1971), and a good description of it is found in Amson (1975). The unfolding of the singularity is derived by the stationary points of a potential function (which is discussed in section 5) of the form:

$$\min E = x_1^2 x_2 + x_2^4 + c_1 x_1^2 + c_2 x_2^2 - c_3 x_1 - c_4 x_2, \tag{1}$$

where x_1 and x_2 are the transformed behavioral variables and the $c_j (j = 1, 4)$ are the transformed control parameters. Such transformations may be indices of proportional deviation from a mean value among the observed values of these variables in the areal units of the urban settings.

The potential function achieves a minimum when the following first order conditions hold:

$$2x_1 x_2 + 2c_1 x_1 - c_3 = 0, \tag{2.1}$$

$$x_1^2 + 4x_2^3 - 2c_2 x_2 - c_4 = 0. \tag{2.2}$$

The degenerate set, the condition that produces the singularities, is derived when the Hessian is zero:

$$(x_2 + c_1)(6x_2^2 + c_2) - x_1^2 = 0. \tag{3}$$

Variations of the parabolic umbilic catastrophe (its compactified and weakly compactified forms:

$$\min E = x_1^2 x_2 + x_1^4 + x_2^4 + c_1 x_1^2 + c_2 x_2^2 - c_3 x_1 - c_4 x_2,$$

$$\min E = x_1^2 x_2 + x_1^4/4 + x_2^4/4 + c_1 x_1^2 + c_2 x_2^2 - c_3 x_1 - c_4 x_2,$$

correspondingly) can be found in Woodcock and Poston (1974).

In Figure 1 the form of the bifurcation set in three dimensions indicates the location of the set M_2 (areas in which the behavioral variables may obtain more than one value and where by a small change in the values of the parameters the bahavior of the system may suddenly change mode). The set M_2 define areas in the urban morphology in which in a short time span sudden transitions may occur in areal units of an urban setting, either from high quality level capital stock (or high level of utility households) to low quality stock (or vacancy), or vice versa. It may also

[6] The proper catastrophes to be used as tools of urban analysis may still have to be developed if the number of critical independent control variables is higher than four. Only the seven elementary catastrophes have been disseminated beyond the laboratories of a few analysts. For an excellent portrait of these seven catastrophes, and some insight with reference to urban settings, the reader is referred to Amson (1975) particularly sections 5, 6. These seven catastrophes occur only when the number of control variables does not exceed four. Various aspects of urban analysis may require structural singularities involving more than just four parameters.

indicate that under the same control conditions two areal units at this part of the urban morphology may exhibit different behavior (mode). In section (*a*) the set M_2 is shown in the space of the control variables: income (c_1), internal rate of return (c_2) and the social rate of discount (c_3). The set M_2 occurs at positive values of c_2 (i.e., in areal units with relatively high internal rate of return) and negative c_1 (low level of income). In section (b) the set M_2 is shown in the space of the control variables: rate of return c_2, social rate of discount c_3 and the population to capital density c_4. The set M_2 appears when c_4 is positive (i.e., in areal units with high density) and the rate of return of private capital investment is high. The area M_2 defines points on the bifurcation set that correspond to minima (two) of the potential function.

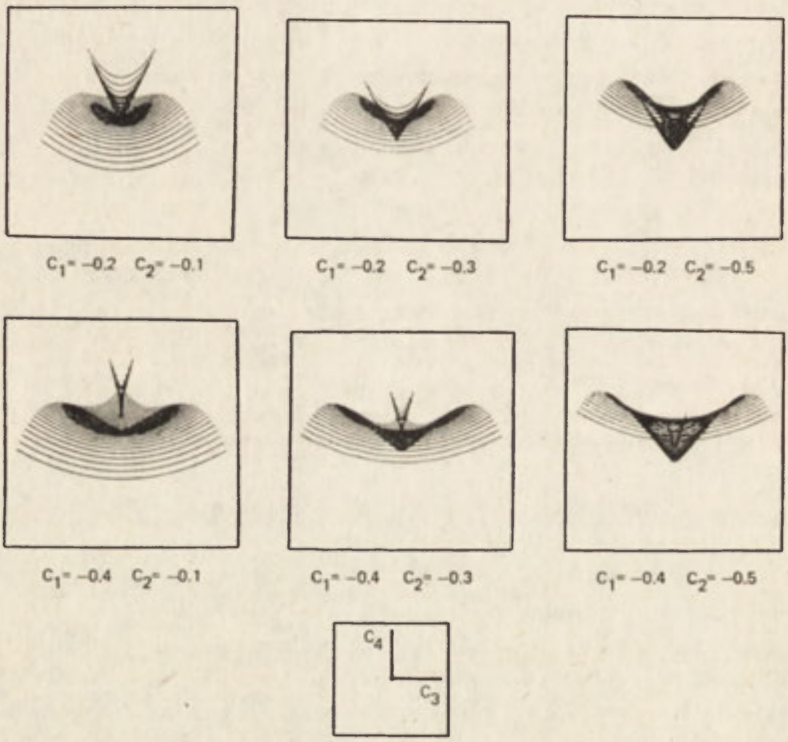


Fig. 2. Various sections of the bifurcation set of the parabolic umbilic catastrophe
Source: A.E.R. Woodcock and T. Poston

Another area relevant to the analysis is the part of the bifurcation set where one minimum and one maximum, or the area in which only one maximum may occur of the potential function. Points that maximize the value of the potential function are not stable points. It is suggested here that contrary to physical systems in the capitalist urban morphology such unstable points are attainable, and that unstable behavior is an inherent pattern of such morphology. The set M_2 corresponds to *stable* slum mode, whereas areas of maximum of the potential function corresponds to *unstable* slums. An unstable slum is the one that: a very small perturbation of the controls in an urban areal unit will lead to slum disappearence and the behavior of the unit to become stable, or to vanish (i.e., the areal unit to be vacated). Contrary to such unstable slums are the other type of (stable) slums that do not necessarily disappear with infinitesimal changes of the values of the control variables.

The part of the bifurcation set that exhibits possible unstable slum behavior is located in the negative values of the rate of return, sections (a) and (b) of Figure 1. It also occurs, as it can be seen from section (b), in negative values of the income variable, i.e., in areal units with low income residents.

A closer look at the form of the bifurcation set can be obtained from Figure 2, taken from Woodcock and Poston (1974). The social rate of discount (c_3) and the density (c_4) indices are plotted as the income (c_1) and the rate of return (c_2) indices vary. They depict the areas where a maximum and a minimum, or only a maximum can occur in the bifurcation set. Here the income index is negative (low income), as is the rate of return index. The area inside the cusp-like form has one maximum and one minimum, the area outside has one maximum. Moving along any row the impact on the bifurcation set due to increases in the rate of return is recorded. As it becomes more and more unattractive to invest in housing, in low income neighborhoods, the area where stable non-slum and/or unstable behavior can occur penetrates into the area where either only unstable or not observable behavior can occur.

Moving along in any column, the impact of decreases in income is recorded. Such decreases result in increased ranges of density along which the urban morphology experiences slum (of the unstable type) or non-slum (stable) modes of behavior. All sections of the bifurcation set are symmetrical along the density axis. This implies that the sign of the density index has no effect on the singularities of the urban morphology, only its absolute magnitude.

The similarity of this qualitative analysis to the observed urban settings makes this catastrophe type relevant for the study of the capitalist urban morphology. A point should be made that the actual urban settings (the areal paths) and their dynamic paths are represented, as indicated, by paths in the bifurcation set. The specific functions that produce such paths are not yet established. *As a result of such functions the density of such paths in the bifurcation set is not necessarily uniform throughout the surface of the bifurcation set.* In fact there may be some areas in the surface of the urban morphology not yet observed, and therefore, paths of existing areal units do not pass through such areas. On the other hand, there might be other parts often encountered by paths of existing urban settings. What creates such density distribution of paths in the bifurcation set of capitalist urban morphology is a question of significant interest, but one not explored in the literature. This paper will not address it, except to state that the issue is a fruitful area for further research.

The two subareas discussed above in the bifurcation set are not the only ones of interest. There are other aspects of the bifurcation set that were not discussed here, for example the ghost-town phenomenon identified by Amson (1974), or the element of morphogenesis identified by Thom (1975).

5. SLUMS IN URBAN MORPHOLOGY: THE POTENTIAL FUNCTION

The question now turns to the final step, the identification of the nature of the potential function regulating the areal units of the urban morphology, that reproduces such structural singularities. The different economic agents of urban systems operate under some objectives (individual) and/or collective and under different constraints. Economists refer to such functions as utility and profit (or cost) functions at the micro level, and as social welfare functions at the macro level. The contribution of catastrophe theory is that it proposes such a regulating function at the macro level, for each type of catastrophe, without entering the discussion of how such a function has been derived from the micro scale. Instead, the micro scale is examined at its aggregate behavior from above, given the general form of a potential function that regulates the system and that would reproduce its observable behavior. In the case of the urban morphology that has to meet the criteria of consistency outlined, the objective function is the one corresponding to the mushroom catastrophe identified in (1). As it was demostrated it reproduces the slum incidence in the urban morphology.

The interpretation of the potential function may be that it represents a minimizing social cost formulation at each areal unit. Another approach to interpreting the potential function is to consider the dual catastrophe (false mushroom) and interpret its maximizing potential as the maximization of an aggregate social welfare function on *each areal unit.*

The above interpretation of the potential function now sheds some more light onto the nature of the slum incidence, as it relates to the two areas of the bifurcation set discussed. In M_2 it is shown that the slum mode is stable and thus *it corresponds to points where the aggregate social cost function in these areal units achieves a minimum through a stable slum mode of dwelling.* However, the area of the bifurcation set where the incidence of slum occurs as unstable behavior is the area *that corresponds to points where the aggregate social cost function attains a maximum through such unstable slum dwelling.*

6. CONCLUSIONS

A theory of urban morphology in the context of a capitalist setting was presented and its aspects as they pertain to the existence of slums and their nature were analyzed. Slums here were viewed as a mode of living in a capitalist urban economy which, together with other non-slum modes, competes for urban space. It was shown that in such a setting there are two types of slums, stable and unstable, that can be studied within the framework of catastrophe theory. These slums as well as various types of discontinuous behavior in urban areas correspond to specific structural singularities associated with the mushroom catastrophe. It was suggested that this specific catastrophe closely approximates the capitalist urban morphology.

It was proposed that catastrophe theory is a radically different approach to urban studies from the behaviorally based deductive-reductional mode of inquiry. The inherent difficulties involved in the standard model of analysis of urban economics were discussed and the advantages of the structuralist approach were indicated.

REFERENCES

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

Amson, J., 1974, Equilibrium and Catastrophic Models of Urban Growth, in *London Papers in Regional Science, 4. Space-Time Concepts in Urban and Regional Models*, ed. E. L. Cripps, Pion, 108–128.

Amson, J., 1975, Catastrophe Theory: A Contribution to the Study of Urban Problems?, *Environment and Planning B*, 2, 177–221.

Davis, O. and A. Whinston, 1961, The Economics of Urban Renewal, *Journal of Law and Contemporary Problems*, Winter, 26, 105–117.

Godwin, A., 1971, Three-dimensional Pictures of Thom's Parabolic Umbilic, Institut des Hautes Études Scientifiques, *Publications Mathématiques*, 40, Paris.

Kain, J. (ed.), 1969, *Race and Poverty: the Economics of Discrimination*, Prentice-Hall.

Mills, E. S., 1972a, *Urban Economics*, Scott and Foresman.

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

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

Thom, R., 1975, *Structural Stability and Morphogenesis*, W. A. Benjamin.

Wheaton, W., et al. (eds.) 1966, *Urban Housing*, The Free Press.

Wilson, J. (ed.), 1966, *Urban Renewal: the Record and the Controversy*, MIT Press.

Woodcock, A. E. R. and T. Poston, 1974, A Geometrical Study of the Elementary Catastrophes, in *Lecture Notes in Mathematics*, Springer-Verlag.

GEOGRAPHIA POLONICA 42, 1979

URBAN GROWTH WITH AGGLOMERATION ECONOMIES AND DISECONOMIES

Burkhard von Rabenau

Department of City and Regional Planning, The Ohio State University, Columbus, Ohio, USA

I. INTRODUCTION

There exist few analytical models of urban growth. This seems to stem less from a lack of interest in urban growth but from the difficulty of modelling the elements that give an economy its urban character. Not unlike regional or small national economies, cities are open; factors and products move relatively freely between urban area and nation. Labor may in-and outmigrate and the urban labor force may grow at a rate different from its local natural rate; capital growth is not restricted by savings generated within the urban area; and much of the urban output may be exported to the rest of the nation.

A characteristic of the urban economy which is not shared by open economies in general, is the spatial concentration of economic activities. Indeed, Mills (1972) makes this the defining characteristic of cities. Such concentration exists because of locational advantages or increasing returns to scale. Similarly, Arrow (1973) bases the existence of cities on agglomeration effects reflecting increasing returns and externalities available to spatially close firms. It is, therefore, desirable to allow in a model of urban growth for agglomeration economies and diseconomies and/or a land market with spatial dimensions.

In the literature there exist two explanations of the growth/decay of an open — urban or regional — economy. In export base theory the export demand by the rest of the nation for goods produced in the urban area determines the urban size via an export multiplier. This theory in itself is entirely static. An urban growth path may be obtained, however, from a given path of export demand. For this, one need only assume that the urban labor and capital stock adjust at any moment through appropriate in-or outflows so as to satisfy the given export and resulting domestic demand. This approach is taken in applied work and simple models of urban growth, for example Czamanski (1964). More complex models are based on the observation that urban areas typically cannot adjust instantaneously to a size required by a given export demand. Delays may result from the immobility of labor considered by Niedercorn and Kain (1963), or from that of capital, see Rabenau (1976). In either case an urban growth path is obtained from difference or differential equations that describe factor changes over time.

An alternative to export base theory has been developed by Borts (1960), Siebert (1969), and others. Here differences in urban or regional growth rates are explained as a consequence of factor price differentials. Factors move between urban area and nation in the direction of higher returns. This lowers return differentials below a level that would have existed in the absence of such factor movements. Analytical models along this line have been formulated by Hanson and Rabenau (1976), and Rabenau (1973, 1974). The model presented here extends this work.

Section 2 of this paper describes an urban area which produces a single multi-purpose output with two factors. The growth of the two factors is described by two differential equations. It is the sum of urban internal growth and growth due to movements between urban area and nation. The analysis of the urban growth path is simplified by the assumption that the urban area is small relative to the nation; hence factor returns in the nation are stationary and unaffected by the factor movements. Section 3 describes the growth path for the case of constant returns. In Sections 4 and 5 agglomeration effects are considered. For ease of exposition the cases of agglomeration economies and diseconomies are examined separately in Section 4. Section 5 analyzes the probably common and most relevant case that an urban area exhibits agglomeration economies for small and diseconomies for large sizes. It should be considered the most realistic and interesting of all cases. Conclusions follow in Section 6.

2. ASSUMPTIONS

An urban economy is considered producing a single output $Q(t)$ with capital $K(t)$ and labor $L(t)$ via a constant return Cobb-Douglas production function,

$$Q(t) = A[K(t)]^{\alpha}[L(t)]^{\beta}, \ \alpha+\beta = 1, \tag{1}$$

where A is the urban productivity index. The urban economy is located within a national economy, but is small enough not to influence the wage rate w_N and rental rate r_N prevailing in the rest of the nation.[1] These national rates are stationary and may differ from the respective urban rates $w(t)$ and $r(t)$.

Capital and labor grow according to two differential equations.

$$\dot{L}(t) = nL(t)+M(t), \tag{2}$$

$$\dot{K}(t) = sQ(t)+E(t) \tag{3}$$

with initial conditions $K(0) = K_0$ and $L(0) = L_0$. Hence, growth of the urban labor force is the sum of natural growth at rate n and (net-) migration $M(t)$ between city and rest of the nation. Similarly, capital growth is the sum of urban savings $sQ(t), 0 < s < 1$, and (net-) capital flows $E(t)$ between city and nation. Without

[1] This assumption must hold for any time t. Hence, an urban area cannot forever grow at a higher than national growth rate without eventually violating the assumption that it be small relative to the nation. If this occurs, the model does not apply. An analogous assumption is made in the analysis of the growth of firms, e.g. Treadway (1969). Such models similarly do not apply to, say General Motors.

$M(t)$ and $E(t)$ the urban economy becomes identical to the closed economy considered by Solow (1956).

Factor flows $M(t)$ and $E(t)$ depend on two variables, the factor price differential between nation and urban area and the size of each factor's stock in the urban area. In particular, following one of two formulations by Siebert (1969), it is assumed that the net migration rates of labor $M(t)/L(t)$ and capital $E(t)/K(t)$ are proportional to the wage and rate of return differentials, respectively, between urban area and nation, so

$$M(t) = q\big(w(t)-w_N\big)L(t), \tag{4}$$

$$E(t) = v\big(r(t)-r_N\big)K(t), \tag{5}$$

where q and v are positive constants, called mobility coefficients by Siebert.

If transport cost were zero and information perfect, one would expect factor flows to be infinite whenever factor prices differ between nation and urban area. Instead (4) and (5) imply, more realistically, that factor flows are finite and increasing with the factor price differential as well as with the amount of the factor already in place in the urban area. To justify this, one may assume that the unit transportation cost increases with total factor shipments made at t. Then these shipments will be finite and just equal to the amount at which expected unit gains equal unit transport cost. To the extent that expected gains from the shipment of a factor unit are positively related to the factor price differential at t, total shipments will increase with this differential as suggested by (4) and (5).[2]

(4) and (5) may also be justified by the presence of imperfect information. For example, if $w(t)-w_N < Q$, workers will search for work outside town. Search efforts and the probability of finding a job will probably increase as $w(t)-w_N$ decreases. Mutliplying the probability of finding a job by the number of people who search, that is $L(t)$, yields the flow of movers as suggested in (4). If $w(t)-w_N > 0$ similar forces operate. In particular, people moving to the urban area depend on information and help from those who are already there. The extent to which people in the rest of the nation are aware of the price differential also depends on $L(t)$. So does the ease with which any given number of people can be integrated and absorbed into the urban area. Hence, one would expect immigration to be positively related to the urban labor force. Similar arguments can be made for capital flows.

To complete the model assume that factors are paid the value of their marginal products. Then if the price of output is unity, differential equations (2) and (3) are given by

$$\dot{L}(t) = nL(t)+q\big(\beta A[K(t)]^{\alpha}[L(t)]^{\beta-1}-w_N\big)L(t), \tag{6}$$

$$\dot{K}(t) = sQ(t)+v\big(\alpha A[K(t)]^{\alpha-1}[L(t)]^{\beta}-r_N\big)K(t). \tag{7}$$

In (6) it is assumed that $n-qw_N < 0$ implying that outmigration from the urban area exceeds natural growth if the urban wage rate is zero.

---

[2]As an example, let the unit transport cost of capital be $T = [bE(t)]/K(t)$. Assume capital owners expect the rate of return differential to be stationary; thus the present value of gains from shipment expected over the infinite life of a capital unit are $G = \big(r(t)-r_N\big)/g$, where g is the discount rate. Equating T and G and solving for $E(t)$ yields (5), letting $1/(bg) = q$.

3. SOLUTION WITH CONSTANT RETURNS TO SCALE

The capital labor ratio $k = K/L$ chagnes over time according to

$$\dot{k}(t) = \left\{\frac{\dot{K}(t)}{K(t)} - \frac{\dot{L}(t)}{L(t)}\right\} k(t). \tag{8}$$

Expressing $\dot{K}/K$ and $\dot{L}/L$ as functions of k, using (6) and (7), yields

$$\dot{k}(t) = \{sA[k(t)]^{-\beta} + v(\alpha A[k(t)]^{-\beta} - r_N)\}k(t) - \{n + q(\beta A[k(t)]^{\alpha} - w_N)\}k(t). \tag{9}$$

As a function of k, $(\dot{K}/K)(k)$ is strictly decreasing with $(\dot{K}/K)(0) = \infty$ and $(\dot{K}/K)(\infty) = -vr_N < 0$ while $(\dot{L}/L)(k)$ is strictly increasing with $(\dot{L}/L)(0) = n - qw_N < < 0$ and $(\dot{L}/L)(\infty) > 0$. Hence, a unique k, say k^*, exists such that $\dot{k} = 0$. k^* is stable since $\dot{k} > 0$ for $k < k^*$ but $\dot{k} < 0$ for $k > k^*$. This is illustrated in Figure 1.

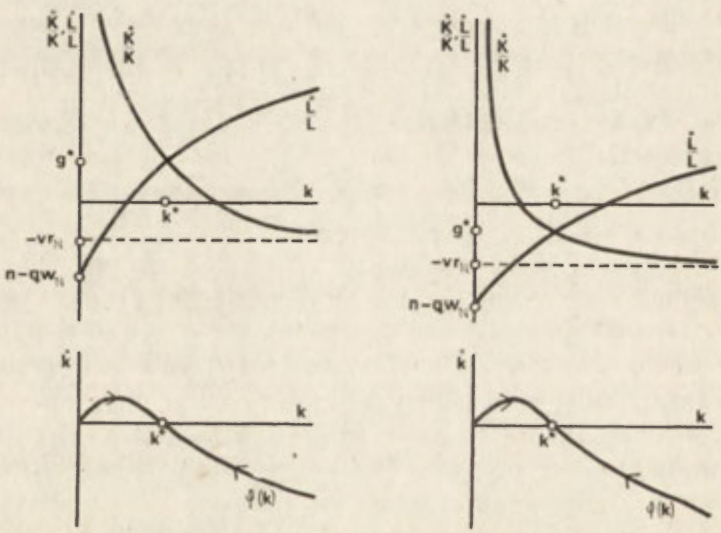


Fig. 1. Phase diagram; growth rates of K, L

At k^*, urban capital and labor grow at a constant rate, say g^*. This growth rate may be positive, zero, or negative and the urban area may grow or decay. Also, at k^* the urban and national rates of return or wage rates are not identical. Hence, along the stable growth or decay path there may be continued in-or outmigration of factors. These results are now summarized.

Proposition 1. Consider an urban area producing a single output Q according to (1). Labor L and capital K grow according to (6) and (7). Then there exists a unique stable equilibrium capital/labor ratio $k^* > 0$ such that for any initial (K_0, L_0) with $k_0 = K_0/L_0$:

$$\lim_{t\to\infty}(k(t) = k^*, \tag{10}$$

$$\lim_{t\to\infty} w(t) = w^* = \beta(k^*)^{\alpha}, \tag{11}$$

$$\lim_{t\to\infty} r(t) = r^* = \alpha(k^*)^{-\beta}, \tag{12}$$

where w^* and r^* may differ from w_N and r_N. The limiting growth rate g^*

$$\lim_{t\to\infty}\frac{\dot{Q}(t)}{Q(t)} = \lim_{t\to\infty}\frac{\dot{K}(t)}{K(t)} = \lim_{t\to\infty}\frac{\dot{L}(t)}{L(t)} = g^* \tag{13}$$

may be positive and the city grows forever. It may be zero and the city ultimately stagnates. Or it may be negative, in which case

$$\lim_{t\to\infty} K(t) = \lim_{t\to\infty} L(t) = 0 \tag{14}$$

and the city is dying.

4. AGGLOMERATION EFFECTS

The existence of cities is frequently thought to be the result of increasing returns to scale, externalities, or more generally of agglomeration effects, see Arrow (1973). It is, therefore, important that an urban growth model consider these effects. This is done here assuming that there are variable returns to the scale of the urban economy but constant returns to scale for each firm. Hence, different from the usual increasing returns to scale, total factor payments exhaust total product, if firms pay factors their marginal value product. Externalities exist in the sense that the expansion of an individual firm in- or decreases the productivity of all other firms in the urban area. If the productivity is increased, we speak of agglomeration economies and if it is decreased of agglomeration diseconomies.

Production functions of a similar nature have been studied by Chipman (1970), and Herberg and Kemp (1969), who consider the case of variable returns to the scale of an industry in the presence of constant returns to the firm. For an urban area such a production function has been considered by Varaiya (1973), who ties increasing returns to the size of the urban labor force. His approach is followed here and production function (1) is modified to make productivity index A a function of L, say $A(L)$. If $A'(L) > 0$ there are agglomeration economies and if $A'(L) < 0$ there are agglomeration diseconomies.

An individual firm i hires capital K_i and labor L_i to produce Q_i. It considers $A(L)$ a parameter and thus faces a constant returns to scale production function.

$$Q_i = A(L)K_i^{\alpha}L_i^{\beta}, \quad \alpha+\beta = 1.$$

Here L, K and Q are the summation over all i of L_i, K_i and Q_i, respectively. If firms pay factors their marginal value product then all firms and thus the urban area will produce with the same capital labor ratio $K_i/L_i = K/L$, Because there are constant returns to scale in K_i and L_i it is possible to obtain the aggregate production function (1)

$$Q = A(L)K^{\alpha}L^{\beta},$$

where $A(L)$ replaces A.

The total marginal product of labor dQ/dL is the sum of an external effect

$(\partial Q/\partial A)\ A'(L)$ which can be positive or negative depending on the sign of $A(L)$, and a direct effect $\partial Q/\partial L > 0$. It seems reasonable to assume that this total marginal product is positive and diminishing. In the following $A(L)$ is assumed to be of the form

$$A(L) = L^{\gamma} \tag{15}$$

and, therefore, γ must satisfy

$$0 < \beta+\gamma < 1. \tag{16}$$

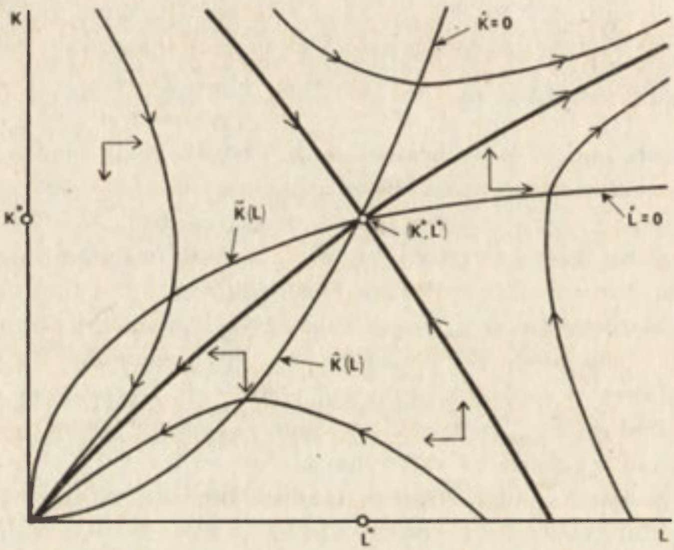


Fig. 2. Phase diagram: agglomeration economies

To study urban growth, a phase diagram in the K–L plane is constructed as shown in Figures 2 and 3 for the cases of agglomeration economies ($\gamma > 0$) and diseconomies ($\gamma < 0$), respectively. Consider in this plane the curve along which $\dot{K} = 0$, given by $\hat{K}(L)$, and the curve for which $\dot{L} = 0$, given by $\tilde{K}(L)$,

$$\hat{K}(L) = \hat{\mu} L^{(\gamma+\beta)/\alpha}, \quad \hat{\mu} = \left(\frac{s+v\alpha}{v r_N}\right)^{1/\beta}, \tag{17}$$

$$\tilde{K}(L) = \tilde{\mu} L^{(\alpha-\gamma)/\beta}, \quad \tilde{\mu} = \left(\frac{q w_N - n}{q\beta}\right)^{1/\alpha}, \tag{18}$$

where $\hat{\mu}$ and $\tilde{\mu}$ are collections of constants. The curves given by (17) and (18) divide the K–L plane into four regions. It is easily verified that $\hat{K}(0) = \tilde{K}(0) = 0$. Moreover, for $\gamma > 0$ ($\gamma < 0$) $\hat{K}(L)$ is a convex (concave) function with $\hat{K}'(0) = 0$ ($\hat{K}'(0) = \infty$) and $\tilde{K}(L)$ is a convave (convex) function with $\tilde{K}'(0) = \infty$ ($\tilde{K}'(0) = 0$). Hence, a unique equilibrium point (K^*, L^*) exists with $K^* > 0$, $L^* > 0$. Examination of the characteristic equation of (6) and (7) at (K^*, L^*) shows that the equilibium point is a saddle-point if $\gamma > 0$ and a stable point if $\gamma < 0$.[3] Finally, since $d\dot{K}/dK < 0$ ($d\dot{L}/dL > 0$) it follows that K decreases (L increases) above and increases (decreases) below curve $\dot{K} = 0$ ($\dot{L} = 0$).

This completes the analysis of the phase diagram. Arrows in each of the regions indicate the direction a path can take. In particular, if there are agglomeration economies then the saddle-point (K^*, L^*) can be reached only along its stable branches.

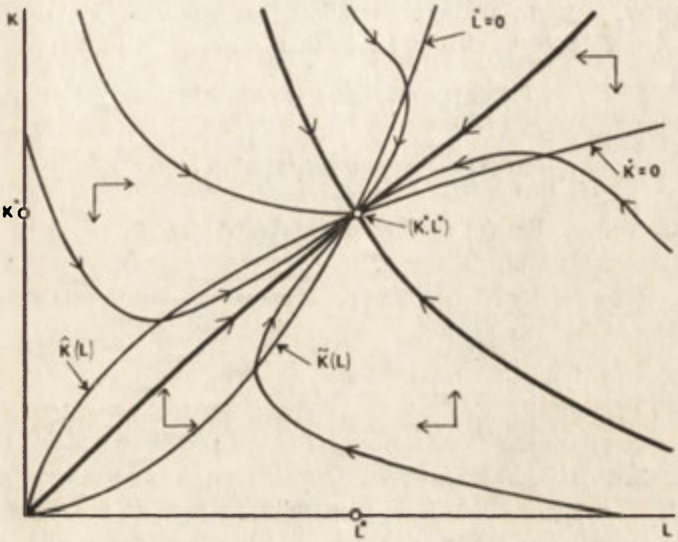


Fig. 3. Phase diagram: agglomeration diseconomies

Observe that these branches serve as a locus of minimum thresholds: from any initial (K_0, L_0) below this curve the urban area decays and the economy's path leads ultimately to $(K, L) = (0, 0)$. In order to grow the urban economy must start from an initial point (K_0, L_0) above the stable branches which serve as a takeoff stage into

[3] This result holds independently of the particular form of $A(L)$ assumed in (15). At (K^*, L^*) it holds that

$$\frac{vr_N}{a+va} = A(L)K^{\alpha-1}L^{\beta}, \quad \frac{qw_N-n}{q\beta} = A(L)K^{\alpha}L^{\beta-1}.$$

Using these expressions we obtain

$$\frac{\partial\dot{K}}{\partial K}+\frac{\partial\dot{L}}{\partial L} = -\beta vr_N+q\beta A'(L)K^{\alpha}L^{\beta}-\alpha(qw_N-n),$$

$$\frac{\partial\dot{K}}{\partial K}\frac{\partial\dot{L}}{\partial L}-\frac{\partial\dot{K}}{\partial K}\frac{\partial\dot{L}}{\partial L} = -q\beta vr_N A'(L)K^{\alpha}L^{\beta}.$$

The characteristic equation

$$\lambda^2-\left\{\frac{\partial\dot{K}}{\partial K}+\frac{\partial\dot{L}}{\partial L}\right\}\lambda+\left\{\frac{\partial\dot{K}}{\partial K}\frac{\partial\dot{L}}{\partial L}-\frac{\partial\dot{K}}{\partial K}\frac{\partial\dot{L}}{\partial L}\right\} = 0$$

has therefore two real roots which are of unequal sign if $A'(L) > 0$ and of equal and positive sign if $A'(L) < 0$ proving the assertion.

self-sustained growth. If instead there are agglomeration diseconomies, then the equilibrium nodal point will be approached from any starting point in infinite time. These results are now summarized.

Proposition 2. An urban economy produces output Q according to (1), (13) and (14). Capital and labor grow according to (4) and (5). Then a unique equilibrium point (K^*, L^*) exists with $K^* > 0$ and $L^* > 0$. This point is a saddle-point if $\gamma > 0$ and a stable nodal point if $\gamma < 0$.

In particular, if there are agglomeration diseconomies ($\gamma < 0$), then for any initial (K_0, L_0), $K_0 > 0$, $L_0 > 0$ it holds that

$$\lim_{t\to\infty}\big(K(t), L(t)\big) = (K^*, L^*), \tag{19}$$

$$\lim_{t\to\infty} w(t) = w^* = \beta A(L^*)(K^*)^{\alpha}(L^*)^{\beta-1}, \tag{20}$$

$$\lim_{t\to\infty} r(t) = r^* = \alpha A(L^*)(K^*)^{\alpha-1}(L^*)^{\beta}, \tag{21}$$

where w^* and r^* may differ from w_N and r_N. The limiting growth rate g^* is zero.

$$\lim_{t\to\infty}\frac{\dot{K}(t)}{K(t)} = \lim_{t\to\infty}\frac{\dot{L}(t)}{L(t)} = g^* = 0. \tag{22}$$

Assume instead agglomeration economies. If (K_0, L_0) falls on a stable branch of the saddle-point, then the equilibrium point (K^*, L^*) will be approached and (15)–(18) hold. The stable branches constitute a minimum threshold curve. If (K_0, L_0) lies below this curve then the urban economy will ultimately decay with

$$\lim_{t\to\infty}\big(K(t), L(t)\big) = (0, 0) \tag{23}$$

but if (K_0, L_0) lies above this curve, then the urban economy will grow with

$$\lim_{t\to\infty}\big(K(t), L(t)\big) = (\infty, \infty), \tag{24}$$

5. A SPECIAL CASE: AGGLOMERATION ECONOMIES AND DISECONOMIES COMBINED

So far agglomeration economies and diseconomies have been treated separately assuming that an urban economy falls either into the former or the latter category. Instead, it seems more reasonable to assume that there are agglomeration economies for small and diseconomies for large urban areas. This is frequently suggested in the literature and is similar to an assumption of a U-shaped long-run average cost curve.

To analyze this possibility we construct a continuous function $A(L)$ with $A'(L) > 0$ for $L < \bar{L}$ and $A'(L) < 0$ for $L > \bar{L}$. Hence, below $\bar{L}$ there is production under agglomeration economies and above $\bar{L}$ there are diseconomies. To make optimal use of the previous derivations $A(L)$ is chosen similar to (15), given by

$$A(L) = \left(\frac{L}{\bar{L}}\right)^{\gamma}, \quad \gamma = \begin{cases} \gamma_1 > 0, & L < \bar{L}, \\ \gamma_2 < 0, & L > \bar{L}, \end{cases} \tag{25}$$

where γ satisfies (16). The function $A(L)$ is shown in Figure 4.

To construct a phase diagram consider again the curves $\dot{K} = 0$ and $\dot{L} = 0$ given by $\hat{K}(L)$ and $\tilde{K}(L)$, respectively

$$\hat{K}(L) = \hat{\mu}\left(\frac{1}{\bar{L}}\right)^{\gamma/\beta} L^{(\gamma+\beta)/\beta}, \tag{26}$$

$$\tilde{K}(L) = \tilde{\mu}\left(\frac{1}{\bar{L}}\right)^{-\gamma/\alpha} L^{(\alpha-\gamma)/\alpha}. \tag{27}$$

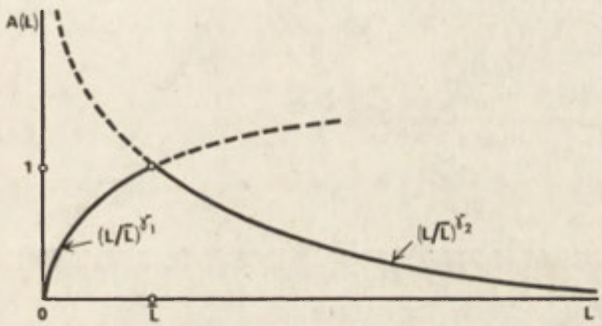


Fig. 4. The function $A(L)$

$\hat{K}(L)$ and $\tilde{K}(L)$ will be given for all L the subscript 1 if $\gamma = \gamma_1 > 0$ and subscipt 2 if $\gamma = \gamma_2 < 0$. The subscript 0 will be used if the functions satisfy (25) with $\gamma = \gamma_1$ for $L < \bar{L}$ and $\gamma = \gamma_2$ for $L > \bar{L}$. Obviously, then $\hat{K}_0(L)$ is simply equal to $\hat{K}_1(L)$ for $L < \bar{L}$ and equal to $\hat{K}_2(L)$ for $L > \bar{L}$. The same holds correspondingly for $\tilde{K}_0(L)$. At $\bar{L}$, $\hat{K}_1(L) = \hat{K}_2(L)$ and $\tilde{K}_1(L) = \tilde{K}_2(L)$ so that $\hat{K}_0(L)$ and $\tilde{K}_0(L)$ are continuous at $\bar{L}$ where agglomeration economies and diseconomies are joined.

$\hat{K}(L)$ and $\tilde{K}(L)$ as given in (26) and (27) have the same form as given in (17) and (18) except for the multiplicative factor involving $\bar{L}$. Hence, the properties and $\hat{K}(L)$ and $\tilde{K}(L)$ shown previously under assumptions of agglomeration economies and diseconomies apply now to the functions $\hat{K}_1(L)$, $\tilde{K}_1(L)$ and $\hat{K}_2(L)$, $\tilde{K}_2(L)$, respectively. In particular then, there exists a unique intersection point, say (K_1, L_1), for the functions $\hat{K}_1(L)$ and $\tilde{K}_1(L)$, and a unique intersection point (K_2, L_2) for the functions $\hat{K}_2(L)$ and $\tilde{K}_2(L)$. Moreover, these points satisfy

$$\hat{K}_0(\bar{L}) \{\gtreqless\} \tilde{K}_0(\bar{L}) \Leftrightarrow L_1 \{\lesseqgtr\} \bar{L} \{\lesseqgtr\} L_2, \tag{28}$$

which readily follows from the definition of $\hat{K}_0(L)$ and $\tilde{K}_0(L)$ and their continuity at $\bar{L}$. Finally, note that $\hat{K}_0(\bar{L}) = \hat{\mu}\bar{L}$ and $\tilde{K}_0(\bar{L}) = \tilde{\mu}\bar{L}$, so

$$\hat{\mu} \{\gtreqless\} \tilde{\mu} \Leftrightarrow \hat{K}_0(\bar{L}) \{\gtreqless\} \tilde{K}_0(\bar{L}). \tag{29}$$

Hence, three cases can be distinguished as shown in Figures 5–7, each corresponding to one of the inequalities in (28) and (29).

First, if $\hat{\mu} > \tilde{\mu}$, then $L_1 < \bar{L} < L_2$. Hence, (K_1, L_1) lies in the area for which $\gamma = \gamma_1$ and, therefore, $\hat{K}_0(L) = \hat{K}_1(L)$ and $\tilde{K}_0(L) = \tilde{K}_1(L)$ here. Also, (L_2, K_2) lies in the area for which $\gamma = \gamma_2$ and hence $\hat{K}_0(L) = \hat{K}_2(L)$ and $\tilde{K}_0(L) = \tilde{K}_2(L)$. So indeed there are two equilibrium points for curves $\hat{K}_0(L)$ and $\tilde{K}_0(L)$, $(K^*, L^*) = (K_1, L_1)$ being a saddle-point and $(K^*, L^*) = (K_2, L_2)$ being a stable nodal point.

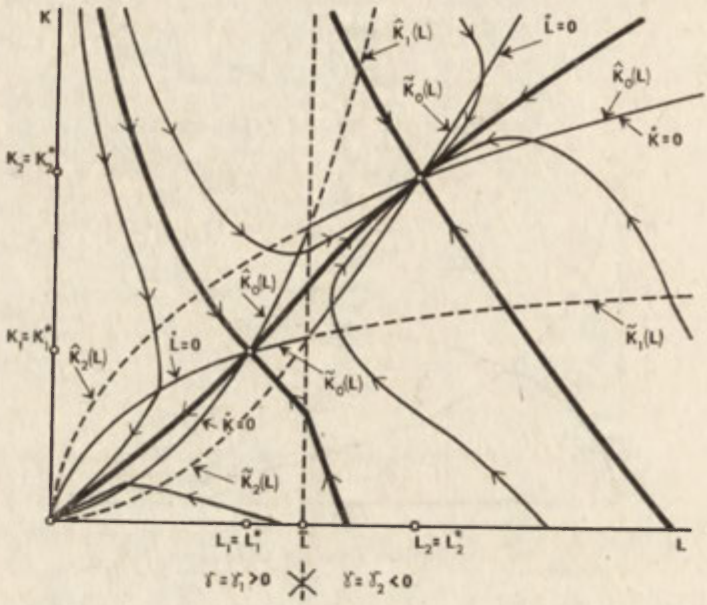


Fig. 5. Phase diagram: $\hat{\mu} > \tilde{\mu}$. Agglomeration economies and diseconomies

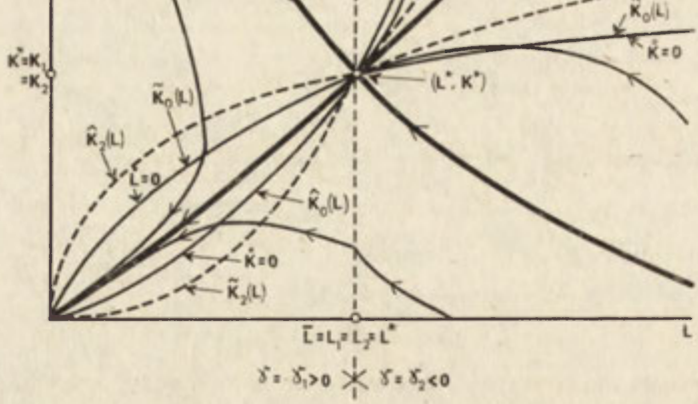


Fig. 6. Phase diagram: $\hat{\mu} = \tilde{\mu}$. Agglomeration economies and diseconomies

This is illustrated in Figure 5. There are five regions, three of which are enclosed by $\hat{K}_0(L)$ and $\tilde{K}_0(L)$ and two below or above both these curves. Noting that $\dot{K} < 0$ ($\dot{L} > 0$) above and $\dot{K} > 0$ ($\dot{L} < 0$) below curve $\hat{K}_0(L)$ ($\tilde{K}_0(L)$) it is possible to construct paths in the K–L plane.

Consider now the second case, that $\hat{\mu} < \tilde{\mu}$ and, therefore, $L_1 > \bar{L} > L_2$. Then $\hat{K}_0(L)$ and $\tilde{K}_0(L)$ have no intersection point other than the origin, for (K_1, L_1) lies in the area for which $\gamma = \gamma_2$ and (K_2, L_2) lies in the area for which $\gamma = \gamma_1$. In this case, from any initial (K_0, L_0) the city will ultimately decay to $(K, L) = (0, 0)$. This is illustrated in Figure 7.

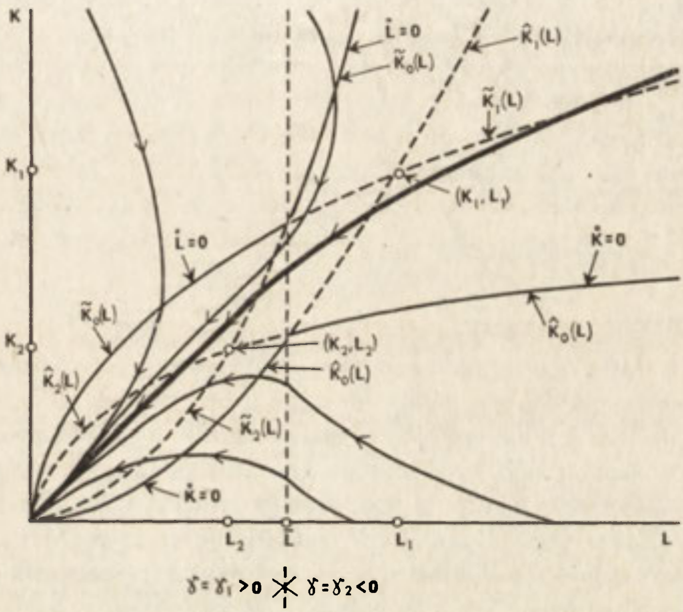


Fig. 7. Phase Diagram: $\hat{\mu} < \tilde{\mu}$. Agglomeration economies and diseconomies

Finally, there is the rather special case that $\hat{\mu} = \tilde{\mu}$ and, therefore, $L_1 = \bar{L} = L_2$, so $\hat{K}_0(L)$ and $\tilde{K}_0(L)$ have a unique intersection point, which for $L > \bar{L}$ has the characteristics of a nodal point and for $L < \bar{L}$ has those of a saddle-point. This case is illustrated in Figure 6. The following proposition restates these results and summarizes the properties of growth and decay paths for each of the cases.

Proposition 3. Consider an urban economy which produces according to (1) and (25), and grows according to (6) and (7). Agglomeration economies prevail if the town's labor force L is small, say below $\bar{L}$, but diseconomies exist for $L > \bar{L}$. Individual urban firms produce with constant returns to scale and factor payments exhaust total urban product. Then there exists a neighbhorhood around $(K, L) = (0, 0)$ such that for any initial (K_0, L_0) falling into this neighborhood the urban economy decays approaching $(K, L) = (0, 0)$ in infinite time. Also, the urban economy may not grow without upper bound and indeed will decay for (K_0, L_0) large enough.

In addition to $(K, L) = (0, 0)$ there exists either two, one or no positive equilibrium point such that $\dot{K} = 0$ and $\dot{L} = 0$. In particular, if $\hat{\mu} < \tilde{\mu}$, then two equilibrium points exist, say (K_1^*, L_1^*) and (K_2^*, L_2^*) with $0 < K_1^* < K_2^*$ and $0 < L_1^* < L_2^*$, being a saddle-point and a stable nodal point respectively. At (K_1^*, L_1^*) agglomeration economies exist and at (K_2^*, L_2^*) there are diseconomies. The stable branches of the saddle-point represent a minimum threshold. For any initial (K_0, L_0) below this threshold the urban economy decays towards $(K, L) = (0, 0)$, for initial (K_0, L_0) above this threshold (K_2^*, L_2^*) is approached. Only if (K_0, L_0) falls on a stable branch will (K_1^*, L_1^*) be approached.

If instead $\hat{\mu} > \tilde{\mu}$, then no equilibrium point exists with positive capital and labor stock. From an initial (K_0, L_0) the urban economy will decay, approaching $(K, L) = (0, 0)$ in infinite time. Finally, if $\hat{\mu} = \tilde{\mu}$, then there exists an equilibrium point (K^*, L^*), $K^* > 0$, $L^* > 0$, and a minimum threshold line passes through this point. For initial (K_0, L_0) below this line the urban economy decays. For (K_0, L_0) above this line the urban economy moves towards (K^*, L^*), However, this point is itself a minimum threshold. Only by historical accident can an urban economy move beyond the threshold line and a 'push' beyond this line does not lead to self-sustained growth.

6. CONCLUSIONS

The growth of a one sector urban economy has been analyzed through phase diagrams. The case of constant returns to scale, treated in neoclassical models, has been briefly examined. It has been shown that in this case the urban area approaches a steady state characterized by a constant capital/labor ratio as well as constant growth rate, wage and rental rates. In this equilibrium state an urban area may grow, stagnate, or decay; wage and rental differentials between urban area and the rest of the nation are not usually eliminated, so in- and outmigration exists in equilibrium.

The focus of the paper has been on agglomeration effects. Clearly, this is the more relevant case for urban areas. While empirical evidence is scarce there seems nevertheless to be some agreement, that urban areas of small size exhibit agglomeration economies while very large urban areas show diseconomies. A production function of this type has been analyzed in Section 5, and it has been shown that three parameter cases must be distinguished. In two of these cases an urban area decays or at most stagnates (see Fig. 6 and 7). These cases are not without interest. However, an urban area may only fall into one of these categories, if it reached its current size due to forces exogenous to the model, or if its parameters satisfied originally the third case (see Fig. 5). Otherwise, the urban area does not exist. The third parameter case, therefore, seems to be the most relevant to an explanation of urban growth. It allows for growth, decay and stagnation of an urban area, and illustrates well a number of points in the literature of urban growth.

First, a number of economists, such as Rosenstein-Rodan (1957) have argued that small urban areas or regions need a big push to enter a period of self-sustained growth. Below some minimum threshold size cities seem to stagnate or decay. It has been shown here that such a threshold size indeed exists in the presence of agglomeration economies. This threshold is defined by the level of both capital and labor, rather than by only one variable, say population or employment size. To overcome the threshold, the government must raise either the capital stock alone, or the urban employment level, or both.

Second, it has been shown that there exists an upper limit to the size of an urban area. This limit, of course, may vary among cities. An urban area may exceed this limit, but without outside intervention it will decline over time to a lower level equilibrium size.

An urban area may, therefore, decay for two reasons: because it has become too big and hence must fall back to its positive equilibrium size; or it may decay because it is below the minimum threshold size. Policy prescriptions will differ in the two cases. In the latter case a one time intervention by a national government can bring about self-sustained growth. In contrast, in the former case a similar action will be without success and continuous intervention is necessary if negative growth rates are to be prevented.

It has sometimes been argued that neoclassical models with an emphasis on steady state growth do not well describe or explain urban growth. Urban areas do not grow at a constant rate with a constant capital/labor ratio. Nor do they approach such a steady state over time, except possibly in stagnation. Such criticism is justified. However, it is shown here that the neoclassical assumptions may be modified to yield a better description of urban growth and decay. It is not necessary to abandon the neoclassical model entirely.

The model presented in this paper is extremely simple and should be extended in a number of directions. The inclusion of land as a factor and the division of urban production into two sectors producing for export and domestic consumption may add to the urban character of the model, which so far is insufficiently developed except for the consideration given to agglomeration effects. The Cobb-Douglas production function may be generalized to a linear homogeneous production function and the agglomeration effect which now depends only on the level of employment may be made to also vary with the urban capital stock. Finally, it must be pointed out that results may be quite different if national wage and rental rates were allowed to vary with factor flows from the urban area to the nation.

Author's note

This paper is based in part on research undertaken in 1973, supported by National Science Foundation grant GK-1065×2 to the Electronics Research Laboratory, University of California, Berkeley. The author wishes to thank P. Varaiya for his comments on earlier versions of the paper.

REFERENCES

Arrow, K., 1973, Criteria, Institutions, and Function in Urban Development Decisions, in A. H. Pascal (ed.), *Thinking About Cities.*

Borts, G. H., 1960, The Equalization of Returns and Regional Growth, *The American Economic Review*, 50 (June), 319–347.

Chipman, J. S., 1970, External Economies of Scale and Competitive Equilibrium, *Quarterly Journal of Economics*, 84, 3 (August), 347–385.

Czamanski, S., 1964, A Model of Urban Growth, *Papers of the Regional Science Association*, 13, 177–200.

Hanson, D. A. and B. v. Rabenau, 1976, Economic Growth in a Mixed Urban Economy: Infrastructure Investment and Labor Migration, *IEEE Transaction on Automatic Control*, 21 (April).

Herberg, H. and M. C. Kemp, 1969, Some Implications of Variable Returns to Scales, *The Canadian Journal of Economics*, 2:3 (August), 403–415.

Mills, E., 1972, *Studies in the Structure of the Urban Economy Resources for the Future*, The Johns Hopkins Press, Baltimore.

Muth, R. F., 1968, Differential Growth Among Large U. S. Cities, in J. P. Quirk and A. M. Zarley (eds.), *Papers in Quantitative Economics*, University Press of Kansas.

Niedercorn, J. and J. F. Kain, 1963, An Econometric Model of Metropolitan Development, *Papers and Proceedings of the Regional Science Association*, 11, 123–143.

Niehaus, J., 1963, Economic Growth with Two Endogenous Factors, *Quarterly Journal of Economics*, 77.

Rabenau, B. von. 1976, Optimal Growth of a Factory Town, *Journal of Urban Economics*, 3, 97–112.

Rabenau, B. von. 1974, Growth of an Urban Economy, *Proceedings of the IEEE Conference on Decision and Control.*

Rabenau, B. von, 1973, *Urban Dynamic Models: Maintenance and Replacement of Buildings, Spatial Development, and Economic Growth*, unpublished Dissertation, University of California.

Rosenstein-Rodan, P. N., 1957, *Notes on the Theory of the 'Big Push'*, M. I. T. Center for International Studies, March.

Siebert, H., 1969, *Regional Economic Growth: Theory and Policy*, International Textbook Company, Scranton.

Solow, R. M., 1956, A Contribution to the Theory of Economic Growth, *Quarterly Journal of Economics*, 70, 65–94.

Treadway, A. B., 1969, On Rational Entrepreneurial Behavior and the Demand for Investment, *Review of Economic Studies*, 36, 227–239.

Varaiya, P., 1973, *Monopsonistic/Competitive Production, Joint/ Separate Land Ownership and City Size*, mimeo.

Williamson, J. G. and J. A. Swanson, 1966, The Growth of Cities in the American Northeast, 1820–70, in *Explorations in Entrepreneurial History*, 4, Supplement, 3–101.

GEOGRAPHIA POLONICA 42, 1979

THE PATTERN AND TIMING OF LAND DEVELOPMENT IN A LONG RUN EQUILIBRIUM URBAN LAND USE MODEL

ALEX ANAS

Department of Civil Engineering, Northwestern University, Evanston, Il., USA

In the United States, regional scientists, economists, urban planners and geographers have contributed to the development of a standard static model of urban land use pioneered by Alonso (1964). Although all contributors agree that the model is static, a large number of contributors have all assumed that the model represents a 'long run equilibrium.' The difference between a static and a long run equilibrium model is a crucial one. While a static model determines equilibrium under the assumption that all growth occurs in an extremely brief instant and that all decisions are made simultaneously, a long run equilibrium model must show that a stationary (equilibirium) state arises at the end of some long time horizon T (> 0). Thus, long run equilibrium analyses cannot assume that all decisions are made simultaneously but *must show* how sequential decisions lead to an equilibrium.

A perusal of the literature highlights the extent of the confusion. Alonso for one did not commit himself to a long run equilibrium interpretation and carefully steered around the problem by writing as follows:

The city in which the individual arrives is a simplified city. It lies on a featureless plain, and transportation is possible in all directions. All employment and goods and services are available only at the center of the city. Land is bought and sold by free contract, without any institutional restraints *and without having its character fixed by any structures existing upon the ground.* (Alonso, 1964, p. 18, emphasis added.)

More than a decade later, a modern contributor, Wheaton, (1977), uses Alonso's static model as if it is a long run equilibrium model and as if the model represents the way in which things will evolve in the long run and *not* the way in which things will be if an instant city could be built. He prefaces his analysis as follows:

In the short run, urban land use is rigid and consumer location decisions are based on the characteristics of the standing housing stock and the resultant pattern of spatial externalities. As the time horizon lengthens, capital becomes mobile and land use is determined primarily by a *long-run* trade-off between travel and residential density. (Wheaton, 1977, page 139, emphasis added.)

Numerous other examples can be found where authors have interpreted Alonso's static model as a long run equilibrium model. The main argument of this paper is that this interpretation is possibly erroneous. This argument is capsulated in the following proposition.

Proposition: If the standard static model of urban land is a *long run equilibrium* model, then a spatiotemporal model exists in which decisions are made sequentially and which at the end of a *long* time horizon, produces the same stationary spatial outcome that the standard static model produces.

This paper constructs such a model which produces a long run equilibrium but is also consistent with the static model of urban land use. Unfortunately the implication is that it is not possible to have a *market* model which produces an equilibrium which also agrees with Alonso's static model. Those authors who assumed the above proposition to be correct have thus made (sometimes willingly) an unreasonable and unjustified leap of faith. This paper shows that agreement between the static model and the long run equilibrium model developed here is possible only under very restrictive assumptions which require nonmarket land use controls. Thus, it is shown that the static model has no value in long run prediction. It can only predict what is an equilibrium now if all decisions could be made from scratch at the current time.

The paper does show, however, that a spatiotemporal decentralized planning procedure which agrees with the static model does exist. This procedure produces an optimal land use pattern which at time T is identical to that of the of the static model. The way this land use pattern develops, however, is not necessarily an outward expansion from the center of the city, but involves discontinuous and/or scattered land development.

To obtain these answers, a simple scenario of urban growth is developed. To summarize: A new city is embedded into a larger national economy. The ultimate population this city will reach is small compared to that of the national economy. The same utility level prevails both within and outside the city. Households are thus indifferent between living within or outside the city. Competitive landowner-speculators anticipate the ultimate long run equilibrium. As the city's employment opportunities and services expand over time more households move into the city until at some time an ultimate population is reached. A larger number of households cannot be accommodated at the prevailing utility level. We preserve all of the essential aspects of a static urban economy. Household income, tastes, transportation costs, building costs and the nonurban land rent are the basic 'parameters' of the model. These are kept invariant over time while we examine the pattern and timing of land development under exogenous population growth.

In part I we develop the essential aspects of competitive residential land speculation. Each landowner jointly decides both how to determine the density of development, and also when to release nonurban land into housing. In part II the demand side is examined and the conditions for households' equilibrium are derived. In part III it is shown that the speculative market can be controlled by issuing building permits in order to assure that the supply of housing units at any point in time does not exceed the demand. Part IV develops a linear programming formulation of the socially efficient land development pattern which maximizes the present value of the urban land. It is shown that land speculation can be controlled via a combination of location specific land taxes and time specific development taxes. Finally, in part V a graphical representation of the efficient development pattern is developed under the assumption of continuous population growth. Part VI then criticizes the long standing assumption that the above mentioned proposition is true. It is argued that deviations from the assumption of the model developed here will produce more satisfactory spatiotemporal models, but that none of these will be a long run equilibrium equivalent of the static model. Thus, all past attempts to draw long run predictions from Alonso's static model appear to have been severely mistaken.

I. SPECULATIVE LAND DEVELOPMENT IN A STATIC ECONOMY

In formal models of urban land use, each location is distinguished by its radial distance, x, to an employment center where all jobs are located. At some initial time $t = 0$ all land around this center is in farming. However, if landowners are competitive they will speculate about the future use of their land in housing. In general, the present value at time $t = 0$ of a unit sized land parcel at distance x from the center will be

$$V(x, t^*, Q) = R_0 \int_0^{t^*} e^{-\alpha s} ds + R(x, Q) \int_{t^*}^{\infty} e^{-\alpha s} ds - e^{-\alpha t^*} C. \tag{1}$$

In (1), R_0 is the opportunity rent of land in farming while $R(x, Q)$ is the rent per unit of land expected from a residential lot of size Q. The present value, V, depends on the choice of lot size Q and the choice of the development date t^*, as well as the location of the lot, x, and the prevailing economic rate of discount, α. The developed value of land also depends on the present value of the cost of construction per unit of land C.

There are several assumptions in (1):

Assumption A1: The agricultural rent R_0 is invariant over time and location.

Assumption A2: The construction cost per unit of land, C, is invariant over time and location. C will not change with location if the housing industry is constant returns to scale and if the amount of capital per unit of land is the same in all housing.

Assumption A3: The rent from residential use, $R(x, Q)$, is invariant over time at every location.

Assumption A4: Once built at time t^* a residential unit is immediately occupied and does not become vacant at any future date.

Integrating (1):

$$V(x, t^*, Q) = (1/\alpha)\,[R(x, Q) - R_0 - \alpha C]e^{-\alpha t^*} + (1/\alpha)R_0. \tag{2}$$

This has the properties.

$$V(x, 0, Q) = [R(x, Q) - \alpha C]/\alpha, \tag{2a}$$

$$\lim_{t^* \to \infty} V(x, t^*, Q) = R_0/\alpha, \tag{2b}$$

$$\partial V(x, t^*, Q)/\partial t^* = -[R(x, Q) - R_0 - \alpha C]e^{-\alpha t^*} < 0, \tag{2c}$$

$$\partial^2 V(x, t^*, Q)/\partial t^{*2} = \alpha[R(x, Q) - R_0 - \alpha C]e^{-\alpha t^*} > 0. \tag{2d}$$

These properties are shown in Figure 1.

Competitive land speculators must solve the following problem:

$$\max_{t^*,Q} V(x, t^*, Q). \tag{3}$$

It follows from (2a)–(2c) and from Figure 1 that for each speculator the preferred development date is $t^* = 0$.

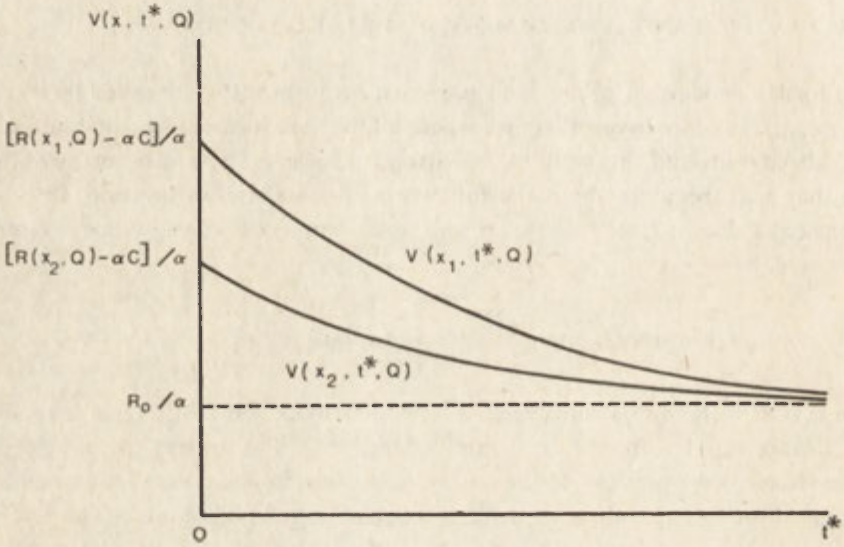


Fig. 1

Property 1: Under A1–A3, each land speculator irrespective of location will want to release land into housing at the earliest possible date when assumption A4 is valid, i.e., no vacancy can result in the future.

Property 2: Because of the monotonic relationship between $V(x, t^*, Q)$ and $R(x, Q)$ given t^*, it follows that each speculator need only be concerned about choosing that lot size which maximizes rent $R(x, Q)$ irrespective of the time of development. This is precisely what occurs in the standard static model where the time dimension is suppressed.

Thus, assumptions A1–A4 enable us to strike an equivalence between the standard static model and the spatio temporal model (s) that must be equivalent to it.

II. HOUSEHOLD EQUILIBRIUM

On the demand side, households are myopic renters and are concerned about their instantaneous welfare. Each household can achieve the same welfare level within or outside the city. Households welfare, U, is a concave function of lot size, Q, and consumption expenditures, Z.

$$U = U(Z, Q). \tag{4}$$

The household's budget constraint is,

$$Y - Z - R(x, Q)Q - T(x) = 0, \tag{5}$$

where Y is instantaneous income, $T(x)$ is commuting cost to and from work as a monotonically increasing function of distance, x. Equation (4) can be transformed to express the indifference curve for utility level $\bar{U}$ as

$$Z = Z(\bar{U}, Q), \tag{6}$$

where,

$$\partial Z/\partial \bar{U} > 0 \text{ and } \partial Z/\partial Q < 0. \tag{6'}$$

Substituting (6) into (5) and solving for $R(x, Q)$ we obtain the household's bid rent function,

$$R(x, Q) = [Y - Z(\bar{U}, Q) - T(x)]/Q. \tag{7}$$

From property 2, this is the expression each land speculator must maximize through appropriate choice of lot size Q.

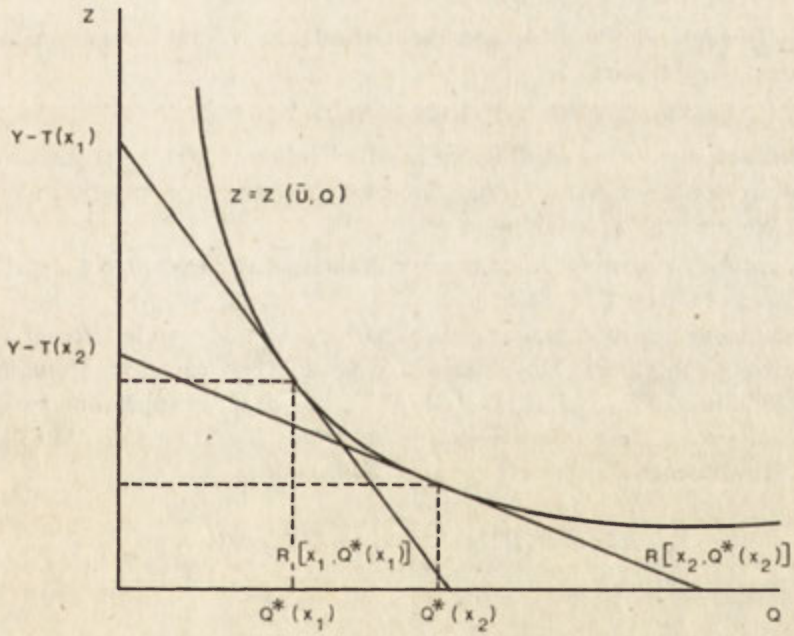


Fig. 2

We now inquire about the conditions that must hold on the demand side so that $R(x, Q)$ is invariant over time. The following is a sufficient set of conditions:

Assumption A5: All households are identical in income Y and in tastes and Y, $\bar{U}$ and $T(x)$ remain invariant over time.

It then follows that $R(x, Q)$ remains invariant over time as per Assumption A3.[1]

[1] More precisely A5 is sufficient but not necessary for $R(x, Q)$ to be time invariant. The necessary and sufficient condition is that $Y - T(x)$, and $\bar{U}$ remain time invariant.

Now applying property 2, lot sizes will be determined by maximizing $R(x, Q)$ at each x,

$$\underset{Q}{\mathrm{Max}} R(x, Q) = [Y - Z(\bar{U}, Q) - T(x)]/Q \tag{8}$$

which yields

$$Q^*(x) = Q[Y - T(x), \bar{U}]. \tag{8'}$$

This is equivalent to maximizing the slope of the household's budget constraint AB in Figure 2 which implies the tangency at point M. From the household's viewpoint, $R(x, Q^*)$ being fixed, point M corresponds to the maximization of utility which yields the familiar marginal condition

$$\left.\frac{\partial Z(\bar{U}, Q)}{\partial Q}\right|_{Q^*} = R(x, Q^*). \tag{8''}$$

Since $T(x_2) > T(x_1)$ for any $x_2 > x_1$, it follows that $R[x_2, Q^*(x_2)] < R[x_1, Q^*(x_1)]$. This well known result states that rents decline with distance from the center and is readily observed in Figure 2. The result that lot size must increase with location, $Q^*(x_2) > Q^*(x_1)$, is also observed.

The city will reach a long run household population $N(T)$ at some ultimate time T. At any intermediate time t,

$$N(0) < N(t) < N(T), \quad 0 < t < T,$$

where $N(0)$ represents the small initial population,[2] and $N(t)$ the population accumulated up to time t. We note that for A4 to be fulfilled, population must never decline. We state this as assumption A6.

Assumption A6: City population never declines, i.e. $\dot{N}(t) \geq 0, 0 \leq t \leq T$. There is no growth for $t > T$.[3]

The ultimate equilibrium allocation of $N(T)$ households can be derived using the well known static model. This amounts to solving two equations simultaneously for $\bar{x}$, the ultimate city radius, and for $N(T)$, the ultimate population holding capacity of the city. These equations are given below. Note that since $N(T)$ is found and the growth schedule is given, T is also endogenous.

$$V(\bar{x}, t^*, Q^*(\bar{x})) = R_0 \int_0^\infty e^{-\alpha s} ds = R_0/\alpha, \tag{9}$$

and

$$N(T) = 2\pi \int_0^{\bar{x}} Q[Y - T(u), \bar{U}]^{-1} u du. \tag{10}$$

Equation (9) states that at the boundary land must yield the same long run value in housing or agriculture. Simplifying the left hand side of (9) we find that an equivalent condition can be stated in terms of instantaneous rents. This is

$$R(\bar{x}, Q[Y - T(\bar{x}), \bar{U}]) - \alpha C = R_0. \tag{11}$$

[2] Without any loss of generality $N(0) = 0$.

[3] If population did decline for some time interval, it would be impossible to satisfy A4, the assumption of no vacancies without demolishing buildings.

Conversely, integrating both sides of (11) we could derive equation (9) which holds for any $0 \leq t^* \leq T$.

Equation (10) states that all households must be located within a circle of radius $\bar{x}$. We can now state a third property.

Property 3: Once the city is in equilibrium at time T according to (9) and (10), it then follows from A5 and A6 that it will remain in that state for all $t > T$.

Again, we must note that if assumption A5 were to be removed, A3 would not hold either, and the equivalence with the standard static model would be lost. A question could be raised as follows: "If the utility level is fixed over time, then why isn't it the case that the city grows to ultimate size immediately ?". If this happened T would shrink to zero and we would be back at the static model. The answer is that population does not grow to ultimate size immediately because job opportunities and the capacity of urban services grow slowly and (by assumption) exogenously. It is these that determine the population growth shedule $N(t)$ of assumption A5. In a general model with several sectors, the growth of urban services and job would also have to be made endogenous.

We now turn to an examination of the spatial patterns that must be implied in this model which has been kept equivalent to the static model. We find that a public body must control the expectations of land developers and the pattern of land development to assure a long run equilibrium.

III. CONTROLLED LAND SPECULATION

Can a planning authority determine a decentralized land development procedure that will assure a long run equilibrium for time $t \geq T$ as determined by (9) and (10)? The answer is in the affirmative and the decentralized procedure need consist of only two basic rules:

Rule 1: The planning authority must assure that the number of residential units developed at each time t is equal to the number of households arriving at that time. One way to achieve this is to issue the right number of building permits at each time.

Rule *1* eliminates the possibility of vacancies in the market. Each land speculator is assured of occupancy as soon as a permit is issued.

Rule 2: The planning authority must not issue any building permits to land speculators anticipating a rent greater than $R(x, Q^*(x))$ at location x (or a utility level $U < \bar{U}$). Rent, $R(x, Q^*(x))$, and utility level, $\bar{U}$, are both consistent with long run equilibrium computed from (9) and (10).

The spatial pattern of development can be completely random, i.e., building permits can be issued randomly (via a lottery) across the locations from 0 to $\bar{x}$. By property 1 a land speculator will develop as soon as a permit is issued and by property 2 and rule 2 the land speculator at x will choose lot size $Q^*(x)$ consistent with $\bar{U}$, since from Figure 2,

$$R(x, Q[Y-T(x), \bar{U}]) > R(x, Q[Y-T(x), \hat{U}]), \quad \hat{U} > \bar{U}.$$

If a land speculator anticipates a utility level lower than the ultimate $\bar{U}$, households would not afford this development and it would remain forever vacant, since $\bar{U}$ is the opportunity welfare level of households moving into the city.

IV. DECENTRALIZED PLANNING AND SOCIALLY EFFICIENT LAND DEVELOPMENT

A possible objective for the planning authority is to issue building permits over space and time in such a fashion as to maximize the aggregate present value of the entire urban land. This would have the effect of maximizing the fiscal property tax base of the urban government. Clearly, a random pattern of building permit distribution need not satisfy this goal of efficient land development.

Formally, the objective of maximizing the present value of the urban land can be posed as a linear programming problem, using discrete notation for both time and space. Suppose the land area is divided into $l = 1, \ldots, L$ concentric annular districts of width Δ_l and total land area $A_l = 2\pi x_l \Delta_l$ each, where x_l is the radius of the lth annulus. We could for convenience number the districts in terms of increasing distance from the center, i.e.,

$$x_1 < x_2 < \ldots < x_L.$$

Each district can be chosen to have an arbitrarily narrow width Δ_l. The time horizon can be divided into $t = 1, \ldots, T$ arbitrarily brief time periods. We introduce the following discrete notation:

$$Q_l = Q^*(x_l),$$
$$V_{lt} \equiv V\big(x_l, t_l^*, Q^*(x_l)\big),$$

where V_{lt} is the present value of the lth district located at radius x_l if developed at time t_l^*. Further, let n_{lt} be the number of households added to district l at time t and n_t the number added to the entire urban area at time t. The following relations hold by definition of the problem:

$$\sum_{l}^{L} 2\pi x_l \Delta_l = \pi \bar{x}^2 \tag{12}$$

and

$$\sum_{t}^{T} n_t = N(T). \tag{13}$$

The public authority's objective can be stated as the following linear programming problem with $\{n_{lt}\colon l = 1, \ldots, L,\ t = 1, \ldots, T\}$ as the primal variables and $\{Q_l V_{lt}\colon l = 1, \ldots, L,\ t = 1, \ldots, T\}$ as the primal objective function coefficients:

$$\text{Max} \sum_{t}^{T} \sum_{l}^{L} (Q_l V_{lt}) n_{lt} \tag{14}$$

subject to:

$$\sum_{l}^{L} n_{lt} \leqslant n_t, \quad t = 1, \ldots, T, \tag{14a}$$

$$\sum_{t}^{T} Q_l n_{lt} \leqslant A_l, \quad l = 1, \ldots, L, \tag{14b}$$

$$n_{lt} \geqslant 0, \quad l = 1, \ldots, L, \quad t = 1, \ldots, T. \tag{14c}$$

Although (14a) and (14b) are stated as inequalities it follows from (10) that if the total land area is chosen equal to $\pi \bar{x}^2$ according to (12), then no land remains vacant, i.e. at the optimum all (14b) hold as equalities. If some households are not located at time t then (14a) would be an inequality at the optimum. But this would necessarily result in some vacant land and since aggregate land values would be increased if no land was left vacant, all (14a) must also hold as equalities at the optimum solution.

The dual problem is as follows:

$$\text{Min} \sum_{l}^{L} \mu_l A_l + \sum_{t}^{T} \lambda_t n_t \tag{15}$$

subject to:

$$\mu_l Q_l + \lambda_t \geqslant Q_l V_{lt}, \quad t = 1, \ldots, T, \quad l = 1, \ldots, L, \tag{15a}$$

$$\mu_l \geqslant 0, \quad l = 1, \ldots, L, \tag{15b}$$

$$\lambda_t \geqslant 0, \quad t = 1, \ldots, T, \tag{15c}$$

where μ_l and λ_t are the dual variables associated with (14b) and (14a), respectively.

From this dual we can see that the following rule describes the decentralized planning procedure that must be followed to attain the socially efficient optimum pattern of land development. Rule 3 below replaces rules 1 and 2.

***Rule** 3*: The planning authority announces at time $t = 0$ the developed value of land, V_{lt}, in each district l that will result if that district is developed at time t.[4] For each district the authority charges a land tax of μ_l^* per unit of land to be paid by the speculator at time $t = 0$. Finally, a development tax of Λ_t is imposed on every unit developed at time t. The development tax for each time period is announced and collected at the time of development and is computed in such a way that $\lambda_t^* = e^{-\alpha t} \Lambda_t$.

If this rule is enforced, speculators in some district l will develop at such time t that they just break even. Formally, the return from development must just equal the sum of land and development taxes incurred:

$$\mu_l^* Q_l + \lambda_t^* = Q_l V_{lt}. \tag{16}$$

[4] The stream of developed land values need not be announced if the market is competitive and all speculators have perfect information about the future and use the same discount rate as the planning authority.

The proof follows directly from the complementary slackness conditions of the duality theorem in linear programming.[5]

$$n_{lt}^*(\mu_l^* Q_l + \lambda_t^* - Q_l V_{lt}) = 0, \tag{17}$$

which implies,

$$n_{lt}^* \geqslant 0 \text{ when } \mu_l^* Q_l + \lambda_t^* - Q_l V_{lt} = 0, \tag{17a}$$

$$n_{lt}^* = 0 \text{ when } \mu_l^* Q_l + \lambda_t^* - Q_l V_{lt} > 0. \tag{17b}$$

We also know from the duality theorem that the aggregate implication of (17) is

$$\sum_t^T \sum_l^L Q_l V_{lt} n_{lt}^* = \sum_l^L \mu_l^* A_l + \sum_t^T \lambda_t^* n_t. \tag{18}$$

This states that maximizing the present value of land is equivalent to minimizing the present value of the sum of land and development taxes.

Rule 3 provides an alternative to rules 1 and 2 which required the public authority to issue building permits in order to achieve a decentralized outcome. In a controlled market with speculative foresight, announcing a schedule of land and development taxes achieves the same effect. The development tax of rule 3 must be set so that each developer breaks even at the time that developer would have made maximum profit if there was no development tax. Then, each developer will develop at the break even time knowing that normal profits is the best that can be expected.

If rules 1 and 2, or rule 3 are not enforced, then a market solution cannot be assured since all developers would want to develop at time $t = 0$. This would lead unavoidably to overbuilding, to vacancies and the equivalence with the static model would be lost. Alternatively, if rents were to adjust so that no over-building occurred, the'lot size decision' would no longer be separable from the 'timing' decision and the equivalence with the static model would be lost once again.

V. THE SPATIAL PATTERN OF EFFICIENT LAND DEVELOPMENT

The dual problem can be reformulated using continuous representation for time and preserving the discrete notation for the annular districts:

$$\underset{\{\lambda(t),\mu_l, l=1,\dots,L\}}{\text{Min}} \sum_l^L \mu_l A_l + \int_0^T \lambda(s) n(s) ds \tag{19}$$

subject to

$$\mu_l Q_l + \lambda(t) \geqslant Q_l V_l(t), \quad l = 1, \dots, L, \tag{19a}$$

$$\mu_l \geqslant 0, \quad l = 1, \dots, L, \tag{19b}$$

$$\lambda(t) \geqslant 0, \quad 0 \leqslant t \leqslant T. \tag{19c}$$

This formulation is valid as long as population grows smoothly over time, without 'jumps', i.e. $n(t)$ is a continuous function of time, whereas in the discrete time

[5] For the duality theorem see Fujita (1976).

linear programming formulation it was not necessary that the pattern of population growth be continuous. The population cumulated up to time t is expressed as

$$N(t) = \int_0^t n(s)\,ds, \tag{20}$$

with

$$\dot{N}(t) = n(t), \tag{20a}$$

$$N(0) = 0, \quad N(T) = N. \tag{20b}$$

With this specification of the problem the development tax $\lambda(t)$ will also be a continuous function of time. Continuity of the urban space is approximated by letting $L \to \infty$. This implies that each annular district is of infinitesimal width since all districts are contained within a circle of finite radius $\bar{x}$.

We may now rewrite the constraint (19a) using (2) in the right hand side:

$$\mu_l Q_l + \lambda(t) \geqslant (1/\alpha)Q_l[R(x_l, Q_l) - R_0 - \alpha C]\mathrm{e}^{-\alpha t} + (1/\alpha)R_0 Q_l, \tag{21}$$

where $Q_l \equiv Q^*(x_l)$. We abbreviate by making the following definitions:

$$B_l \equiv (\mu_l - R_0/\alpha)Q_l \tag{21a}$$

and

$$M_l(t) \equiv (1/\alpha)e^{-\alpha t}[R(x_l, Q_l) - R_0 - \alpha C]Q_l. \qquad (21b)^{6}$$

Thus

$$B_l + \lambda(t) \geqslant M_l(t). \tag{22}$$

From (21b) we can redefine $M_l(t)$ as the product of a time dependent function and a location dependent function.

$$M_l(t) = F(t)G_l(x_l), \tag{23}$$

where

$$F(t) \equiv (1/\alpha)e^{-\alpha t} \tag{24a}$$

and

$$G_l(x_l) \equiv [R(x_l, Q_l) - R_0 - \alpha C]Q_l. \tag{24b}$$

Property 4: For each district $l = 1, \ldots, L-1$ the corresponding $M_l(t)$ time profile has the following characteristics:

$$\dot{M}_l(t) < 0, \quad l = 1, \ldots, L-1, \qquad (25a)^{7}$$

$$\ddot{M}_l(t) > 0, \quad l = 1, \ldots, L-1. \tag{25b}$$

For the last district $l = L$ we already know from (11) that

$$R(x_L, Q_L) = R_0 + \alpha C, \tag{26}$$

---

[6] $M_l(t)$ is the value of land developed at time t in one housing unit in l net of the nonurban opportunity worth of that land, i.e.

$$M_l(t) = V_l(t)Q_l - R_0 Q_l/\alpha.$$

[7] See (2c), (2d).

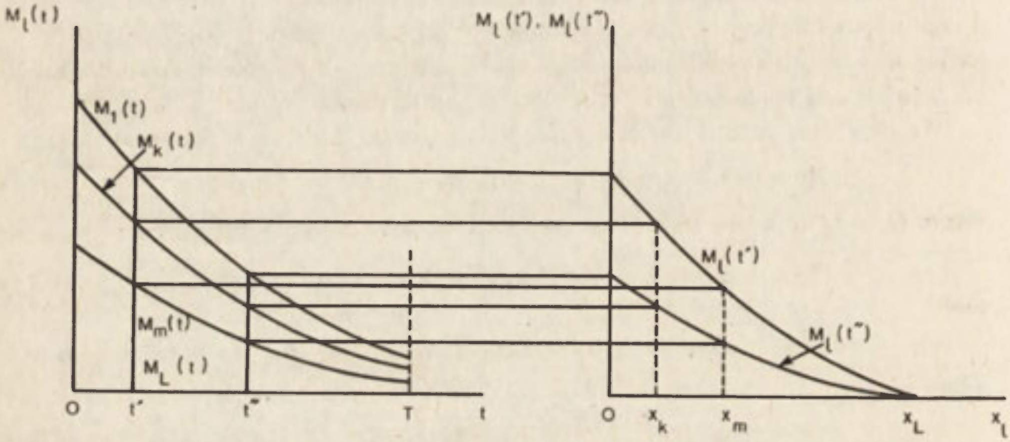
which implies $G_L(x_L) = 0$ and

$$M_L(t) = 0, \tag{27a}$$

$$\dot{M}_L(t) = 0, \tag{27b}$$

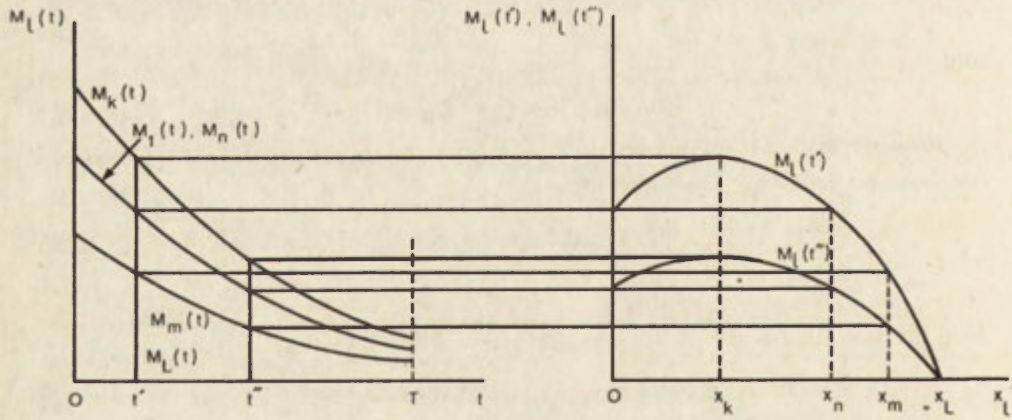
$$\ddot{M}_L(t) = 0. \tag{27c}$$

Property 5: At each time t the profile of $M_l(t)$ across the districts $l = 1, \ldots, L$ could either decrease monotonically until $M_L(t) = 0$, or it could first increase up to some district $l = k$ at distance x_k and then decrease until $M_L(t) = 0$.



(a)

CASE A : $M_1(t) > M_2(t) > \ldots > M_k(t) > \ldots > M_L(t)$.



(b)

CASE B : $M_1(t) < M_2(t) < \ldots < M_k(t) > \ldots > M_L(t)$.

Fig. 3

To prove this we note that,

$$dM_l(t)/dx_l = F(t)\left(dG_l(x_l)/dx_l\right)$$
$$= F(t)\{[R(x_l, Q_l) - R_0 - \alpha C](dQ_l/dx_l) + Q_l(dR(x_l, Q_l)/dx_l)\} \gtreqless 0 \quad (28)^8$$

since $dQ_l/dx_l > 0$ and $dR(x_l, Q_l)/dx_l < 0$.

From (27a) we know that for large l (near $l = L$) $dM_l(t)/dx_l < 0$, but for small l (near $l = 1$) $dM_l(t)/dx_l \gtrless 0$. It also follows that the district k where $M_l(t)$ attains a maximum is the same at every time t. This follows from,

$$\operatorname*{Max}_l\{M_l(t),\ l = 1, \ldots, L\} = F(t)\operatorname*{Max}_l\{G_l(x_l),\ l = 1, \ldots, L\}. \quad (29)$$

Property 5 implies two possible cases. These are:

Case A:

$$M_1(t) > M_2(t) > \ldots > M_{k-1}(t) > M_k(t) > M_{k+1}(t) > \ldots > M_L(t),\ 0 \leqslant t \leqslant T.$$

Case B:

$$M_1(t) < M_2(t) < \ldots < M_{k-1}(t) < M_k(t) > M_{k+1}(t) > \ldots > M_L(t),\ 0 \leqq t \leqslant T.$$

Both of these cases are shown in Figure 3 where for case A the uppermost profile is $M_1(t)$ whereas for case B the uppermost is $M_k(t)$.

Property 6: $d\dot{M}_l(t)/dx_l = \dot{F}(t)\left(dG_l(x_l)/dx_l\right)$. At every time t:

For Case A,

$$d\dot{M}_l(t)/dx_l > 0, \quad \text{for all } x_l.$$

For Case B,

$$d\dot{M}_l(t)/dx_l < 0, \quad \text{for } x_l < x_k,$$
$$d\dot{M}_l(t)/dx_l = 0, \quad \text{for } x_l = x_k,$$
$$d\dot{M}_l(t)/dx_l > 0, \quad \text{for } x_l > x_k,$$

where $M_k(t) = \max\limits_l\{M_l(t),\ l = 1, \ldots, L\}$.

For both cases A and B this implies that the slope of $M_l(t)$ evaluated at time t becomes flatter as we move from the highest $M_l(t)$ schedule to the lowest. To state this formally, pick any two districts i and j and any time t.
Then,

$$M_i(t) > M_j(t) \Rightarrow |\dot{M}_i(t)| > |\dot{M}_j(t)|, \quad 0 \leqslant t \leqslant T. \quad (30)$$

This is also equivalent to

$$M_i(t) > M_j(t) \Rightarrow \ddot{M}_i(t) > \ddot{M}_j(t) \geqslant 0, \quad 0 \leqslant t \leqslant T. \quad (31)$$

We now turn to several properties about the spatial pattern of land development:

Property 7: The tax schedule $B_l^* + \lambda^*(t)$ lies above the $M_l(t)$ profile and touches this from above at the development date t_l^* for that district.

This property is an expression of (22) which follows from the complementary

[8] Strictly speaking, the profile could produce more than one peak, but sufficient conditions which guarantee the above two cases only, probably exist. Analyzing only the two cases of property 5, provides all the insight that is needed. The reader can easily generalize to cases where more than one peak exists. This will become clear below.

slackness condition (17). The speculators in district l can just afford to develop at time t_l^* as shown in Figure 4, and will incur losses if they develop at other times. Formally,

$$B_l^* + \lambda^*(t_l^*) = M_l(t_l^*), \quad l = 1, ..., L. \tag{32}$$

From property 7 we can also state that

$$[\dot{\lambda}^*(t)]|_{t_l^*} = [\dot{M}_l(t)]|_{t_l^*}, \quad l = 1. .... L. \tag{33}$$

If (33) did not hold it would be possible to find a t' such that

$$B_l^* + \lambda^*(t') < M_l(t').$$

This would violate (22).

Property 8: The land tax μ_l^* adjusts from district to district in such a manner that the tax schedule $B_l^* + \lambda^*(t)$ touches the corresponding $M_l(t)$ schedule at the relevant development date t_l^*.

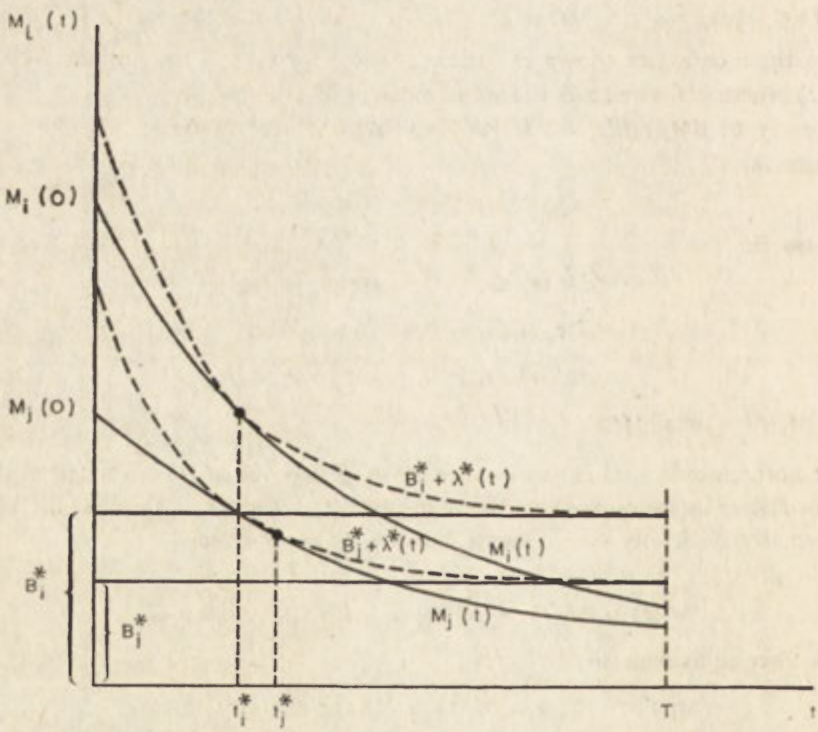


Fig. 4

This is also seen in Figure 4, and follows from (32).

Property 9: The development tax, $\lambda^*(t)$, must decline with time, i.e. $\dot{\lambda}^*(t) \leqslant 0$, $0 \leqslant t \leqslant T$.

Also seen in Figure 4, this follows from (33). Since the right hand side is known to be negative from (25a), the left hand side must also be negative for all values of t_l^*. Since there is some development at every time t, it follows that t_l^* eventually assumes

all values in $[0, T]$, thus $\lambda^*(t)$ is downward sloping, i.e. $\dot{\lambda}^*(t) \leqslant 0$ for all $0 \leqslant t \leqslant T$.
Property 10: For any $0 \leqslant t \leqslant T$,

$$\ddot{\lambda}^*(t) > \ddot{M}_l(t) \geqslant 0, \quad l = 1, \ldots, L. \tag{34}$$

Let $M_k(t) = \max_l [M_l(t), \; l = 1, \ldots, L]$. Then, we already know from (31) that

$$\ddot{M}_k(t) > [\ddot{M}_l(t), \; l \neq k]. \tag{35a}$$

By property 7 $B_k^* + \lambda^*(t)$ must touch $M_k(t)$ from above which implies that

$$\ddot{\lambda}^*(t) > \ddot{M}_k(t). \tag{35b}$$

Then, (34) follows from (35a) and (35b).

Property 11: The spatial pattern of land development is as follows: Development begins at $t = 0$ in district j where $M_j(t) = \max_l [M_l(t), \; l = 1, \ldots, L]$, then spreads to those districts with the next highest $M_l(t)$ schedules, and so on until the district $l = L$, with the lowest schedule is developed at time $t = T$.

This is proven by contradiction. Suppose that development in district j begins at $t_j^* > 0$. It then follows from (33) that

$$-[\dot{\lambda}^*(t)]|_{t<t_j^*} > -[\dot{M}_j(t)]|_{t<t_j^*}. \tag{36a}$$

Since it is also known from (30) that for any t,

$$-\dot{M}_j(t) > -\dot{M}_m(t), \quad M_j(t) > M_m(t) \tag{36b}$$

it follows that

$$-[\dot{\lambda}^*(t)]|_{t<t_j^*} > -[\dot{M}_m(t)]|_{t<t_j^*}, \tag{36c}$$

where m is any district other than j. This implies that (33) will never hold for $0 \leqslant t < t_j^*$. But this is not possible since some district must be developed at each time. It then follows that $t_j^* > 0$ cannot be correct and that $t_j^* = 0$ for j, $M_j(t) = \max_l [M_l(t), \; l = 1, \ldots, L]$. In a similar way we would prove that the district i with the next highest $M_l(t)$ profile, i.e. $M_i(t) = \max_l [M_l(t), \; l \neq j]$, must be developed at $t_i^* = t_j^* + \varepsilon$ where ε is an arbitrarily small positive number.

This general property of spatial land development can be translated into two specific patterns, one for case A, the other for case B.

Property 11A: If the $M_l(t)$ profiles are as in case A, development starts at $t = 0$ at the central district $l = 1$ and spreads outward continuously until it reaches the last district $l = L$ at time T.

To prove this we apply property 11 and recall from property 5 that for case A the $M_l(t)$ profiles are ordered as the districts are ordered.

Property 11B: If the $M_l(t)$ profiles are as in case B, development starts at $t = 0$ at the district $l = k$ (at distance $x = x_k$) where

$$M_k(t) = \max_l [M_l(t), \; l = 1, \ldots, L].$$

Development then proceeds continuously and simultaneously towards the center ($l < k$) and towards the city boundary ($l > k$). Districts m and n where $m < k$

and $n > k$ and $M_m(t) = M_n(t)$, are developed simultaneously. Development reaches the city center and district j ($j > k$) at some time $t_1^*(= t_j^*)$ but continues to move beyond district j until it reaches the last district $l = L$ at time $t = T$.

To prove this we again apply property 11 recalling that the highest $M_l(t)$ schedule is that of district k at distance x_k where $G_l(x_l)$ is a maximum. By property 11 development must begin in this district at time $t = 0$ moving to the two neighboring districts with slightly lower $M_l(t)$ and moving to all other district pairs in order of decreasing $M_l(t)$, until the center ($l = 1$) and a district n ($n > k$) where $M_1(t) = M_n(t)$ are reached simultaneously at time $t_1^*(= t_j^*)$. Thereafter development moves to each slightly more outlying suburban district with slightly lower $M_l(t)$ until district $l = L$ is reached at time $t = T$.

We now expand on property 8 to derive the spatial variation in B_l^* and the land tax μ_l^*.

Property 12: B_l^* is highest for the district j where $M_j(t) = \max_l[M_l(t),\ l = 1, \ldots, L]$ and is successively lower for each next district with a slightly lower $M_l(t)$ schedule.

This follows from recalling property 11. The district with the highest $M_l(t)$ is developed first. Then, for the district with the next highest $M_l(t)$ to be developed next, B_l^* must be lowered as we can see from Figure 4.

This property too is translated into specific patterns for cases A and B.

Property 12A: For case A, B_l^* is highest for $l = 1$ and declines monotonically with distance across all other districts.

Property 12B: For case B, B_l^* is highest for $l = k$ where $M_k(t) = \max_l[M_l(t),\ l = 1, \ldots, L]$ and declines monotonically both towards the center and towards the urban border.

Both properties 12A and 12B follow from 12 and 11A, 11B.

Property 13A: For case A, μ_l^* is highest for $l = 1$ and declines monotonically with distance across all other districts.

Property 13B: For case B, μ_l^* could either decline monotonically with distance from the center, or it could first increase up to some intermediate district m and decrease monotonically thereafter. This district m is closer to the center than district k where $M_k(t) = \max_l[M_l(t),\ l = 1, \ldots, L]$.

To prove 13A, 13B we recall from (21a) that

$$\mu_l^* = B_l^*/Q_l^* + R_0/\alpha.$$

Differentiating,

$$d\mu_l^*/dx_l = -B_l^* Q_l^{-2}(dQ_l/dx_l) + Q_l^{-1}(dB_l^*/dx_l). \tag{37}$$

For case A $dB_l^*/dx_l < 0$ for all x_l. Thus both terms in (37) are negative implying

$$d\mu_l^*/dx_l < 0 \quad \text{for case A, all } x_l.$$

For case B we separately analyze $x_l > x_k$, $x_l = x_k$ and $x_l < x_k$. For $x_l > x_k$ the result is as in case A.

$$d\mu_l^*/dx_l < 0 \quad \text{for case B, all } x_l > x_k.$$

For $x_l < x_k$, $dB_l^*/dx_l > 0$ which implies a positive second term in (37) and

$$d\mu_l^*/dx_l \gtreqless 0 \quad \text{for case B, all } x_l < x_k.$$

Finally, $dB_l^*/dx_l = 0$ at $x_l = x_k$ by property 12B and thus,

$$d\mu_l^*/dx_l < 0 \quad \text{at } x_l = x_k.$$

This last fact implies that if μ_l^* peaks it does so between the center and district k.

One more concept needs to be explored. That is the spatial variation in the initial present value of urban land parcels (i.e. at $t = 0$). We define the initial present value of land in district l as

$$V_l^*(0) = V_l(t_l^*) = M_l(t_l^*)/Q_l + R_0/\alpha. \tag{38a}$$

From the complementary slackness condition we know that

$$V_l^*(0) = \mu_l^* + \lambda^*(t_l^*)/Q_l. \tag{38b}$$

Property 14A: Under case A the initial present value of land, $V_l^*(0)$, decreases monotonically with distance from the center.

Property 14B: Under case B the initial present value of land could either decrease monotonically with distance from the center, or it could first increase monotonically up to some intermediate district m and decrease monotonically thereafter. This district m is closer to the center thean district k where $M_k(t) = \max_l[M_l(t),$ $l = 1, \ldots, L]$.

To prove both 14A and 14B we differentiate (38b) with respect to t_l^*, the district development date:

$$dV_l^*(0)/dx_l = d\mu_l^*/dx_l + Q_l^{-1}(d\lambda^*(t_l^*)/dt_l^*)(dt_l^*/dx_l) - \lambda^*(t_l^*)Q_l^{-2}(dQ_l/dx_l). \tag{39}$$

For cases A and B we know that $d\lambda^*(t_l^*)/dt_l^* < 0$ and $dQ_l/dx_l > 0$. For case A, $d\mu_l^*/dx_l < 0$ by property 13A and $dt_l^*/dx_l > 0$ by property 11A. Thus, each term in (39) is negative, which implies

$$dV_l^*(0)/dx_l < 0 \quad \text{for case A, call } x_l.$$

For case B we must separately analyze the regions $x_l < x_k$ and $x_l > x_k$ and the point $x_l = x_k$ where $M_k(t) = \max_l[M_l(t), l = 1, \ldots, L]$. For $x_l > x_k$, $d\mu_l^*/dx_l < 0$ and $dt_l^*/dx_l > 0$. Thus all three terms are negative implying unambiguously that

$$dV_l^*(0)/dx_l < 0 \quad \text{for case B, } x_l > x_k.$$

For $x_l < x_k$, $d\mu_l^*/dx_l \gtreqless 0$, $dt_l^*/dx_l < 0$. Thus, the first term in (39) is ambiguous in sign while the second is positive and the third is negative. This implies

$$dV_l^*(0)dx_l \gtreqless 0 \quad \text{for case B, } x_l < x_k.$$

For $x_l = x_k$, $d\mu_l^*/dx_l = 0$ by property 13B and $dt_l^*/dx_l = 0$ by 11B. This implies

$$dV_l^*(0)/dx_l < 0 \quad \text{for case B, } x_l = x_k.$$

This last fact also means that it $V_l^*(0)$ peaks it must do so at some district closer to the center than district k.

Finally, one more property of the optimal solution must be proved:

Property 14: The development tax $\lambda^*(t)$ is equal to zero at the ultimate time T, i.e. $\lambda^*(T) = 0$. The land tax in the last district μ_L^* is equal to the nonurban value of land, i.e. $\mu_L^* = R_0/\alpha$.

To prove these consider the following argument. The land tax μ_L^* is the Lagrange multiplier (and shadow price) of land in L. If we increase the supply of land in L from A_L to $A_L + \varepsilon$, violating (12), then this excess ε land increment will be left vacant in the optimal solution since it is the least valuable in the urban space. The contribution of this undeveloped marginal land to the aggregate present value of the urban land will be equal to the nonurban value of land, i.e., R_0/α. Thus, $\mu_L^* = R_0/\alpha$ and $B_L^* = 0$. Since $M_L(t) = M_L(T) = 0$ it follows from (32), that $\lambda^*(T) = 0$. There is yet another way to prove $\lambda^*(T) = 0$. If we increase the population added at time T by an infinitesimal amount from $n(T)$ to $n(T) + \varepsilon$, then since the land is not available this increment of households will not be located, i.e. constraint (14a) for $t = T$ will hold as an inequality at the optimal solution. Since these households are not located their contribution to the objective function will be nill and thus the Lagrangian multiplier associated with the corresponding time period must be zero. Thus, $\lambda_T = 0$ and for the continuous time problem $\lambda(T) = 0$.

VI. IMPLICATIONS OF THIS ESSAY

An important implication of this essay is that slight variations in the assumptions of the model prohibit a long run equilibrium outcome which is consistent with the static model. Suppose, for example, that we relax the assumption of a static economy letting all exogenous variables be functions of time. We must then replace equation (1) with the following:

$$V(x, t^*, Q) = \int_0^{t^*} e^{-\alpha s} R_0(s)\, ds + \int_{t^*}^{\infty} e^{-\alpha s} R(x, s, Q)\, ds - e^{-\alpha t^*} C(t^*), \tag{40}$$

where $R_0(t)$ and $C(t)$ are exogenous functions of time and where

$$R(x, t, Q) = \{Y(t) - Z[\bar{U}(t), Q] - T(x, t)\}/Q, \tag{41}$$

with $Y(t)$, $\bar{U}(t)$ and $T(x, t)$ also exogenous functions of time. Properties 1 and 2 no longer apply.

First, it will be no longer true that each speculator will want to develop as soon as possible. It is very likely that even in the absence of taxes there will be an optimal holding period for each speculator depending on location. Graphically, the curves in Figure 1 need no longer be down-ward sloping but could achieve a maximum anywhere in $[0, \infty)$. If, for example, construction costs decrease rapidly over time speculators will want to wait for some time before developing in order to maximize the present value of their land.

Second, the optimal choice of lot size and development date are no longer separable as in property 2. In fact a different lot size will be optimal depending on the date of development. Since maximizing the value of land with respect to lot size is no longer equivalent to maximizing the household's static bid rent function (7), the urban form at some ultimate time need bear no resemblance to that of the static model.

The above considerations strongly suggest the restrictiveness of interpreting the static urban land use model as a model of long run equilibrium. It follows that the static model is a meaningful description of a new town or an "instant metropolis", but an incorrect description of any but the simplest form of urban growth only when crucial variables remain invariant over time. Even this case, however, has complex implications for the spatial pattern of urban development as was shown in this paper.

In summary, dynamic models which will be developed in the future will not in general be consistent with the static analyses of the past and will demonstrate clearly that static models lead to erroneous results when applied to long run equilibrium predictions. In fact, several models of urban growth developed up to now have demonstrated this by implication or through direct analysis. For two of these see Fujita (1976) and Anas (1978).

REFERENCES

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

Anas, A., 1978, Dynamics of Urban Residential Growth, *Journal of Urban Economics*, 1 (January).

Fujita, M., 1976, Spatial Patterns of Urban Growth, Optimum and Market, *Journal of Urban Economics*, 3, 209–242.

Wheaton, W. C., 1977, Residential Decentralization, Land Rents, and the Benefits of Urban Transportation Investment, *The American Economic Review*, (March), 136–144.

GEOGRAPHIA POLONICA 42, 1979

SPATIAL PATTERNS OF URBAN GROWTH AND CONTRACTION: PROBLEM A

Masahisa Fujita

Department of Regional Science, University of Pennsylvania, Philadelphia, USA

1. INTRODUCTION

In Fujita (1976b), the author proposed a framework for the development of dynamic theory of urban land use, and suggested five special problems of interest. And, in Fujita (1976a), one of these five problems, which is hereafter called Problem A, was solved under special assumptions.

The purpose of this paper is to study the solution of Problem A under more general assumptions. In particular, though it is assumed in Fujita (1976a) that the demand for buildings of each type is nondecreasing with respect to time, this assumption is abandoned in this paper. As will be explained in section 7, this generalization is also necessary to obtain the solution of Problem C in Fujita (1976b) by utilizing the solution of Problem A.

In this paper, we study only the optimal planning problem of urban land use within the framework of Problem A. But, note that, for each optimal planning problem within the framework of Problem A, there exists a corresponding market equilibrium problem which has the same solution (see Proposition 1 in Fujita, 1976b). Hence, the reader must keep in mind that each characteristic of the solution which we obtain in this paper is also a characteristic of the solution for an appropriate market equilibrium problem.

In section 2, we formulate a dynamic planning problem of urban land use (Problem A), and obtain the optimality conditions for that problem by applying the maximum principle in optimal control theory. Next, in section 3, we study some basic properties of the solution for Problem A by introducing the concepts of bid land prices and building utilization plans. Then, in section 4, we obtain optimal utilization plans of buildings by introducing the notion of occupancy set, and examine how slopes of bid land price curves change with time. Next, in section 5, assuming that the city is to be composed of only two types of buildings, we study in detail optimal spatial patterns of urban growth and contraction under various specifications of building demand functions. Finally, in section 6, we briefly examine the case of many building types.

To avoid repetitions, explanations of statements and procedures of analyses are as simplified as possible whenever they can be found in Fujita (1976a) or (1976b).

2. PROBLEM FORMULATION

Suppose a city is to be developed on an isolated featureless plain. We divide the area into g districts, and let

d_l: the distance of district l from the predetermined center of the city,

s_l: the size of district l,

and we assume

$$d_1 < d_2 < \ldots < d_l < \ldots < d_g,$$

$$s_l > 0, \quad l = 1, 2, \ldots, g.$$

Since g can be arbitrarily large, locations within each district are considered to be homogeneous, and the total area $\sum_{l=1}^{g} s_l$ is sufficiently large so that there will be no spatial constraint on the size of the whole city in the future.

The city is to consist of m types of buildings each of which is identified not only by its structure but also by its lot size; and let

$x_{il}(t)$: the number of buildings of type i existing in district l at time t,

$u_{il}(t)$: the number of buildings of type i constructed in district l at time t,

k_i: the lot size of a building of type i which is a constant.

The destruction of buildings is not considered in this paper, and hence $u_{il}(t) \geqslant 0$.

Next, we divide all the building-using activities (such as business activities, commercial activities and households) into n types depending on their characteristics. And we assume in this paper that there is one-to-one correspondence between activity types and building types. That is, we assume activities of type i use only buildings of type i. Thus, index i represents building type i as well as activity type i, and we have $m = n$. We further assume that the total number of behavioral units belonging to each activity type is exogenously given at each time, and let

$N_i(t)$: the number of behavioral units belonging to activity type i at time t,

$y_{il}(t)$: the number of behavioral units of activity type i allocated to buildings of type i in district l at time t.

The objective function employed in this paper is

$$\int_0^\infty e^{-\gamma t}\Big[\sum_i \sum_l \Psi_{il}(t) y_{il}(t) - \sum_i \sum_l B_i(t) u_{il}(t)\Big] dt,$$

where

$B_i(t)$: the construction costs per unit of buildings of type i at time t,

$\Psi_{il}(t)$: the monetary value attached to a unit of building services of type i when it is occupied by activities of type i in district l at time t,

γ: the time-discount rate of the objective value.

Using the above assumptions and notation, a normative problem of urban land use can be summarized as follows.

OPTIMUM PROBLEM A (OPA)

Choose values of construction speed $u_{il}(t)$ and numbers of activity units $y_{il}(t)$ $(\overline{i} = 1, 2, ..., m)$ at each time t so as to maximize the objective function

$$\int_0^\infty e^{-\gamma t}\Big[\sum_i \sum_l \Psi_{il}(t) y_{il}(t) - \sum_i \sum_l B_i(t) u_{il}(t)\Big] dt,$$

subject to the following restrictions.

a) variation of building stock

$$\dot{x}_{il}(t) = u_{il}(t), \quad u_{il}(t) \geqslant 0, \quad i = 1, 2, ..., m, \quad l = 1, 2, ..., g,$$

b) building stock constraint

$$0 \leqslant y_{il}(t) \leqslant x_{il}(t), \quad i = 1, 2, ..., m, \quad l = 1, 2, ..., g,$$

c) total number of behaving-units constraint

$$\sum_l y_{il}(t) = N_i(t), \quad i = 1, 2, ..., m,$$

d) land constraint

$$\sum_i k_i x_{il}(t) \leqslant s_l, \quad l = 1, 2, ..., g,$$

e) initial condition

$$x_{il}(0) = \bar{x}_{il}(0), \quad i = 1, 2, ..., m, \quad l = 1, 2, ..., g.$$

In the above problem, $\dot{x}_{il}(t) = dx_{il}(t)/dt$. And, in the following, we always assume that

Assumption OPA-1. $s_l > 0$, $k_i > 0$, $\gamma > 0$; $B_i(t) > 0$ and $N_i(t) \geqslant 0$ for all $t \geqslant 0$; $B_i(t)$ and $N_i(t)$ are continuously differentiable with respect to t; $de^{-\gamma t}B_i(t)/dt < 0$ for all $t \geqslant 0$; $\max\{N_i(t) | \ t \in [0, \infty)\} < \infty$ and $\sum_i k_i \max\{N_i(t) | \ t \in [0, \infty)\} < \sum_l s_l$, for all $i = 1, 2, ..., m$ and $l = 1, 2, ..., g$.

Assumption OPA-2. Each $\Psi_{il}(t)$ is piecewise continuously differentiable with respect to t, and $\lim_{t\to\infty} e^{-\gamma t}\Psi_{il}(t) = 0$, where $i = 1, 2, ..., m$ and $l = 1, 2, ..., g$.

For convenience we set

$$b_i(t) = e^{-\gamma t}B_i(t), \quad \psi_{il}(t) = e^{-\gamma t}\Psi_{il}(t). \tag{2.1}$$

Then, under the above two assumptions, the optimality conditions for problem OPA can be summarized as follows by applying the maximum principle in optimal control theory.[1]

---

[1] For the maximum principle relevant to OPA, see Theorems 23 and 24 in Pontryagin et al. (1962), Theorem 2.1 in Chapter 8 of Hestenes (1966), section 16 of Boltyanskii (1968), section II.6 of Arrow and Kurz (1970), and Halkin (1974).

OPTIMALITY CONDITION FOR PROBLEM A (OCA)

For $u_{il}(t)$ and $y_{il}(t)$ ($i = 1, 2, \ldots, m$, $l = 1, 2, \ldots, g$, $t \geqslant 0$) to be optimal construction and allocation processes for OPA, it is necessary and sufficient that there exist $x_{il}(t)$ and auxiliary variables $r_{il}(t)$, $q_l(t)$, $p_{il}(t)$, $p_l(t)$ ($i = 1, 2, \ldots, m$, $l = 1, 2, \ldots, g$, $0 \leqslant t$) such that

(i) rental market equilibrium condition: for $i = 1, 2, \ldots, m$ and $l = 1, 2, \ldots, g$,

$$r_{il}(t) = \max\{\overline{\psi}_{il}(t)+q_l(t), 0\},$$

$$(r_{il}/(t)-\overline{\psi}_{il}(t)-q_l(t))y_{il}(t) = 0,$$

$$0 \leqslant y_{il}(t) \leqslant x_{il}(t),$$

$$(x_{il}(t)-y_{il}(t))r_{il}(t) = 0,$$

$$\sum_l y_{il}(t) = N_i(t),$$

(ii) construction market equilibrium condition

$$p_{il}(t) \leqslant b_i(t)+k_i p_l(t), \quad i = 1, 2, \ldots, m, \quad l = 1, 2, \ldots, g,$$

$$p_{il}(t) = b_i(t)+k_i p_l(t) \text{ if } u_{il}(t) > 0,$$

(iii) asset market equilibrium condition

(iii-1) building

$$\dot{p}_{il}(t) \leqslant -r_{il}(t), \quad i = 1, 2, \ldots, m, \quad l = 1, 2, \ldots, g,$$

$$\dot{p}_{il}(t) = -r_{il}(t) \text{ if } x_{il}(t) > 0,$$

(iii-2) land

$$\dot{p}_l(t) \leqslant 0, \quad l = 1, 2, \ldots, g,$$

$$\dot{p}_l(t) = 0 \text{ when } \sum_i k_i x_{il}(t) < s_l,$$

(iv) variation of building stock and land constraint

$$\dot{x}_{il}(t) = u_{il}(t), \quad u_{il}(t) \geqslant 0, \quad i = 1, 2, \ldots, m, \quad l = 1, 2, \ldots, g,$$

$$\sum_i k_i x_{il}(t) \leqslant s_l, \quad l = 1, 2, \ldots, g,$$

(v) transversality condition: for $i = 1, 2, \ldots, m$ and $l = 1, 2, \ldots, g$,

(v-1) initial

$$x_{il}(0) = \overline{x}_{il}(0),$$

(v-2) terminal

$$\lim_{t\to\infty} p_{il}(t) = 0, \quad \lim_{t\to\infty} p_l(t) = 0.$$

To make easier the economic interpretation of the above optimality condition, let us introduce the following notation:

$$R_{il}(t) = e^{\gamma t} r_{il}(t),\ Q_j(t) = e^{\gamma t} q_j(t),$$
$$P_{il}(t) = e^{\gamma t} p_{il}(t), \text{ and } P_l(t) = e^{\gamma t} p_l(t). \tag{2.2}$$

Then, using (2.1) and (2.2), OCA can be restated as follows.

OPTIMALITY CONDITION FOR PROBLEM A (OCA*)

(i) rental market equilibrium condition: for $i = 1, 2, \ldots, m$ and $l = 1, 2, \ldots, g$,

$$R_{il}(t) = \max\{\Psi_{il}(t)+Q_j(t), 0\},$$
$$(R_{il}(t)-\Psi_{il}(t)-Q_j(t))y_{il}(t) = 0,$$
$$0 \leqslant y_{il}(t) \leqslant x_{il}(t),$$
$$(x_{il}(t)-y_{il}(t))R_{il}(t) = 0,$$
$$\sum_l y_{il}(t) = N_i(t),$$

(ii) construction market equilibrium condition

$$P_{il}(t) \leqslant B_i(t)+k_i P_l(t), \quad i = 1, 2, \ldots, m, \quad l = 1, 2, \ldots, g,$$
$$P_{il}(t) = B_i(t)+k_i P_l(t) \quad \text{if} \quad u_{il}(t) > 0,$$

(iii) asset market equilibrium condition

(iii-1) building

$$\dot{P}_{il}(t) \leqslant \gamma P_{il}(t)-R_{il}(t), \quad i = 1, 2, \ldots, m, \quad l = 1, 2, \ldots, g,$$
$$\dot{P}_{il}(t) = \gamma P_{il}(t)-R_{il}(t) \quad \text{if} \quad x_{il}(t) > 0,$$

(iii-2) land

$$\dot{P}_l(t) \leqslant \gamma P_l(t), \quad l = 1, 2, \ldots, g,$$
$$\dot{P}_l(t) = \gamma P_l(t) \quad \text{if} \quad \sum_i k_i x_{il}(t) < s_l,$$

(iv) variation of of building stock and land constraint

$$\dot{x}_{il}(t) = u_{il}(t), \quad u_{il}(t) \geqslant 0, \quad i = 1, 2, \ldots, m, \quad l = 1, 2, \ldots, g,$$
$$\sum_i k_i x_{il}(t) \leqslant s_l, \quad l = 1, 2, \ldots, g,$$

(v) transversality condition: for $i = 1, 2, \ldots, m$ and $l = 1, 2, \ldots, g$,

(v-1) initial

$$x_{il}(0) = \bar{x}_{il}(0),$$

(v-2) terminal

$$\lim_{t\to\infty} e^{-\gamma t} P_{il}(t) = 0, \quad \lim_{t\to\infty} e^{-\gamma t} P_l(t) = 0.$$

The economic meanings of the above optimality conditions become clear if we compare them with the equilibrium conditions for the competitive market problem associated with OPA. Consider the urban land market consisting of building renters of each type i ($i = 1, 2, \ldots, m$), builders of buildings, owners of buildings, land owners, building speculators and land speculators. Then, it can be shown that OCA* constitutes a part of the equilibrium condition for that urban land market.[2] The title attached to each set of conditions in OCA* (and OCA) has been derived from

[2] For details, see section 2 of Fujita (1976b).

that market equilibrium condition. And, auxiliary variables $R_{il}(t)$, $Q_j(t)$, $P_{il}(t)$ and $P_l(t)$ in OCA* represent, respectively, building rent, subsidy (or, tax) for the building rent, building price and land price at time t. Hence, variables $r_{il}(t)$, $q_i(t)$, $p_{il}(t)$ and $p_l(t)$ in OCA represent discounted values of these price variables, respectively. Thus, OCA states the optimality condition in terms of discounted prices, and OCA* in terms of current prices.

Next, observe in OPA that, for any given set of functions $\eta_i(t)$, $i = 1, 2, \ldots, n$, $t \geqslant 0$, we have

$$\int_0^\infty e^{-\gamma t}\Big[\sum_l\sum_i(\Psi_{il}(t)+\eta_i(t))y_{il}(t)-\sum_l\sum_i B_i(t)u_{il}(t)\Big]dt$$

$$=\int_0^\infty e^{-\gamma t}\Big[\sum_l\sum_i\Psi_{il}(t)y_{il}(t)-\sum_l\sum_i B_i(t)u_{il}(t)\Big]dt+\int_0^\infty e^{-\gamma t}\eta_i(t)N_i(t)dt \quad \text{from (c) of OPA.}$$

Hence, we get

Proposition A.1. If a set of functions. $u_{il}(t)$, $x_{il}(t)$ and $y_{il}(t)$ ($i = 1, 2, \ldots, m$, $l = 1, 2, \ldots, g$, $0 \leqslant t < \infty$) represents a solution for OPA with parameters $\overline{\Psi}_{il}(t)$ ($i = 1, 2, \ldots, m$, $l = 1, 2, \ldots, g$, $0 \leqslant t < \infty$), it is also a solution for OPA with parameters $\overline{\Psi}_{il}(t)+\eta_i(t)$ ($i = 1, 2, \ldots, m$, $l = 1, 2, \ldots, g$, $0 \leqslant t < \infty$) where each $\eta_i(t)$ is any piecewise continuously differentiable function of t.

That is, the solution for OPA is independent of the absolute level of "bid rent", $\overline{\Psi}_{il}(t)$, for each renter type i.

3. BID RENT CURVES AND BID PRICE CURVES

In the rest of paper, we study the characteristics of the solution for OPA.

3.1. BID RENT AND BID RENT CURVE

It must first be noted that we do not have land rent in our problem. That is, since it is assumed in OPA that land is serviceable only when buildings are constructed on it and buildings cannot be destructed, any piece of land on which buildings are once constructed never appear again in the land market. Hence, in this paper, *rent* means *building rent*.

In the context of OCA, we set

$$\psi_{il}(q_i, t) = \overline{\psi}_{il}(t)+q_i, \tag{3.1}$$

and call $\psi_{il}(q_i, t)$ the (discounted) bid *rent* (of renter type i) for buildings of type i in district l. This is a function of *rent-subsidy* q_i for renter type i.

From (i) of OCA, we obtain the following relationships between rents and bid rents (on the optimal path).

Property A.1. Rent $r_{il}(t)$ is nonnegative, and is equal to bid rent $\psi_{il}(q_i(t), t)$ when the latter is positive.[3]

Property A.2. Renters of type i can occupy buildings of type i in district l only when their bid rent $\psi_{il}(q_i(t), t)$ is equal to building rent $r_{il}(t)$ in that district.

Property A.3. If there is vacancy in buildings of type i in district l, rent $r_{il}(t)$ is zero.

We often use the following assumption on parameter function $\overline{\Psi}_{il}(t)$.

Assumption OPA-3. For every $i = 1, 2, \ldots, m$ and $t \geqslant 0$,

a) $\overline{\Psi}_{il}(t)$ is decreasing with respect to d_l, or

b) $\overline{\Psi}_{il}(t) = \overline{\Psi}_i(t) - A_i(t)d_l$, where $A_i(t) > 0$ and $A_i(t)$ is continuous with respect to t.

In assumption (b) above, we often use the following expression for convenience.

$$A_i(t) = A_i\alpha_i(t) \text{ where } A_i = A_i(0),\ \alpha_i(0) = 1,\ \alpha_i(t) > 0$$
$$\text{for all } t, \text{ and } \alpha_i(t) \text{ is continuous with respect to } t. \tag{3.2}$$

Next suppose that a building of type i is occupied in a district l' at time t (i.e., $y_{il'}(t) > 0$). Then, from Properties A.1 and A.2, $r_{il'}(t) = \psi_{il'}(q_i(t), t) \geqslant 0$. Hence, from definition (3.1) and Property A.1, under either (a) or (b) in Assumption OPA-3, it must be true that $r_{il}(t) = \psi_{il}(q_i(t), t) > 0$ for every $l < l'$. Therefore, from Property A.3, $x_{il}(t) = y_{il}(t)$ for every $l < l'$.

Conversely, suppose that a building of type i in a district l' is vacant at time t (i.e., $x_{il'}(t) > y_{il'}(t)$). Then, from Properties A.1 and A.3, $0 = r_{il'}(t) \geqslant \psi_{il'}(q_i(t), t)$. Therefore, under either one of (a) or (b) in Assumption OPA-3, $\psi_{il}(q_i(t), t) < 0$ for every $l > l'$, which implies from Properties A.1 and A.2 that $y_{il}(t) = 0$ for every $l > l'$. Threfore, we conclude that

Property A.4. Under either (a) or (b) of Assumption OPA-3, buildings of each type are utilized in order of distance from the city center at each time. That is, when a building of a certain type is occupied in a district, all the buildings of that type are occupied in all the districts closer than that district (to the city center); when a building of a certain type is vacant in a district, all the buildings of that type are vacant in all the districts further than that district.

3.2. BID LAND PRICES AND LAND PRICE CURVES

Let $\delta_t^i(\tau)$ be an arbitrary function of time parameter $\tau (t \leqslant \tau < \infty)$ of which value is either 0 or 1 at each τ. We call function $\delta_t^i(\tau)$ a *utilization plan* of a building of type i which will be constructed at time t.

The implication of function $\delta_t^i(\tau)$ is

$$\delta_t^i(\tau) = \begin{cases} 0\text{: the building is vacant at time } \tau, \\ 1\text{: the building is occupied at time } \tau. \end{cases}$$

Notice that we intentionally do not attach index l to function $\delta_t^i(\tau)$. That is, function $\delta_t^i(\tau)$ is independent of district.

[3] In the following, Property A. α means a property of the optimal path for problem OPA.

Next, given a rent subsidy function $q_i(\tau)$ $(0 \leqslant \tau < \infty)$, bid rent function $\psi_{il}(q_i(\tau), \tau)$ $(0 \leqslant \tau < \infty)$ is obtained from (3.1). For simplicity, we often represent $\psi_{il}(q_i(\tau), \tau)$ by $\psi_{il}(\tau)$. That is,

$$\psi_{il}(\tau) = \psi_{il}(q_i(\tau), \tau) = \overline{\psi}_{il}(\tau) + q_i(\tau). \tag{3.3}$$

Then, given a bid rent function $\psi_{il}(\tau)$ $(0 \leqslant \tau < \infty)$ and a utilization plan $\delta_t^i(\tau)$ $(t \leqslant \tau < \infty)$, we set

$$p_l^i(t|\ \psi_{il}(\tau),\ \delta_t^i(\tau)) = \frac{\int_t^\infty \psi_{il}(\tau)\delta_t^i(\tau)d\tau - b_i(t)}{k_i}, \tag{3.4}$$

and call function p_l^i the ***bid land price function*** for constructors of buildings of type i in district l. Function $p_l^i(t|\ \psi_{il}(\tau),\ \delta_t^i(\tau))$ shows how much net revenue will result from a unit of land after time t if buildings of type i are constructed in district l and the utilization of each building is $\delta_t^i(\tau)$ $(t \leqslant \tau < \infty)$, and if the supposed level of bid rent is $\psi_{il}(\tau)$ $(t \leqslant \tau < \infty)$.

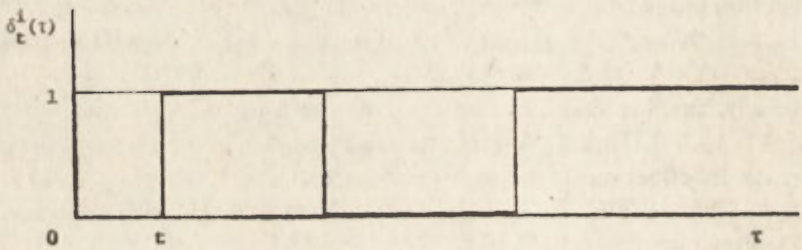


Fig. 1. Building utilization plan $\delta_t^i(\tau)$.

The relation between bid land prices and actual land prices can be summarized as follows.

Property A.5. Suppose that $p_l(t)$ $(l = 1, 2, \ldots, g, 0 \leqslant t < \infty)$ and $q_i(t)$ $(i = 1, 2, \ldots, m, 0 \leqslant t < \infty)$ are functions in OCA corresponding to the solution for problem OPA. Then, in each district l $(l = 1, 2, \ldots, g)$ and at each time t $(0 \leqslant t < \infty)$, we have

$$p_l(t) \geqslant p_l^i\big(t|\ \psi_{il}(q_i(\tau), \tau),\ \delta_t^i(\tau)\big) \tag{3.5}$$

for any building type i and for any utilization plan $\delta_t^i(\tau)$ $(t \leqslant \tau < \infty)$. And, if buildings of type i are being constructed in district l at tiem t on the solution path for OPA, and if $\delta_t^i(\tau)$ $(t \leqslant \tau < \infty)$ represents the utilization of one of these buildings on that solution, then we have

$$p_l(t) = p_l^i\big(t|\ \psi_{il}(q_i(\tau), \tau),\ \delta_t^i(\tau)\big), \tag{3.6}$$

where $\delta_t^i(\tau) = 1$ when that building is occupied at time τ and $\delta_t^i(\tau) = 0$ if not occupied.

Proof. From (i), (ii) and (iii-1) of OCA, we have for each i, l, τ and t that

$$r_{il}(\tau) = \max\{\overline{\psi}_{il}(\tau)+q_i(\tau), 0\} = \max\{\psi_{il}(q_i(\tau), \tau), 0\}, \tag{a}$$

$$\dot{p}_{il}(\tau) \leqslant -r_{il}(\tau), \tag{b}$$

$$p_l(t) \geqslant (p_{il}(t)-b_i(t))/k_i. \tag{c}$$

Hence, recalling (v-2) of OCA, we get

$$\begin{aligned} p_l(t) &\geqslant (p_{il}(t)-b_i(t))/k_i \\ &= \Big(\int_t^\infty -\dot{p}_{il}(\tau)d\tau - b_i(t)\Big)/k_i \\ &\geqslant \Big(\int_t^\infty r_{il}(\tau)d\tau - b_i(t)\Big)/k_i \\ &= \Big(\int_t^\infty \max\{\psi_{il}(q_i(\tau), \tau), 0\}d\tau - b_i(t)\Big)/k_i \\ &\geqslant \Big(\int_t^\infty \psi_{il}(q_i(\tau), \tau)\delta_t^i(\tau)d\tau - b_i(t)\Big)/k_i \qquad \text{for any } \delta_t^i(\tau) \\ &= p_l^i(t|\ \psi_{il}(q_i(\tau), \tau), \delta_t^i(\tau)) \qquad \text{for any } \delta_t^i(\tau). \end{aligned}$$

Next, from (ii) of OCA, if $u_{il}(t) > 0$, then (c) is satisfied with equality. And, if $u_{il}(t) > 0$ for $t \in (t-\varepsilon, t+\varepsilon)$ for some $\varepsilon > 0$, then $x_{il}(t) > 0$; and hence, from (iii-1) of OCA, relation (b) is satisfied with equality. Thus, if we set

$$\delta_t^i(\tau) = \begin{cases} 1 & \text{if } r_{il}(\tau) = \psi_{il}(q_i(\tau), \tau), \\ 0 & \text{if } r_{il}(\tau) > \psi_{il}(q_i(\tau), \tau), \end{cases} \tag{3.7}$$

then we obtain

$$p_l(t) = p_l^i\big(t|\ \psi_{il}(q_i(\tau), \tau), \delta_t^i(\tau)\big).$$

Hence, recalling Properties A.2 and A.3, we see the latter half of Property A.5 holds to be true. Q.E.D.

Given a set of bid rent functions $\psi_{il}(\tau)$ $(l = 1, 2, \ldots, g,\ 0 \leqslant \tau < \infty)$ and a utilization plan $\delta_t^i(\tau)$ $(t \leqslant \tau < \infty)$ at time t, bid land price $p_l^i(t|\ \psi_{il}(\tau), \delta_t^i(\tau))$ can be calculated from (3.4) for each district l $(l = 1, 2, \ldots, g)$. Hence, we obtain a curve on the urban space, and this curve is called a *bid land price curve.* By using this term, Property A.5 can be restated as follows.

Property A.6. Let $q_i(\tau)$, $0 \leqslant \tau < \infty$, be the rent subsidy function corresponding to the solution of problem OPA $(i = 1, 2, \ldots, m)$. Then, at any time t, bid rent curve $p_l^i\big(t|\ \psi_{il}(q_i(\tau), \tau), \delta_t^i(\tau)\big)$ is never above the land price curve $p_l(t)$ in any district for any building type i under any utilization plan $\delta_t^i(\tau)$. When buildings of type i are being constructed in a district at time t and function $\delta_t^i(\tau)$ $(t \leqslant \tau < \infty)$ represents the utilization of one of these buildings, then bid rent curve $p_l^i\big(t|\ \psi_{il}(q_i(\tau), \tau), \delta_t^i(\tau)\big)$ touches the land price curve $p_l(t)$ from below in that district.

Next, from (iii-2) and (v-2) of OCA, we have

Property A.7. Land price $p_l(t)$ never increases with time.

Property A.8. Land price $p_l(t)$ does not decrease while there remains vacant land in that district.

Property A.9. Land price $p_l(t)$ is zero at the terminal time, that is, $p_l(\infty) = 0$ for every l.

And, from Properties A.6 to A.7, we get

Property A.10. Suppose buildings of type i are being constructed in a district at time t. Then, the bid rent curve $p_l^i\big(t|\ \psi_{il}(q_i(\tau), \tau), \delta_t^i(\tau)\big)$ of building type i touches both the initial land price curve $p_l(0)$ and land price curve $p_l(t)$ from below in that district; and curve $p_l(t)$ lies between the two curves, $p_l(0)$ and $p_l^i\big(t|\ \psi_{il}(q_i(\tau), \tau), \delta_t^i(\tau)\big)$. Here $q_l(\tau)$ $(0 \leqslant \tau < \infty)$ and $\delta_t^i(\tau)$ $(t \leqslant \tau)$ are, respectively, the rent subsidy function for demand type i (corresponding to the solution of OPA) and the optimal utilization plan for one of these buildings.

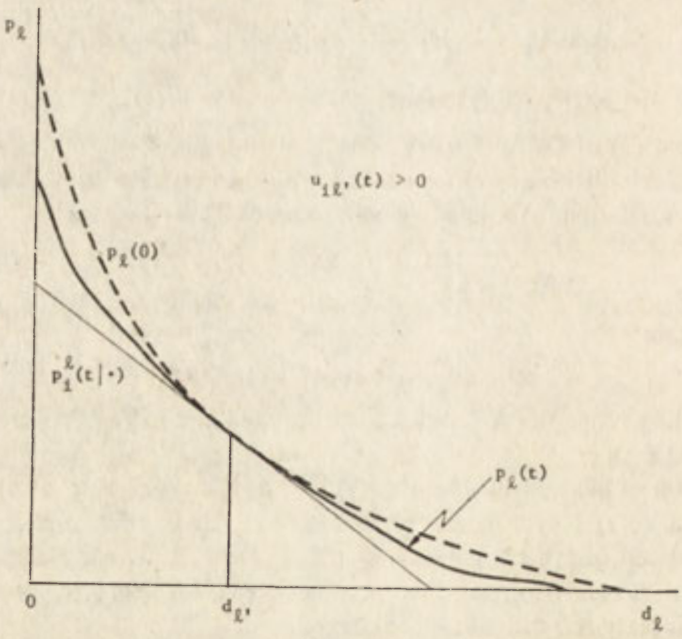


Fig. 2. Relation between initial land price curve $p_l(0)$, land price curve $p_l(t)$ and bid land price curve $p_l^i(t|\cdot)$ corresponding to the optimal path

From the above property, the relation between the three curves, $p_l(0), p_l(t)$ and $p_l^i\big(t|\ \psi_{il}(q_i(\tau), \tau), \delta_t^i(\tau)\big)$ can be depicted as in Figure 2. And, from Properties A. 8 and A.9, we see that

Property A.11. When a district has vacant land throughout the plan period (i.e., for all $t \geqslant 0$), the land price in that district is zero from the beginning.

Therefore, combining A.6 to A.11, we obtain

Property A.12. Let $q_l(\tau)$ $(0 \leqslant \tau < \infty)$ be the rent subsidy function corresponding to the solution of problem OPA. Then,

$$\text{if } \sum_i k_i x_{il}(0) < s_l, \text{ then} \quad p_l(0) = \max\{\max_{i,t,\delta_t^i(\tau)} p_l^i\big(t \mid \psi_{il}(q_l(\tau), \tau), \delta_t^i(\tau)\big), 0\}. \tag{3.8}$$

That is, if a district l has some vacant land at the initial time $t = 0$, then the land price $p_l(0)$ is equal to the maximum of all the bid land prices $p_l^i\big(t \mid \psi_{il}(q_l(\tau), \tau), \delta_t^i(\tau)\big)$ for all $i = 1, 2, \ldots, m$, $t \in [0, \infty)$, and for all possible $\delta_t^i(\tau)$; or, equal to zero when that maximum value is nonpositive. On the other hand,

$$\text{if } \sum_{i=1}^{m} k_i \bar{x}_{il}(0) = s_l, \text{ then} \quad p_l(0) \geqslant \max\{\max_{i,t,\delta_t^i(\tau)} p_l^i\big(t \mid \psi_{il}(q_l(\tau), \tau), \delta_t^i(\tau)\big), 0\}. \tag{3.9}$$

That is, if a district l has no vacant land initially, any price which is no less than the maximum value of all the bid rent curves (or, zero) can serve as an initial land price (corresponding to the solution path of OPA in that district). Therefore, as a special case of the above property, we have

Property A.13. Suppose that every district has vacant land at time $t = 0$. Then, the initial land price curve $p_l(0)$ coincides with the upper envelope of all the bid land price curves $p_l^i\big(t \mid \psi_{il}(q_l(\tau), \tau), \delta_t^i(\tau)\big)$ for all $i = 1, 2, \ldots, m$, $t \in [0, \infty)$ and for all possible $\delta_t^i(\tau)$. Here, $q_l(\tau)$ $(0 \leqslant \tau < \infty)$ is the rent subsidy function corresponding to the solution of OPA.

Next, from the definition of bid land prices by (3.4) and from the nature of integral calculus, it is clear that

Property A.14. Under (a) (or (b)) of Assumption OPA-3, any bid land price curve $p_l^i(t \mid \psi_{il}(\tau), \delta_t^i(\tau))$ is decreasing with respect to distance d_l. And, it is convex with respect to d_l when function $\psi_{il}(\tau)$ is convex with respect to d_l at every time τ $(0 \leqslant \tau < \infty)$.[4] In particular, under (b) of Assumption OPA-3,

$$\partial p_l^i(t \mid \psi_{il}(\tau), \delta_t^i(\tau))/\partial d_l = -\frac{a_l}{k_i}\sigma_i(\delta_t^i(\tau)), \tag{3.10}$$

where

$$\sigma_i(\delta_t^i(\tau)) = \int_t^\infty e^{-\gamma\tau}\alpha_i(\tau)\delta_t^i(\tau)d\tau, \tag{3.11}$$

and hence each $p_l^i(t \mid \psi_{il}(\tau), \delta_t^i(\tau))$ is a straight line with a downward slope. Therefore, combining the above property with Property A.13, we obtain

[4] Let $f_l(\cdot)$ be a function defined for each district l $(l = 1, 2, \ldots, g)$. Then, by the statements, "$f_l(\cdot)$ is convex with respect to d_l" or "$f_l(\cdot)$ is strictly convex with respect to d_l," we mean, respectively, that

$$f_l(\cdot) \geqq \alpha_l f_{l'}(\cdot) + (1-\alpha_l) f_{l''}(\cdot) \quad \text{for every } l'' < l < l',$$

or

$$f_l(\cdot) > \alpha_l f_{l'}(\cdot) + (1-\alpha_l) f_{l''}(\cdot) \quad \text{for every } l'' < l < l',$$

where $\alpha_l = (d_l - d_{l''})/(d_{l'} - d_{l''})$.

Property A.15. Suppose that every district has vacant land at time $t = 0$. And denote by $\bar{l}$ the nearest district to the city center among all the districts in which $p_l(0) = 0$. Then, under (a) (or (b)) of Assumption OPA-3, initial land price curve $p_l(0)$ is decreasing between $l = 1$ and $l = \bar{l}$, and is zero for all $l \geqslant \bar{l}$. In particular, under (b) of Assumption OPA-3, land price curve $p_l(0)$ is convex on the interval between $l = 1$ and $l = \bar{l}$.

Finally, from (iii-1) and (v-2) of OCA, we obtain the following properties of building prices.

Property A.16. Building price $p_{il}(t)$ should decrease at least as much as rent $r_{il}(t)$ at each time, and it should decrease exactly as much as rent $r_{il}(t)$ when there are buildings of that type in that district.

Property A.17. Building price $p_{il}(t)$ is zero at the terminal time, that is $p_{il}(\infty) = 0$ for every i and every l.

4. OPTIMAL UTILIZATION PLANS OF BUILDINGS AND SLOPES OF BID LAND PRICE CURVES

Suppose that buildings of type i are being constructed in district l at time t. Then, from Property A.10, it must be true that

$$p_l(0) = p_l(t) = \frac{\int_t^{\infty} \psi_{il}(q_i(\tau), \tau)\delta_t^i(\tau)d\tau - b_i(t)}{k_i}. \tag{4.1}$$

And, for building to be constructed in district l at time t, there should remain some vacant land there. Hence, from Properties A.7 and A.8,

$$\dot{p}_l(t) = 0.$$

From (4.1), the following relation must hold to be true.

$$\psi_{il}(q_i(t), t)\delta_t^i(t) = -\dot{b}_i(t).$$

On the other hand, recall from Assumption OPA-1 that we always assume in this paper

$$\dot{b}_i(t) < 0 \quad \text{for all } t \geqslant 0. \tag{4.2}$$

Hence, it must be true that

$$\psi_{il}(q_i(t), t) = -\dot{b}_i(t) > 0 \quad \text{and} \quad \delta_t^i(t) = 1. \tag{4.3}$$

Therefore, recalling Properties A.1 to A.2, we conclude that

Property A.18. When buildings of type i are being constructed in a district $l_i(t)$ at time t, then the following relations must hold to be true.

(i) $r_{il}(t) = -\dot{b}_i(t) > 0$ for $l = l_i(t)$.

(ii) There are no vacant buildings of type i in district $l_i(t)$ at time t.

(iii) No building (of type i which is constructed in district $l_i(t)$ at time t) can be vacant immediately following its construction.

Hereafter, we denote by $l_i(t)$ the district, if it exists, in which buildings of type i are being constructed at time t. Then, from (i) of Property A.18 and Properties A.2 and A.4, we get

Property A.19. Under (a) (or (b)) of Assumption OPA-3, if it exists, the district of construction, $l_i(t)$, for each building type i is unique at each time t.
And, from Property A.4 and (ii) of Property A.18, we see that

Property A.20. Under (a) (or (b)) of Assumption OPA-3, when buildings of type i are being constructed in a district, $l_i(t)$, at time t, there are no vacant buildings of type i in districts l, $1 \leqslant l \leqslant l_i(t)$, at that time.
It can also be shown that

Property A.21. Under (a) (or (b)) of Assumption OPA-3, the construction district $l_i(t)$ moves outwards at each time (if it exists). That is, if both $l_i(t')$ and $l_i(t'')$ exist, then

$$l_i(t') \leqslant l_i(t'') \quad \text{whenever } t' < t''.$$

The validity of the above property can intuitively be demonstrated by the following counterexample.[5] Suppose, on the contrary, we have the following relation on a construction plan.

$$l_i(t') = l' > l_i(t'') = l'' \quad \text{and} \quad t' < t''.$$

Then, instead of this construction plan, we employ a new construction plan which is exactly the same as the original plan except for the following point. That is, on this new construction plan, we set

$$l_i(t') = l'' \quad \text{and} \quad l_i(t'') = l',$$

and keep the rest of the plan exactly the same as the original plan. Then, since we have under (a) (or (b)) of Assumption OPA-3 that $\bar{\psi}_{il''}(t) > \bar{\psi}_{il'}(t)$ at any t and since $t' < t''$, the value of the objective function in OPA should be greater under the new plan than under the original plan. Thus, the original plan cannot be optimal.

In the rest of this paper, we study the spatial patterns of urban growth and contraction under the following assumption.

Assumption OPA-4. $\bar{x}_{il}(0) = 0$ for all $i = 1, 2, ..., m$ and $l = 1, 2, ..., g$. That is, no building exists at time $t = 0$.[6]

Under this assumption, each optimal utilization plan $\delta_i^l(\tau)$ takes a simple form as shown below.

Suppose that buildings of type i are being constructed in district $l_i(t)$ at time t (on the optimal path). Then, if we adopt (a) (or (b)) of Assumption OPA-3, this implies from Property A.21 that there are no buildings of type i in those districts beyond $l_i(t)$ at time t (i.e., $x_{il}(t) = 0$ for every $l > l_i(t)$). And, from Property A.20, there is no vacant building of type i in those districts l where $l \leqslant l_i(t)$. Therefore, it must be true that

$$\sum_{l=1}^{g} x_{il}(t) = N_i(t) \quad \text{if} \sum_{l=1}^{g} u_{il}(t) > 0, \tag{4.4}$$

[5] For a rigorous proof of Property A.21, see the proof of Property 28 in Fujita (1976b). Property 28 includes Property A.21 as a special case.

[6] We do not need to take this assumption literally. For example, suppose that the original city (at time $t = 0$) forms a densely built-up area within a certain distance from the city center. Then, Assumption OPA-4 means that we are going to construct new buildings only in the suburbs of the existing buil-up area.

which implies

$$\sum_{l=1}^{g} u_{il}(t) = N_i(t) \quad \text{if} \sum_{l=1}^{g} u_{il}(t) > 0. \tag{4.5}$$

And, from Assumption OPA-4 and since buildings cannot be destructed, (4.4) also implies that

$$N_i(t) = \max\{N(\tau)|\ \tau \leqslant t\} \quad \text{if} \sum_{l=1}^{g} u_{il}(t) > 0. \tag{4.6}$$

And, from Property A.21,

$$\begin{aligned} x_{il}(\tau) &= x_{il}(t) \quad \text{for } l = 1, 2, \ldots, l_i(t)-1 \text{ and for all } \tau \geqslant t, \\ x_{il}(\tau) &\geqslant x_{il}(t) \quad \text{for } l \geqslant l_i(t) \text{ and for all } \tau \geqslant t. \end{aligned} \tag{4.7}$$

Therefore, on the optimal path, we can arrange the utilization of buildings of type i in district $l_i(t)$ so that those buildings of type i which are constructed at time t become vacant at time τ ($\tau > t$) if and only if $N_i(\tau) < N_i(t)$.

Hence, if we introduce the following notation,

$$N_i^*(t) = \max\{N_i(\tau)|\ \tau \leqslant t\}, \tag{4.8}$$

$$\begin{aligned} T_t^i &= \{\tau|\ N_i(\tau) \geqslant N_i^*(t),\ \tau \geqslant t\} \quad \text{when } N_i^*(t) > 0, \\ T_t^i &= \text{closure of } \{\tau|\ N_i(\tau) > 0,\ \tau > t\} \quad \text{when } N_i^*(t) = 0, \end{aligned} \tag{4.9}$$

then we can conclude that

Property A.22. Suppose we have (a) (or (b)) of Assumption OPA-3 and Assumption OPA-4. Then, the optimal construction process, $u_{il}(t)$, $i = 1, 2, \ldots, m$, $l = 1, 2, \ldots, g$, $0 \leqslant t < \infty$, must satisfy the following relation at each time t

$$\sum_{l=1}^{g} u_{il}(t) = \dot{N}_i^*(t), \quad i = 1, 2, \ldots, m.^{7} \tag{4.10}$$

And, there is no vacant building of type i in any district when $\sum_{l=1}^{g} u_{il}(t) > 0$. And, the optimal utilization plan $\delta_t^i(\tau)$ ($t \leqslant \tau < \infty$) for buildings of type i constructed at time t (if they exist) can be given by

$$\delta_t^i(\tau) = \begin{cases} 1 & \text{if } \tau \in T_t^i, \\ 0 & \text{if } \tau \notin T_t^i. \end{cases} \tag{4.11}$$

Function $N_i^*(t)$ and an example of *occupancy set* T_t^i are depicted in Figure 3. Relation (4.10) tells us how many buildings of each type should be constructed in the entire city at each time. And, relation (4.11) enables us to specify the optimal utilization plans of buildings (constructed at each time on the optimal path) without knowing districts of these buildings.

From definitions (4.8) and (4.9), we see that

$$\text{if } t' < t'', \text{ then } T_{t'}^i \supset T_{t''}^i, \text{ and } T_{t'}^i \neq T_{t''}^i \text{ whenever } \dot{N}_i^*(t') > 0. \tag{4.12}$$

Hence, from Properties A.14 and A.22, we get

[7] More precisely, $\sum_{l=1}^{g} u_{il}(t) = \dot{N}_i^*(t_+)$, where $\dot{N}_i^*(t_+) = \lim_{\varepsilon>0,\varepsilon\to 0} \dot{N}_i^*(t+\varepsilon)$.

Property A.23. Suppose we have (b) of Assumption OPA-3 and Assumption OPA-4. Then, slope $(A_i/k_i)\sigma_i(\delta_t^i(\tau))$ of bid land price line $p_l^i(t \mid \psi_{il}(q_i(\tau), \tau), \delta_t^i(\tau))$ is nonincreasing at any time, and it is decreasing (with respect to time) whenever $\dot{N}_i^*(t) > 0$. Here, $\delta_t^i(\tau)$ is the optimal utilization plan specified by (4.11).

Note that slope $(A_i/k_i)\sigma_i(\delta_t^i(\tau))$ does not necessarily always change continuously. That is, from (3.11) and (4.11), we have

Property A.24. Suppose we have (b) of Assumption OPA-3 and Assumption OPA-4. Then, slope $(A_i/k_i)\sigma_i(\delta_t^i(\tau))$ of bid rent curve for buildings type i changes continuously at time t if and only if set T_t^i (see (4.9)) changes continuously at that time.[8] Here, $\delta_t^i(\tau)$ is given by (4.11).

An example of time at which set T_t^i changes discontinuously can be found in Figure 3.

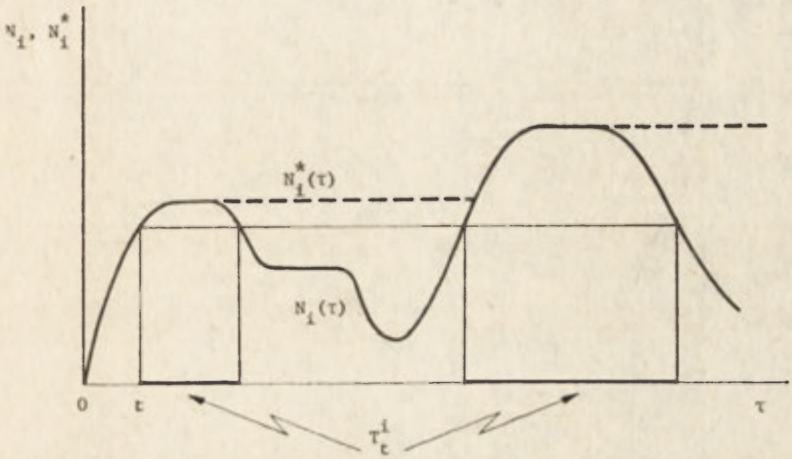


Fig. 3. Demand functions $N_i(t), N_i^*(t)$ and occupancy set T_t^i

Next, suppose that initial land prices $p_{l_1}(0), p_{l_2}(0)$ and $p_{l_3}(0)$ were located on a straight line (on $p_l - d_l$ coordinates space), where $l_1 < l_2 < l_3$ and $p_l(0) > 0$ for $l = l_1, l_2$, and l_3. Then, from (4.10) and Property A.23, when buildings of type i are being constructed in some district, the bid land price line of that building type could not touch point $p_{l_2}(0)$ from the below more than for an instant; which implies from Properties A.7 and A.8 that vacant land in district l_2 would never be filled up with buildings of any types. This is a contradiction of Property A.11.

Therefore, recalling Property A.15, we get

Property A.25. Suppose we have (b) of Assumption OPA-3 and Assumption OPA-4. Then, initial land price curve $p_l(0)$ is strictly convex on the interval between $l = 1$ and $l = \bar{l}$, where $\bar{l}$ is the nearest district to the city center among all the districts in which $p_l(0) = 0$.[9]

Combining Properties A.10, A.24 and A.25, we get

---

[8] That is, $\lim_{\varepsilon > 0, \varepsilon \to 0} (A_i/k_i)\sigma_i(\delta_{t+\varepsilon}^i(\tau)) = \lim_{\varepsilon > 0, \varepsilon \to 0} (A_i/k_i)\sigma_i(\delta_{t-\varepsilon}^i(\tau))$ if and only if $\lim_{\varepsilon > 0, \varepsilon \to 0} T_{t+\varepsilon}^i = \lim_{\varepsilon > 0, \varepsilon \to 0} T_{t-\varepsilon}^i$.

[9] For the definition of strict convexity, see footnote 4. Reader may notice that Property A. 25 is true under a weaker assumption that every district has some vacent land at the initial time.

Property A.26. Suppose we have (b) of Assumption OPA-3 and Assumption OPA-4. Then, if occupancy set T_t^i changes continuously during a time interval, say (t, t), and if $\dot{N}_i^*(t) > 0$ for that time interval, construction site $l_i(t)$ of building type i moves continuously outwards during that time interval.[10]

Next, suppose buildings of types i' and i'' are being constructed, respectively, in districts $l_{i'}(t)$ and $l_{i''}(t)$ at time t. Then, from Properties A.10 and A.25, the relation between bid land price curves $p_l^i(t \,|\, \psi_{il}(q_i(\tau), \tau), \delta_t^i(\tau))$, $i = i'$ and i'', and initial land price curve $p_l(0)$ can be depicted as in Figure 4. Thus, as a relation between slopes of bid land price curves and locations of construction districts among different building types, we have

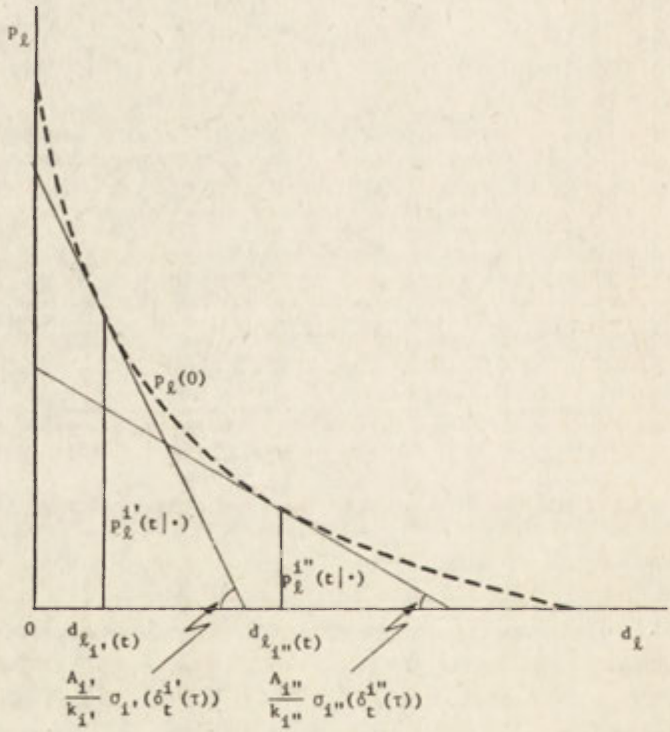


Fig. 4. Relation between locations of construction sites and slopes of bid land price curves

Property A.27. Suppose we have (b) of Assumption OPA-3 and Assumption OPA-4. Then, at each time t, building construction sites $l_i(t)$ ($i = 1, 2, \ldots, m$, if they exist) locate outward from the city center in order of the absolute values of the slopes of bid rent curves $(A_i/k_i)\sigma_i(\delta_t^i(\tau))$ ($i = 1, 2, \ldots, n$, if they exist). That is, if $\sum_{l=1}^{g} u_{il}(t) > 0$ for $i = i'$ and i'', then

[10] That is, if $l_i(t_1) < l' < l_i(t_2)$ and $t_1, t_2 \in (t, \bar{t})$, then there exists a time interval (t', t'') such that $t_1 < t' < t'' < t_2$ and $l_i(t) = l'$ for $t \in (t', t'')$.

$$l_{i'}(t) \leqslant l_{i''}(t) \quad \text{whenever } \frac{A_{i'}}{k_{i'}} \sigma_{i'}(\delta_t^{i'}(\tau)) > \frac{A_{i''}}{k_{i''}} \sigma_{i''}(\delta_t^{i''}(\tau)).^{11} \tag{4.13}$$

Here $\delta_t^i(\tau)$ is specified by (4.11).

Next, let us call district $\bar{l}$ defined in Property A.25 the *fringe district* of the city. Then, since Property A.25 implies that

$$p_l(0) > 0 \quad \text{for } l = 1, 2, \ldots, \bar{l}-1,$$

we immediately see from Property A.11 that

Property A.28. Suppose we have (b) of Assumption OPA-3 and Assumption OPA-4. Then, every district inside of the urban fringe (i.e., $l = 1, 2, \ldots, \bar{l}-1$) should eventually be filled up by buildings of some types.

And, Property A.22 implies that

Property A.29. Under (a) (or (b)) of Assumption OPA-3 and Assumption OPA-4, the number of buildings of type i to be constructed within the plan period is just equal to $\max\{N_i(t) |\ t \in [0, \infty)\}$, where $i = 1, 2, \ldots, m$.

Thus, fringe district $\bar{l}$ is a district such that it satisfies the following relation.

$$\sum_{l=1}^{\bar{l}-1} s_l < \sum_{i=1}^{m} k_i \max\{N_i(t) |\ t \in [0, \infty)\} \leqslant \sum_{l=1}^{\bar{l}} s_l. \tag{4.14}$$

Finally, as a special case, suppose we have

$$\dot{N}_i(t) \geqslant 0 \quad \text{for all } t \geqslant 0.$$

Then, from definitions (4.8) and (4.9),

$$N_i^*(t) = N_i(t), \tag{4.15}$$

$$\begin{aligned} T_t^i &= \{\tau |\ \tau \geqslant t\} \quad \text{when } \dot{N}_i(t) > 0, \\ T_t^i &= \text{closure of } \{\tau |\ \dot{N}_i(\tau) > 0\} \quad \text{when } \dot{N}_i(t) = 0. \end{aligned} \tag{4.16}$$

Therefore, if we define $\underline{t}_i$ $(i = 1, 2, \ldots, m)$ by

$$\underline{t}_i = \inf\{t |\ \dot{N}_i(t) > 0\}, \tag{4.17}$$

we obtain the following corollary to Property A.22.

Property A.30. Suppose we have (a) (or (b)) of Assumption OPA-3 and Assumption OPA-4. And assume that the demand function $N_i(t)$ for building of type i is nondecreasing with respect to time, that is,

$$\dot{N}_i(t) \geqslant 0 \quad \text{for all } t \geqslant 0.$$

Then, the optimal construction process $u_{il}(t)$, $l = 1, 2, \ldots, n$, $0 \leqslant t < \infty$, of buildings of type i should satisfy the following relation

$$\sum_{l=1}^{g} u_{il}(t) = \dot{N}_i(t) \quad \text{for all } t \geqslant 0. \tag{4.18}$$

And, there is no vacant building of type i in any district at any time; and thus the optimal utilization plan for buildings of type i is given by

[11] Note that we may have $l_{i'}(t) = l_{i''}(t)$ since d_l changes discretely in our model.

$$\text{if } t \geqq t_i,\ \delta_t^i(\tau) = 1 \text{ for all } \tau \geqq t,$$
$$\text{if } t < t_i,\ \delta_t^i(\tau) = \begin{cases} 0 & \text{for } t \leqq \tau < t_i, \\ 1 & \text{for } \tau \geqq t_i, \end{cases} \tag{4.19}$$

where t_i is defined by (4.17). Therefore, under (b) of Assumption OPA-3,

$$\frac{\partial p^i\big((t \mid \psi_{il}(q_i(\tau), \tau), \delta_t^i(\tau)\big)}{\partial d_l} = \begin{cases} -\dfrac{A_i}{k_i}\sigma_i(t) \text{ if } t \geqq t_i, \\ -\dfrac{A_i}{k_i}\sigma_i(t_i) \text{ if } t < t_i, \end{cases} \tag{4.20}$$

where

$$\sigma_i(t) \equiv \int_t^\infty e^{-\gamma\tau}\alpha_i(\tau)d\tau. \tag{4.21}$$

5. SPATIAL PATTERNS IN THE CASE OF TWO BUILDING TYPES ($m = 2$)

In this section, we assume that the city is to be composed of only two types of buildings ($i = 1, 2$), and study in detail optimal spatial patterns of urban growth and contraction under various specifications on building demand functions $N_i(t)$ ($i = 1, 2$).

Throughout this section, we employ (b) of Assumption OPA-3 and Assumption OPA-4.

5.1. CONSTRUCTION PROCESSES

We first obtain the optimal construction processes in this section. Since the numbering of building types can be arbitrary, we can suppose without loss of generality that

$$\frac{A_1}{k_1}\sigma_1\big(\delta_0^1(\tau)\big) \geqq \frac{A_2}{k_2}\sigma_2\big(\delta_0^2(\tau)\big), \tag{5.1}$$

where $\delta_0^i(\tau)$ is the optimal utilization plan for buildings of type i specified by (4.11) with $t = 0$ ($i = 1, 2$).

From (4.10), at each time t, buildings of type i must be constructed as much as $\dot{N}_i^*(t)$ in the city. When $\dot{N}_i^*(t) > 0$, from Property A.10, bid land price line $p_l^i\big(t \mid \psi_{il}(q_i(\tau), \tau), \delta_t^i(\tau)\big)$ must touch the initial land price curve $p_l(0)$ from below in construction district $l_i(t)$. On the other hand, when $\dot{N}_i^*(t) = 0$, bid land price curve $p_l^i\big(t \mid \psi_{il}(q_i(\tau), \tau), \delta_t^i(t)\big)$ may not touch the initial land price curve in any district, and hence $l_i(t)$ would not exist. But, for convenience, when $\dot{N}_i^*(t) = 0$, we denote by $l_i(t)$ the district in which the initial land price curve and the bid land price line (for building type i) would first touch to each other if the latter curve were moved upward in parallel;[12] and we consider that zero buildings of type i are being constructed in district $l_i(t)$. Consequently, we can consider that construction district $l_i(t)$ exists for each building type i at any time t.

[12] If there are two such districts, we take one of them arbitrarily.

It is easy to see that relation (4.13) holds to be true also when we generalize the definition of construction district $l_i(t)$ as in the above. Hence, using (4.13) for $t = 0$, we always have from (5.1) that

$$l_1(0) \leqslant l_2(0).$$

From Property A.28, there should exist a time after which no vacant land is left in district 1. Hence, it must be true that

$$l_1(0) = 1 \leqslant l_2(0). \tag{5.2}$$

Next, from Property A.21, district $l_i(t)$ moves toward the suburbs with time.[13] We denote by t_{il}^* the time at which (generalized) construction district $l_i(t)$ moves from district l to $l+1$. Here, t_{1l}^* is defined for $l = 1, 2, \ldots, \bar{l}-1$, and t_{2l}^* for $l = l_2(0)$, $l_2(0)+1, \ldots, \bar{l}-1$.[14] From (4.14), fringe district $\bar{l}$ is determined by the relation

$$\sum_{l=1}^{\bar{l}-1} s_l < \sum_{i=1}^{2} k_i \max\{N_i(t)|\ t \in [0, \infty)\} \leqslant \sum_{l=1}^{\bar{l}} s_l. \tag{5.3}$$

When construction district $l_i(t)$ $(i = 1, 2)$ moves from l to $l+1$, bid land price line of building type i (or its upward-shifted line) touches the initial land price curve $p_l(0)$ in both districts, l and $l+1$; which implies from (3.10) that

$$\frac{A_1}{k_1}\sigma_1(\delta_{t_{1l}^*}(\tau)) = \frac{A_2}{k_2}\sigma_2(\delta_{t_{2l}^*}(\tau)) \quad \text{for } l = l_2(0), \ldots, \bar{l}-1. \tag{5.4}$$

And, from Property A.28, each district within the urban fringe should eventually be filled up by buildings of the two types. Hence, it must be true from (4.10) (and from Property A.21) that

$$k_1 N_1^*(t_{1l}^*) = \sum_{j=1}^{l} s_j \quad \text{for } l = 1, 2, \ldots, l_2(0)-1, \tag{5.5}$$

$$\sum_{i=1}^{2} k_i N_i^*(t_{il}^*) = \sum_{j=1}^{l} s_j \quad \text{for } l = l_2(0),\ l_2(0)+1, \ldots, \bar{l}-1. \tag{5.6}$$

Therefore, district $l_2(0)$ is obtained as the nearest district to the city center among all the districts for which the values of t_{1l}^* and t_{2l}^* determined by the simultaneous equations, (5.4) and (5.6), are both positive. After determining $l_2(0)$, moving times t_{1l}^*, $l = 1, 2, \ldots, l_2(0)-1$, are obtained by solving (5.5) for each t_{1l}^*. And, moving times t_{il}^*, $i = 1, 2, l = l_2(0)$, $l_2(0)+1, \ldots, \bar{l}-1$, are determined by simultaneously solving (5.4) and (5.6) for each l.

By using these t_{il}^*, the optimal construction process $u_{il}(t)$ $(i = 1, 2, l = 1, 2, \ldots, g,\ 0 \leqslant t < \infty)$ for problem OPA can be represented by

$$u_{il}(t) = \begin{cases} \dot{N}_i^*(t) & \text{for } t_{il-1}^* \leqslant t < t_{il}^*, \\ 0 & \text{otherwise.} \end{cases} \tag{5.7}$$

[13] It is clear from Property A. 23 and Property A. 25 that Property A. 21 holds to be true also under the generalized definition of $l_i(t)$.

[14] When t_{il}^* jumps from $l = l'$ to $l = l''$ at time t, we consider that $t_{il'}^* = t_{il'+1}^* = \ldots = t_{il''-1}^* = t$.

And, the number of buildings of each type i constructed in each district l during the plan period is given by

$$N_i^*(t_{il}^*) - N_i^*(t_{il-1}^*). \tag{5.8}$$

5.2. SPATIAL PATTERNS OF URBAN GROWTH: CASE 1

In this subsection, we study spatial patterns of urban growth under the following specifications on building demand functions and slopes of bid rent lines (recall (b) of Assumption OPA-3 and (3.2)).

$$\left.\begin{array}{ll} \dot{N}_i(t) > 0 & \text{for } 0 \leqslant t < \bar{t} \\ \dot{N}_i(t) = 0 & \text{for } t \geqslant \bar{t} \end{array}\right\} \quad i = 1, 2, \tag{5.9}$$

$$\frac{A_1 \alpha_1(t)}{k_1} > \frac{A_2 \alpha_2(t)}{k_2} \quad \text{for all } t \geqslant 0. \tag{5.10}$$

Under assumption (5.9), from Property A.30, there is no vacancy in buildings of any type at any time. And, under assumption (5.10), we have from (4.20) and (4.21) that

$$\frac{A_1}{k_1} \sigma_1(t) > \frac{A_2}{k_2} \sigma_2(t) \quad \text{for all } t \geqslant 0. \tag{5.11}$$

That is, the bid land price line of building type 1 has always steeper slope than that of building type 2. Hence, from (4.13) we have

$$l_1(t) \leqslant l_2(t) \quad \text{for all } t \geqslant 0.^{15} \tag{5.12}$$

That is, construction district of building type 1 locates always closer to the city center than that of building type 2.

For further discussions it is convenient to distinguish the next two alternative cases:

$$\text{case 1-1.} \quad \frac{A_1}{k_1} \sigma_1(\bar{t}) < \frac{A_2}{k_2} \sigma_2(0). \tag{5.13}$$

$$\text{case 1-2.} \quad \frac{A_1}{k_1} \sigma_1(\bar{t}) \geqslant \frac{A_2}{k_2} \sigma_2(0). \tag{5.14}$$

Given values of parameters A_i, k_i, $\alpha_i(\tau)$ ($i = 1, 2$, $\tau \geqslant 0$), case 1-1 occurs when growth period $\bar{t}$ is relatively long, and case 1-2 when $\bar{t}$ is relatively short (recall definition (4.21) and assumption (5.10)).

Let us first examine case 1-1. By using (5.12), the process of the optimal urban growth in case 1-1 can be summarized as follows (refer to Figure 5(ii)):

(1) Construction of buildings of type 2 begins in some district $l_2(0)$ far from the city center. Then, while some vacant land remains there, it moves to the next district $l_2(0)+1$. Likewise, it gradually moves outward leaving some vacant land in each district.

(2) Construction of buildings of type 1 begins in the nearest district to the city center (i.e., in $l = 1$). Then, occupying the whole area in each district, it moves toward

[15] Note that if the size of each district is sufficiently small, we always have strict inequality in (5.12) for $t \in [0, \bar{t})$.

district $l_2(0)-1$. After that district, it gradually moves outward occuping all the vacant land remaining in each district after construction of buildings of type 2. (3) Construction of buildings of type 2 ends in the fringe district $\bar{l}$ at time $\bar{t}$, and construction of buildings of type 1 ends at the same time in some district $l_1(\bar{t})$ which is inside of the urban fringe.

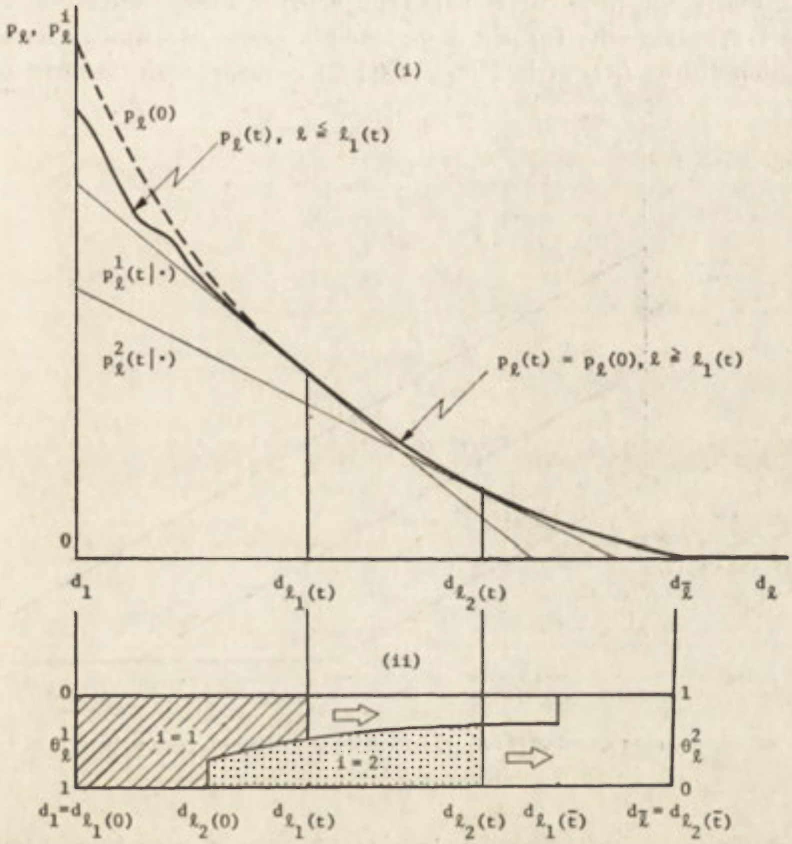


Fig. 5. (i) Relation among curves $p_l(0)$, $p_l(t)$ and $p_l^i(t|\cdot)$ and (ii) spatial pattern of urban growth in case 1-1 $(0 < t < \bar{t})$

Figure 5 (ii) depicts this process of urban spatial growth. The left (right) vertical axis represents land use ratio $\theta_l^1(\theta_l^2)$ in each distruct l, where

$$\theta_l^i = \frac{k_i x_{il}}{s_l}, \quad i = 1, 2.$$

And, the horizontal axis shows the distance of each district from the city center. One of the outstanding characteristics of this growth process is that the construction of buildings of type 2 moves toward the suburbs while leaving a large area vacant in each district. Hence, this growth process of urban space has the property of urban sprawl, and results in a mixture of different buildings in many districts.

Figure 5 (i) depicts the relationship between initial land price curve $p_l(0)$, land price curve $p_l(t)$ at time t and bid land price curves $p_l^i\big(t \mid \psi_{il}(q_i(\tau), \tau)\big), \delta_t^i(\tau)\big)$, $i = 1,2$, where $0 < t < \bar{t}$, and $\delta_t^i(\tau) = 1$ for all $\tau \geqq t$. Initial land price curve $p_l(0)$ is strictly convex between districts 1 to $\bar{l}$; its slope changes smoothly from district 1 to district $\bar{l}$; curve $p_l(0)$ is kinked at district $\bar{l}$. Bid land price curves $p_l^i\big(t \mid \psi_{il}(q_i(\tau), \tau),\ \delta_t^i(\tau)\big)$, $i = 1, 2$, touch land price curves $p_l(t)$ $\big($and $p_l(0)\big)$ in their construction districts, $l_i(t)$, $i = 1, 2$, respectively. The part of the land price curve $p_l(t)$ which is further than construction district $l_1(t)$ of buildings of type 1 coincides with that part of curve

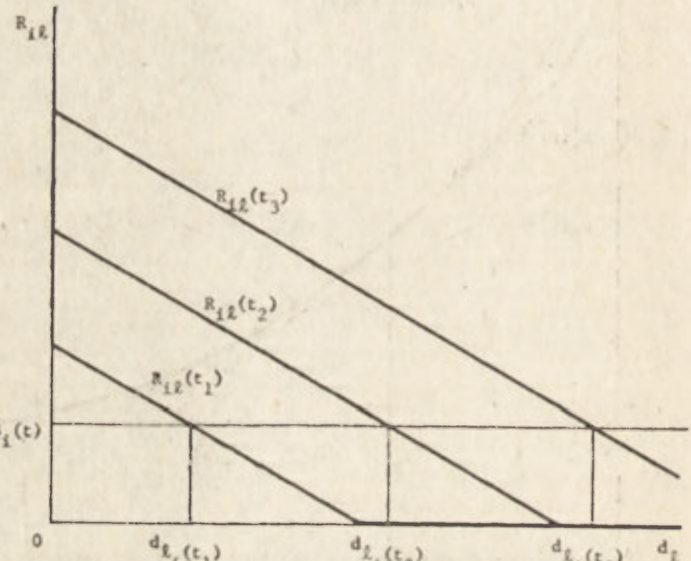


Fig. 6. Variation of building rent curve $R_{1l}(t)$ during the construction period $(0 < t_1 < t_2 < t_3 < < \bar{t})$

$p_l(0)$. But, the part of the land price curve $p_l(t)$ between districts 1 and $l_1(t)$, where no vacant land remains, is essentially indeterminate;[16] any curve lying between curves $p_l(0)$ and $p_l^1\big(t \mid \psi_{1l}(q_1(\tau), \tau),\ \delta_t^1(\tau)\big)$ can serve as that part of land price curve at time t $\big($see Figure 5 (i)$\big)$. Similarly, after time $t^*_{l_1(\bar{t})}$ $\big($i.e., after construction district $l_2(t)$ of buildings of type 2 goes beyond district $l_1(\bar{t})\big)$, the part of land price curve $p_l(t)$ between districts $l_1(\bar{t})$ and $l_2(t)$ is essentially indeterminate; any curve lying between curves $p_l(0)$ and $p_l^2\big(t \mid \psi_{2l}(q_2(\tau), \tau),\ \delta_t^2(\tau)\big)$ can serve as that part of land price curve $p_l(t)$ at that time.

Next, from definition (3.1) and from Properties A.1 and A.18 (i) (and recalling (b)

[16] More precisely, this means: though there exists a unique land price $p_l(t)$ is each of districts $l = 1, 2, \ldots, l_1(t) - 1$ at time t, its value depends on the past values of $p_l(\tau)$ $(0 \leqq \tau < t)$ which are arbitrary (except for the restrictions, $\dot{p}_l(\tau) \leqq 0$, and $p_l(t) \geqq p_l^i(t \mid \cdot)$, $i = 1, 2$).

of Assumption OPA-3), the equation of rent curve $r_{il}(t)$ (in case 1-1) during the construction period, $[0, \bar{t}]$, is given by

$$r_{il}(t) = -\dot{b}_i(t) + e^{-\gamma t} A_i \alpha_i(t)(d_{l_i(t)} - d_l) \quad \text{for } i = 1, 2, 0 \leqq t < \bar{t}.^{17} \tag{5.15}$$

Or, in terms of current prices, we have

$$R_{il}(t) = (\gamma B_i(t) - \dot{B}_i(t)) + A_i \alpha_i(t)(d_{l_i(t)} - d_l) \quad \text{for } i = 1, 2, 0 \leqq t < \bar{t}. \tag{5.16}$$

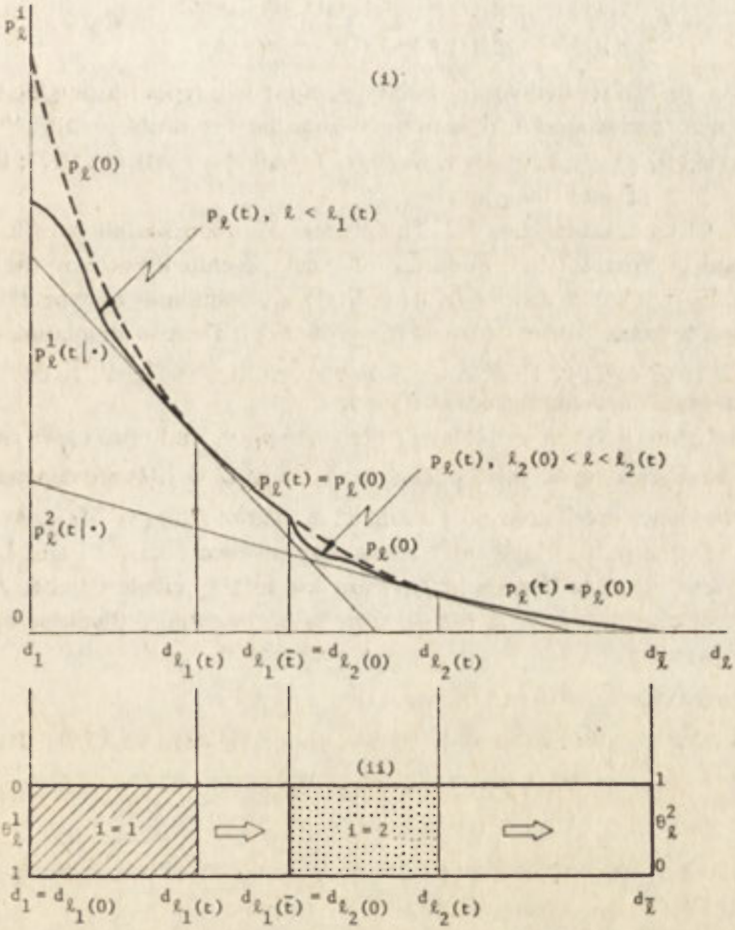


Fig. 7. (i) Relation among curves $p_l(0), p_l(t)$ and $p_l^i(t)$ and (ii) spatial patterm of urban growth in case 1-2 $(0 < t < \bar{t})$

Figure 6 depicts rent curves $R_{il}(t)$ for building type i ($i = 1$ or 2) at times, $t = t_1$, t_2 and t_3, where $0 \leqq t_1 < t_2 < t_3 < \bar{t}$. It is assumed in this figure, for simplicity, that $A_i \alpha_i(t)$ and $\gamma B_i(t) - \dot{B}_i(t)$ are constant, respectively.

[17] More precisely, we have

$$r_{il}(t) = \max\{-\dot{b}_i(t) + e^{-\gamma t} A_i \alpha_i(t)(d_{l_i(t)} - d_l), 0\} \quad \text{for } i = 1, 2, l = 1, 2, \ldots, \bar{l}, 0 \leqq t < \bar{t}.$$

After the end $\bar{t}$ of construction period, time-variation of rent curves for buildings of each type is essentially arbitrary. That is, suppose that $\beta_i(t)$, $i = 1, 2$, are any functions of time such that

$$\begin{aligned} \beta_i(\bar{t}) = b_i(\bar{t}), \quad \beta_i(t) \leqslant b_i(t) \quad \text{and} \quad \dot{\beta}_i(t) \leqslant 0 \\ \text{for all } t \geqslant \bar{t}, \quad \text{and} \quad \lim_{t\to\infty} \beta_i(t) = 0. \end{aligned} \tag{5.17}$$

Then, the following equations

$$\left.\begin{aligned} r_{1l}(t) &= -\dot{\beta}_1(t) + e^{-\gamma t} A_1 \alpha_1(t)(d_{l_1(t)} - d_l), \\ r_{2l}(t) &= -\dot{\beta}_2(t) + e^{-\gamma t} A_2 \alpha_2(t)(d_{\bar{l}} - d_l), \end{aligned}\right\} \quad t \geqslant \bar{t}, \tag{5.18}$$

represent a pair of rent curves for buildings of the two types (during the time period, $t \geqslant \bar{t}$) which correspond to the optimal solution for problem OPA.[18] Of course, functions $\beta_i(t)$, $i = 1, 2$, which equal $b_i(t)$ for all $t \geqslant \bar{t}$ satisfy (5.17); but they are only one pair of such functions.

Next, let us examine case 1-2. The process of urban spatial growth of this case is depicted in Figure 7 (ii). Buildings of type 1 eventually occupy the whole area between district 1 and district $l_2(0)\left(= l_1(\bar{t})\right)$, and buildings of type 2 occupy the whole area between district $l_2(0)$ and fringe district $\bar{l}$. There is on mixture of buildings of different types in any district (except possibly in district $l_2(0)$). And, so called, "Thünen rings" are completed at time $\bar{t}$.

The relations between initial land price curve $p_l(0)$, land price curve $p_l(t)$ at time t and bid land price curves $p_l^i\left(t \mid \psi_{il}(q_l(\tau), \tau), \delta_t^i(\tau)\right)$ $(i = 1, 2)$ are depicted in Figure 7 (i). Initial land price curve $p_l(0)$ is kinked in district $l_2(0)$ $\left(= l_1(\bar{t})\right)$. At each time t $(0 < t < \bar{t})$, the part of land price curve $p_l(t)$ between districts 1 and $l_1(t)$ and the part between districts $l_2(0)$ and $l_2(t)$ are essentially indeterminate, respectively.

For rent curves $(i = 1, 2, t \geqslant 0)$, there is no essential differences between case 1-1 and case 1-2.

5.3. SPATIAL PATTERNS OF URBAN GROWTH: CASE 2

As a generalization of case 1, let us assume (instead of (5.9)) that

$$\left.\begin{aligned} \dot{N}_i(t) > 0 \quad &\text{for } t \in [\underline{t}_i, \bar{t}_i), \\ \dot{N}_i(t) = 0 \quad &\text{for } t \notin [\underline{t}_i, \bar{t}_i), \end{aligned}\right\} \quad i = 1, 2, \tag{5.19}$$

where $\underline{t}_i$ and $\bar{t}_i$ are, respectively, certain times specified exogenously for each i $(i = 1, 2)$. We retain assumption (5.10) as before.

Under (5.19), from Property A.30, there is no vacancy in buildings of any type at any time (as before). And, the slopes $(A_i/k_i)\sigma_i(t)$ of bid land price lines of buildings of two types $(i = 1, 2)$ have relation (5.11). Hence, relation (5.12) is also true.

Figure 8 depicts all the possible urban growth patterns under assumptions (5.10) and (5.19); conditions on parameters which are necessary and sufficient for realiza-

[18] Here, $l_1(t)$ is the fictitious construction district for buildings of type 1 at time t $(t \geqq \bar{t})$, which is obtained by solving simultaneously (5.5) and (5.6), assuming that zero units of buildings of type 1 are being constructed in the city (see the beginning of section 5.1 for the definition of $l_i(t)$ when $N_i^*(t)$ $(= \dot{N}_i(t))$ is zero). Of course, $l_1(t) = \bar{l}$ for $t > t_{1l-1}^*$.

tion of each pattern are summarized on Table 1 (recall that function $\sigma_i(t)$ is defined by (4.21) for $i = 1, 2$). Each diagram in Figure 8 represents the pattern of urban land use at the end of the growth of the city (i.e., at the time, max $\{\bar{t}_1, \bar{t}_2\}$). As in Figure 5 (ii), the vertical axis of each diagram shows land use ratio θ_l^i ($i = 1$ er 2) of buildings of each type; and the horizontal axis represents distance d_l of each district from the city center. In each pattern, construction of buildings of type i ($i = 1, 2$)

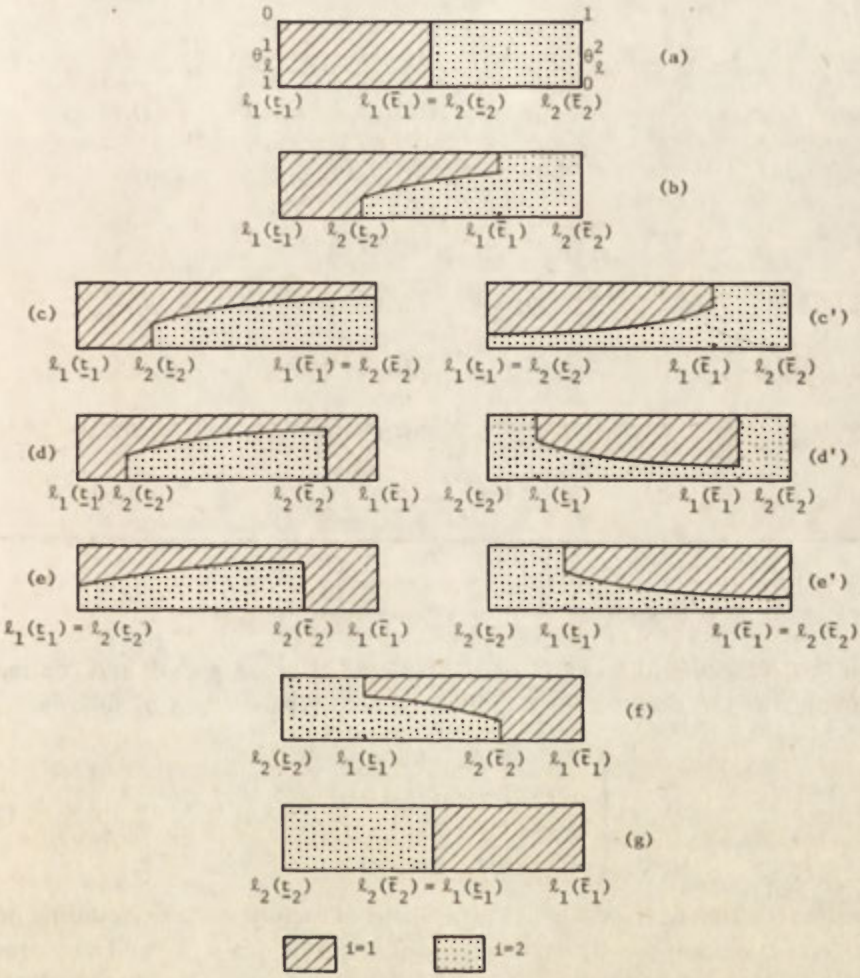


Fig. 8. Spatial patterns of urban land use at time max $\{\bar{t}_1, \bar{t}_2\}$ in case 2

begins in district $l_i(\underline{t}_i)$ at time $\underline{t}_i$, moves gradually outward, and ends in district $l_i(\bar{t}_i)$ at time $\bar{t}_i$. Since occupancy set T_t^i continuously changes during construction period $[\underline{t}_i, \bar{t}_i]$ for each i, we see from Property A.26 that the area occupied by buildings of each type i is spatially connected in any pattern. Except for Patterns (a) and (g), mixture of buildings of different types happens in some districts.

Note that Patterns (a) and (b) include, respectively, cases 1-1 and 1-2 as their special cases.

TABLE 1. Relations between spatial patterns and conditions on parameters in case 2

Pattern	Conditions on parameters
(a)	$\frac{A_1}{k_1}\sigma_1(\bar{t}_1) \geqq \frac{A_2}{k_2}\sigma_2(\underline{t}_2)$
(b)	$\frac{A_1}{k_1}\sigma_1(\underline{t}_1) > \frac{A_2}{k_2}\sigma_2(\underline{t}_2) > \frac{A_1}{k_1}\sigma_1(\bar{t}_1) > \frac{A_2}{k_2}\sigma_2(\bar{t}_2)$
(c)	$\frac{A_1}{k_1}\sigma_1(\underline{t}_1) > \frac{A_2}{k_2}\sigma_2(\underline{t}_2) > \frac{A_1}{k_1}\sigma_1(\bar{t}_1) = \frac{A_2}{k_2}\sigma_2(\bar{t}_2)$
(c′)	$\frac{A_1}{k_1}\sigma_1(\underline{t}_1) = \frac{A_2}{k_2}\sigma_1(\underline{t}_1) > \frac{A_1}{k_1}\sigma_1(\bar{t}_1) > \frac{A_2}{k_2}\sigma_2(\bar{t}_2)$
(d)	$\frac{A_1}{k_1}\sigma_1(\underline{t}_1) > \frac{A_2}{k_2}\sigma_2(\underline{t}_2) > \frac{A_2}{k_2}\sigma_2(\bar{t}_2) > \frac{A_1}{k_1}\sigma_1(\bar{t}_1)$
(d′)	$\frac{A_1}{k_2}\sigma_2(\underline{t}_2) > \frac{A_1}{k_1}\sigma_1(\underline{t}_1) > \frac{A_1}{k_1}\sigma_1(\bar{t}_1) > \frac{A_2}{k_2}\sigma_2(\bar{t}_2)$
(e)	$\frac{A_1}{k_1}\sigma_1(\underline{t}_1) = \frac{A_2}{k_2}\sigma_2(\underline{t}_2) > \frac{A_2}{k_2}\sigma_2(\bar{t}_2) > \frac{A_1}{k_1}\sigma_1(\bar{t}_1)$
(e′)	$\frac{A_2}{k_2}\sigma_2(\underline{t}_2) > \frac{A_1}{k_1}\sigma_1(\underline{t}_1) > \frac{A_1}{k_1}\sigma_1(\bar{t}_1) = \frac{A_2}{k_2}\sigma_2(\bar{t}_2)$
(f)	$\frac{A_2}{k_2}\sigma_2(\underline{t}_2) > \frac{A_1}{k_1}\sigma_1(\underline{t}_1) > \frac{A_2}{k_2}\sigma_2(\bar{t}_2) > \frac{A_1}{k_1}\sigma_1(\bar{t}_1)$
(g)	$\frac{A_2}{k_2}\sigma_2(\bar{t}_2) > \frac{A_1}{k_1}\sigma_1(\underline{t}_1)$

5.4. SPATIAL PATTERNS OF URBAN GROWTH AND CONTRACTION: CASE 3

In this subsection, we obtain spatial patterns of urban growth and contraction assuming that the demand for buildings of each type changes as follows:

$$\dot{N}_i(t) \begin{cases} > 0 & \text{for } t \in [0, \bar{t}), \\ = 0 & \text{for } t \in [\bar{t}, \bar{t}+\omega), \\ < 0 & \text{for } t \in [\bar{t}+\omega, \hat{t}), \\ = 0 & \text{for } t \in [\hat{t}, \infty), \end{cases} \quad i = 1, 2, \tag{5.20}$$

where $\omega > 0$. That is, the demand for buildings of each type increases during period $[0, \bar{t})$, stays constant for $[\bar{t}, \bar{t}+\omega)$, decreases in period $[\bar{t}+\omega, \hat{t})$, and stay constant in the rest of the plan period. Assumption (5.10) is retained also in this subsection.

Let us first examine the next special case:

case 3-1. $N_1(t) = \alpha N_2(t)$ for $t \in [0, \hat{t}]$, and

$$N_1(\hat{t}) = N_2(\hat{t}) = 0, \quad \text{where } \alpha > 0. \tag{5.21}$$

That is, in addition to (5.10) and (5.20) we assume that demands for buildings of two types are proportional to each other at each time, and that the city becomes completely vacant at time $\hat{t}$. An example of demand functions under (5.20) and (5.21) is depicted in Figure 9.

Under assumption (5.21), we have from definition (4.9) that

$$T_t^1 = T_t^2 (\equiv T_t) \quad \text{for all } t \geqslant 0.^{19} \tag{5.22}$$

Thus, from (4.11), optimal utilization plans $\delta_t^i(\tau)$ ($t \leqslant \tau < \infty$, $i = 1, 2$) are given by

$$\delta_t^1(\tau) = \delta_t^2(\tau)(\equiv \delta_t(\tau)) = \begin{cases} 1 & \text{if } \tau \in T_t, \\ 0 & \text{if } \tau \notin T_t. \end{cases} \tag{5.23}$$

Hence, under assumption (5.10), from (3.10) and (3.11), slopes $(A_i/k_i)\sigma_i(\delta_t(\tau))$ of bid land price lines of the two building types have the following relation

$$\begin{aligned} \frac{A_1}{k_1}\sigma_1(\delta_t(\tau)) &> \frac{A_2}{k_2}\sigma_2(\delta_t(\tau)) \quad \text{for } 0 \leqslant t < \bar{t}+\omega, \\ \frac{A_1}{k_1}\sigma_1(\delta_t(\tau)) &= \frac{A_2}{k_2}\sigma_2(\delta_t(\tau)) = 0 \quad \text{for } t \geqslant \bar{t}+\omega. \end{aligned} \tag{5.24}$$

Thus, from (4.13), we have

$$l_1(t) \leqslant l_2(t) \quad \text{for } t \in [0, \bar{t}+\omega).^{20} \tag{5.25}$$

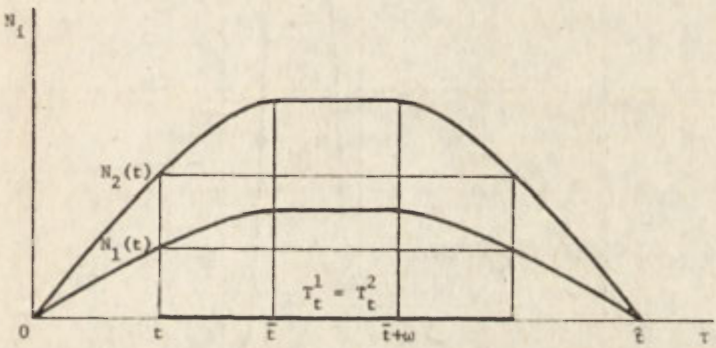


Fig. 9. Building demand functions $N_i(\tau)$, $i = 1, 2$, assumed in case 3-1 ($N_1(\tau) = \alpha N_2(\tau)$)

We have, as in section 5.2, two different spatial patterns depending on: case 3-1 (α):

$$\frac{A_1}{k_1}\sigma_1(\delta_{\bar{t}}(\tau)) < \frac{A_2}{k_2}\sigma_2(\delta_0(\tau)), \tag{5.26}$$

case 3-1 (β):

$$\frac{A_1}{k_1}\sigma_1(\delta_{\bar{t}}(\tau)) \geqslant \frac{A_2}{k_2}\sigma_2(\delta_0(\tau)), \tag{5.27}$$

where

$$\sigma_1(\delta_{\bar{t}}(\tau)) = \int_{\bar{t}}^{\bar{t}+\omega} e^{-\gamma\tau}\alpha_1(\tau)d\tau, \qquad \sigma_2(\delta_0(\tau)) = \int_0^{\hat{t}} e^{-\gamma\tau}\alpha_2(\tau)d\tau. \tag{5.28}$$

---

[19] Recall from definition (4.9) that, for $t = 0$, we have

$$T_0^i = [0, \hat{t}].$$

[20] Again, note that if the size of each district is sufficiently small, we have strict inequality in (5.25) for $t \in [0, \bar{t}+\omega)$.

It is clear from (5.10) and (5.28) that case 3-1 (β) can happen only when $\bar{t}$ is very small and ω is very large; otherwise, we always have case 3-1 (α).

Let us first examine case 3-1 (α). The spatial pattern of growth process (during $0 \leqslant t < \bar{t}$) under condition (5.26) is essentially the same as that of case 1-1. And, at time $\bar{t}$, the spatial pattern of land use takes a form of type (b) in Figure 8; as the length, ω, of the stationary state period approaches zero, the land use pattern at time $\bar{t}$ comes to a form of type (c) in Figure 8. Next, define functions $\hat{p}_l^i (i = 1, 2)$ by

$$\hat{p}_l^i(t) = \frac{\int_t^\infty r_{il}(\tau)d\tau - b_i(t)}{k_i}, \qquad l = 1, 2, \ldots, g. \tag{5.29}$$

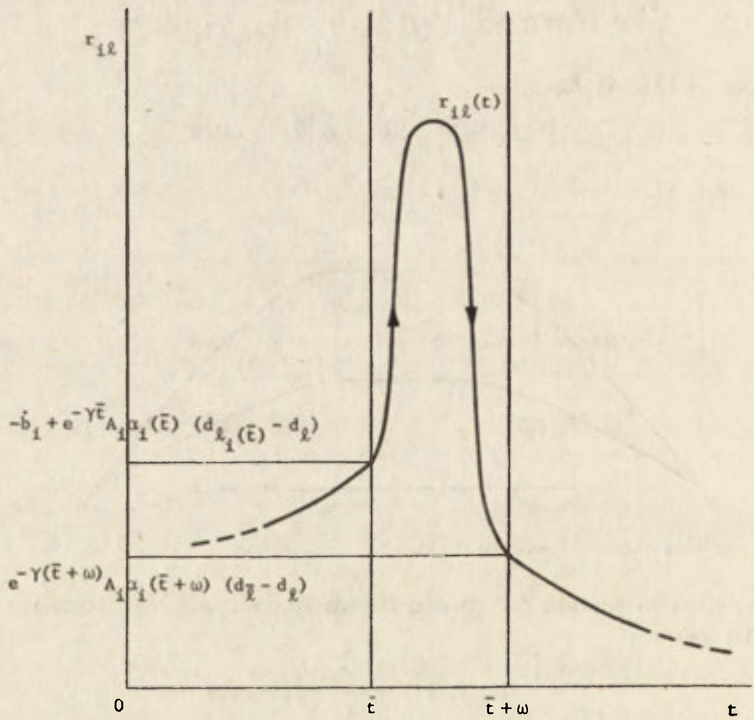


Fig. 10. Variation of building rent $r_{il}(t)$ in a district l $(l < \bar{l})$ within a neighborhood of time period $[\bar{t}, \bar{t}+\omega]$ in case 3

Curve $\hat{p}_l^i(t)$ represents the upper envelope of all the possible bid land prices for building type i at time t. It is easy to see that, under assumption (5.20),

$$\begin{aligned} \hat{p}_l^i(t) &> p_l^i\big(t \mid \psi_{il}(q_i(\tau), \tau),\ \delta_t^i(\tau)\big) \quad \text{for } l < l_i(t), \\ \hat{p}_l^i(t) &= p_l^i\big((t \mid \psi_{il}(q_i(\tau),\tau),\ \delta_t^i(\tau)\big) \quad \text{for } l = l_i(t). \end{aligned} \tag{5.30}$$

where $\delta_t^i(\tau)$ is given by (5.23).[21] Hence, any curve which is lying between two curves $p_l(0)$ and $\hat{p}_l^1(t)$ can serve as a land price curve on the interval $l = 1$ to $l_1(t)$. The

[21] Note that, in cases 1 and 2, we have $\hat{p}_l^i(t) = p_l^i\big(t \mid \psi_{il}(q_i(\tau), \tau),\ \delta_t^i(\tau)\big)$ for all $l \leqq l_i(t)$.

part of the land price curve beyond district $l_1(t)$ is equal to that part of $p_l(0)$ curve. Building rent $r_{il}(t)$ curves during the growth period are given by equation (5.15) as before.

During stationary period $[\bar{t}, \bar{t}+\omega)$, there is, of course, no change in land use pattern nor in building use pattern. But, though its process cannot be determined uniquely, there is a big change in building rent curve $r_{il}(t)$ within period $[\bar{t}, \bar{t}+\omega)$. That is, suppose $\beta_i(t)$, $i = 1, 2$, are functions of time such that

$$\begin{aligned} &\beta_i(\bar{t}) = b_i(\bar{t}), \quad \beta_i(t) \leqslant b_i(t) \quad \text{and} \quad \dot{\beta}_i(t) \leqslant 0 \\ &\text{for all } t \in [\bar{t}, \bar{t}+\omega), \quad \text{and} \quad \beta_i(\bar{t}+\omega) = 0. \end{aligned} \tag{5.31}$$

Then, the following equations

$$\left.\begin{aligned} r_{1l}(t) &= -\dot{\beta}_1(t)+e^{-\gamma t}A_1\,\alpha_1(t)(d_{l_1(t)}-d_l), \\ r_{2l}(t) &= -\dot{\beta}_2(t)+e^{-\gamma t}A_2\,\alpha_2(t)(d_{\bar{l}}-d_l), \end{aligned}\right\} \quad \text{for } t \in [\bar{t}, \bar{t}+\omega), \tag{5.32}$$

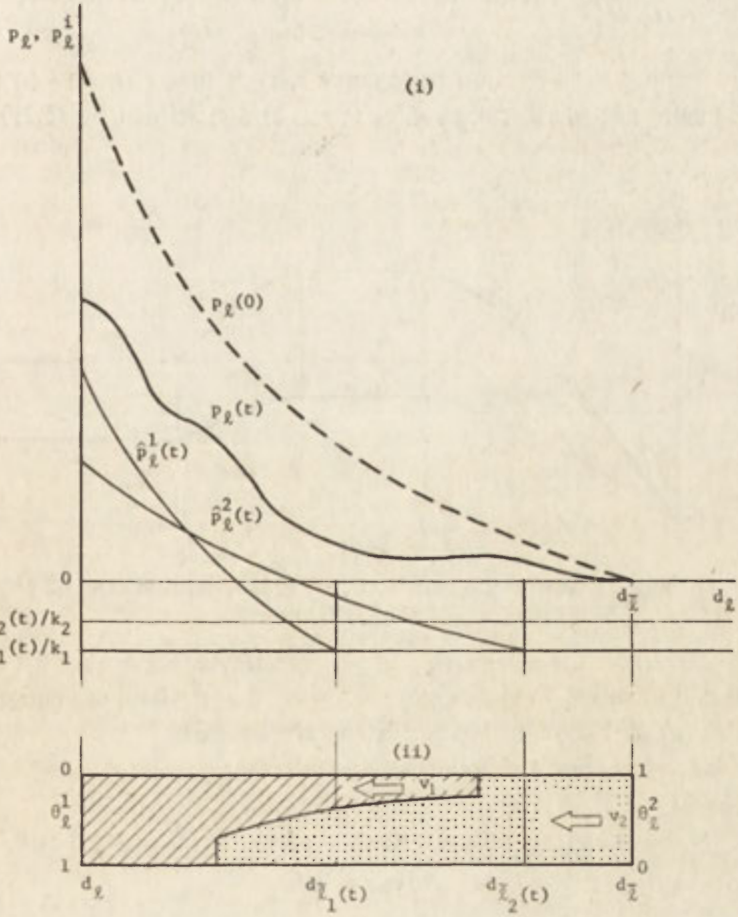


Fig. 11. (i) Building use pattern and (ii) land price curve $p_l(t)$ at a time t of urban contraction in case 3-1 (α) ($t > \bar{t}+\omega$)

represent a pair of rent curves (at each time $t \in [\bar{t}, \bar{t}+\omega)$) corresponding to the solution for problem OPA.[22] An example of time-variation of rent $r_{il}(t)$ in district l ($i = 1$ or 2, $1 \leqslant l < \bar{l}$) is depicted in Figure 10. Since function $\beta_i(t)$ must satisfy condition (5.31), rent $r_{il}(t)$ at each time in period $(\bar{t}, \bar{t}+\omega)$ becomes explosively high as the period of stationary state, ω, approaches to zero. In Figure 10, if $l = \bar{l}$, then $r_{i\bar{l}}(\bar{t}+\omega) = 0$; and if $i = 2$, then $l_i(\bar{t}) = \bar{l}$.

The building use pattern at a time t in the period of urban contraction (i.e., $t > \bar{t}+\omega$) is depicted in Figure 11. Let us denote by $\tilde{l}_i(t)$ ($i = 1, 2$, $t > \bar{t}+\omega$) the closest district to the city center among all the districts where there is vacancy in buildings of type i. District $\tilde{l}_i(t)$ is determined as such district l that satisfies the following relation:

$$N_i(t^*_{il-1}) \leqslant N_i(t) < N_i(t^*_{il}) \quad \text{where } t > \bar{t}+\omega. \tag{5.33}$$

Here t^*_{il-1} and t^*_{il} are determined by (5.5) and (5.6). Using $\tilde{l}_i(t)$, building rent curve $r_{il}(t)$ is given by

$$r_{il}(t) = \begin{cases} e^{-\gamma t} A_i \alpha_i(t)(d_{\tilde{l}_i(t)} - d_l), & \text{for } l = 1, 2, \ldots, \tilde{l}_i(t)-1, \\ 0 & \text{for } l \geqslant \tilde{l}_i(t), \end{cases} \tag{5.34}$$

for $i = 1, 2$ and $t > \bar{t}+\varepsilon$. Land price curve $p_l(t)$ at time t ($t > \bar{t}+\omega$) is depicted in (i) of Figure 11, where curves $\hat{p}^i_l(t)$ ($i = 1, 2$) are defined by (5.29).

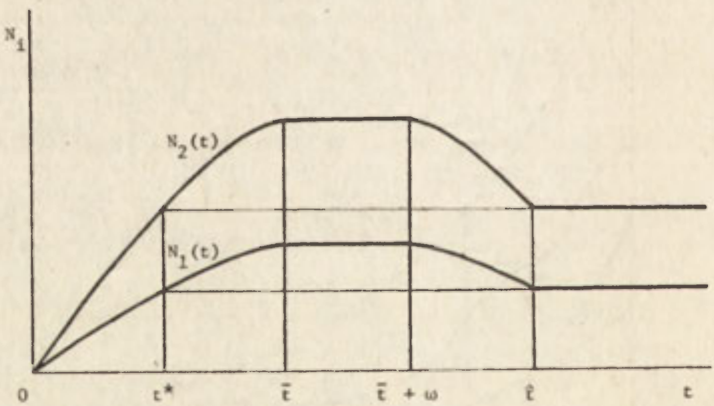


Fig. 12. Building demand functions $N_i(t)$, $i = 1, 2$, assumed in case 3-2 ($N_1(t) = \alpha N_2(t)$)

Spatial pattern of urban growth and contraction process in case 3-1 (β) is quite similar to that of case 3-1 (α) except for the point that the land use pattern at time $\bar{t}$ is given by (a) in Figure 8. Hence, details are omitted.

Next, let us examine the following special case:

case 3-2:

$$N_1(t) = \alpha N_2(t) \quad \text{for } t \in [0, \hat{t}], \quad \text{and} \quad N_1(\hat{t}) = N_2(\hat{t}) > 0, \quad \text{where } \alpha > 0. \tag{5.35}$$

[22] In (5.32), $l_1(t)$ is a fictitious construction district of building type 1. At time $t = \bar{t}+\omega$, we have $l_1(t) = \bar{l}$. For this, refer to footnote 18.

That is, in case 3-2, we assume (5.35) in addition to (5.10) and (5.20). Hence, the only difference between cases 3-1 and 3-2 is that, in case 3-2, the decrease of demand for buildings of each type stops before all the buildings of that type become vacant. An example of demand curves $N_i(t)$ $(i = 1, 2)$ under (5.20) and (5.35) is depicted in Figure 12. Let us define time t^* by

$$N_1(t^*) = N_1(\hat{t}) \quad (\text{i.e., } N_2(t^*) = N_2(\hat{t})) \quad \text{where } t^* < \hat{t}. \tag{5.36}$$

This t^* is depicted in Figure 12. Then, from definition (4.9) (and from Figure 12), we have

$$\left.\begin{aligned} T^i_{t^*_-} &= \{\tau \mid \tau \geqslant t^*\}, \\ T^i_{t^*_+} &= \{\tau \mid t^* \leqslant t \leqslant \hat{t}\}, \end{aligned}\right\} \quad i = 1, 2. \tag{5.37}$$

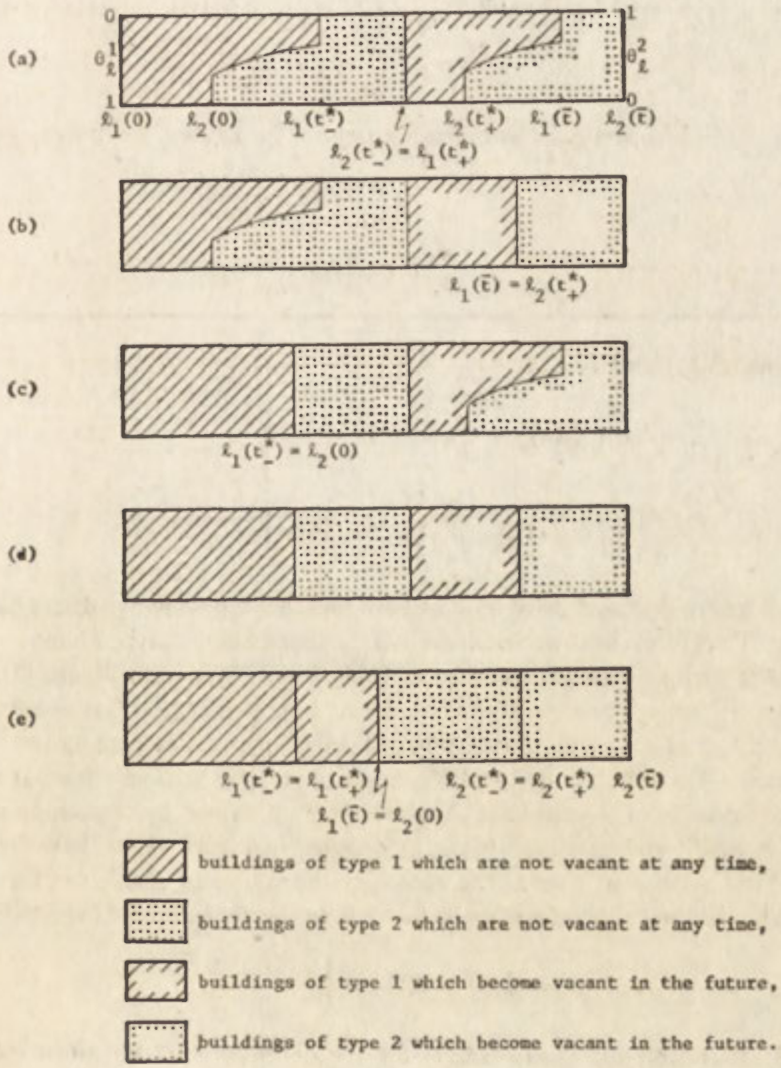


Fig. 13. Spatial patterns of urban land use at time $\bar{t}$ in case 3-2

TABLE 2. Relations between spatial patterns and conditions on parameters in case 3-2

Pattern	Conditions on parameters
(a)	$\frac{A_2}{k_2}\sigma_2(0) > \frac{A_1}{k_1}\sigma_1(t^*) > \frac{A_2}{k_2}\sigma_2(t^*) \geqq \frac{A_1}{k_1}\sigma_1(\delta_{t^*_+}(\tau)) > \frac{A_2}{k_2}\sigma_2(\delta_{t^*_+}(\tau)) > \frac{A_1}{k_1}\sigma_1(\delta_{\hat{t}}(\tau))$
(b)	$\frac{A_2}{k_2}\sigma_2(0) > \frac{A_1}{k_1}\sigma_1(t^*) > \frac{A_2}{k_2}\sigma_2(t^*) \geqq \frac{A_1}{k_1}\sigma_1(\delta_{t^*_+}(\tau)) > \frac{A_1}{k_1}\sigma_1(\delta_{\hat{t}}(\tau)) \geqq \frac{A_2}{k_2}\sigma_2(\delta_{t^*_+}(\tau))$
(c)	$\frac{A_1}{k_1}\sigma_1(t^*) \geqq \frac{A_2}{k_2}\sigma_2(0) > \frac{A_2}{k_2}\sigma_2(t^*) \geqq \frac{A_1}{k_1}\sigma_1(\delta_{t^*_+}(\tau)) > \frac{A_2}{k_2}\sigma_2(\delta_{t^*_+}(\tau)) > \frac{A_1}{k_1}\sigma_1(\delta_{\hat{t}}(\tau))$
(d)	$\frac{A_1}{k_1}\sigma_1(t^*) \geqq \frac{A_2}{k_2}\sigma_2(0) > \frac{A_2}{k_2}\sigma_2(t^*) \geqq \frac{A_1}{k_1}\sigma_1(\delta_{t^*_+}(\tau)) > \frac{A_1}{k_1}\sigma_1(\delta_{\hat{t}}(\tau)) \geqq \frac{A_2}{k_2}\sigma_2(\delta_{t^*_+}(\tau))$
(e)	$\frac{A_1}{k_1}\sigma_1(\delta_{\hat{t}}(\tau)) \geqq \frac{A_2}{k_2}\sigma_2(0)$

Thus, from (3.11) and (4.11),

$$\left.\begin{aligned} \frac{A_i}{k_i}\sigma_i(\delta^i_{t^*_-}(\tau)) &= \frac{A_i}{k_i}\sigma_i(t^*), \\ \frac{A_i}{k_i}\sigma_i(\delta^i_{t^*_+}(\tau)) &= \frac{A_i}{k_i}\int_{t^*}^{\hat{t}} e^{-\gamma\tau}\alpha_i(\tau)d\tau, \end{aligned}\right\} \quad i = 1, 2. \tag{5.38}$$

That is, slopes of bid land price lines of both building types change discontinuously at time t^*. Therefore, construction district $l_i(t)$ of each building type i jumps outward at time t^* (except for case (e) specified in Table 2 and depicted in Figure 13).

The spatial growth process of the city during time period $[0, t^*)$ is essentially the same as that of case 1; and, during time period $[t^*, \hat{t})$, it is essentially the same as that of case 3-1. Consequently, possible land use patterns at time $\hat{t}$ (i.e., at the end of the construction of the city) are those which are obtained by combining possible patterns in case 1 and case 3-1. Hence, as depicted in Figure 13, we have five different land use patterns at time $\hat{t}$. The necessary and sufficient conditions for realization of each pattern are summarized on Table 2. In this table, each $\delta_t(\tau)$ is defined by

$$\delta^1_t(\tau) = \delta^2_t(\tau) \equiv \delta_t(\tau) = \begin{cases} 1 & \text{if } \tau \in T_t, \\ 0 & \text{if } \tau \notin T_t, \end{cases} \tag{5.39}$$

for $t \in [0, \hat{t}]$. Recall (4.21) and (3.11) for the definitions of functions $\sigma_i(t)$ and $\sigma_i(\delta_t(\tau))$.

For example, pattern (a) is a combination of case 1-1 and case 3-1 (α). At time t^*, construction district of building type 1 jumps from $l_1(t_-^*)$ to $l_1(t_+^*)$ which is equal to $l_2(t_-^*)$, and construction district of building type 2 jumps from $l_2(t_-^*)$ to $l_2(t_+^*)$. Initial land price curve $p_l(0)$ is kinked in district $l_2(t_-^*)(= l_1(t_+^*))$. This pattern of growth process will happen when the city grows slowly during period $[0, \bar{t}]$.

For another example, pattern (e) is a combination of case 1-2 and case 3-1 (β). In this pattern, there is no jump of construction district of any building type; and there is no mixture of buildings of different types in any district. Initial land price curve $p_l(0)$ is kinked in districts $l_1(t^*)$, $l_1(\bar{t})$ and $l_2(\bar{t})$. This pattern will be realized only when the city grows very rapidly during period $[0, \bar{t}]$, and stays the same for a long period before the contraction starts.

Next, we examine the following special case:

case 3-3:

$$\hat{N}_1(\hat{t}) = \hat{N}_2(\hat{t}) = 0. \tag{5.40}$$

That is, we drop the assumption of the proportionality of demands for the two types of buildings which is assumed in case 3-1. An immediate consequence of this generalization is that relation (5.22) would no longer be true. Namely, under assumption (5.40) (in addition to (5.10) and (5.20)), we can only assert that

$$T_t^1 = T_t^2 (= T_t) \quad \text{for } t = 0 \text{ and } t \in [\bar{t}, \infty), \tag{5.41}$$

which only implies

$$\frac{A_1}{\dot{\kappa}_1} \sigma_1(\delta_t^1(\tau)) > \frac{A_2}{\dot{\kappa}_2} \sigma_2(\delta_t^2(\tau)) \quad \text{for} \quad t \in [0, \varepsilon) \text{ and } t \in (\bar{t}-\varepsilon', \bar{t}+\omega), \tag{5.42}$$

for some small positive ε and ε', and hence

$$l_1(t) \leqslant l_2(t) \quad \text{for} \quad t \in [0, \varepsilon) \text{ and } t \in (\bar{t}-\varepsilon', \bar{t}+\omega). \tag{5.43}$$

Since relations (5.41) to (5.43) would not always be true for $t \in (0, \bar{t})$, there might exist some differences in the construction process between case 3-1 and case 3-3. For example, it might happen in case 3-3 that

$$l_1(t) > l_2(t) \quad \text{for some } t \in (0, \bar{t}). \tag{5.44}$$

But, because of relations (5.41) to (5.43), spatial patterns of land use which are possible at time $\bar{t}$ in case 3-3 are the same as those in case 3-1 (i.e., patterns (a) and (b) in Figure 8).

5.5. SPATIAL PATTERNS OF URBAN GROWTH AND CONTRACTION: CASE 4

Finally, let us briefly examine the following case which is a generalization of case 3.

$$\dot{N}_i(t) \begin{cases} = 0 & \text{for } t \in [0, \underline{t}_i), \\ > 0 & \text{for } t \in [\underline{t}_i, \bar{t}_i), \\ = 0 & \text{for } t \in [\bar{t}_i, \bar{t}_i+\omega_i), \\ < 0 & \text{for } t \in [\bar{t}_i+\omega_i, \hat{t}_i), \\ = 0 & \text{for } t \in [\hat{t}_i, \infty) \end{cases} \tag{5.45}$$

for $i = 1, 2$, where

$$0 \leqslant \underline{t}_i < \bar{t}_i < \bar{t}_i + \omega_i < \hat{t}. \tag{5.46}$$

We keep assumption (5.10) here.

If we further assume that

case 4-1:

$$N_i(\hat{t}_i) = 0 \quad \text{for } i = 1, 2, \tag{5.47}$$

then it is not difficult to see that the spatial pattern of land use at the time, $\max\{\bar{t}_1, \bar{t}_2\}$, should be one of those patterns in Figure 8. Namely, since there is no discontinuous change in sets T_t^i, $i = 1, 2$, at any time, we see from Property A.26 that the area where buildings of each type are constructed is spatially connected. Hence, patterns (a) to (g) depicted in Figure 8 exhaust all the possible patterns of land use at time, $\max\{\bar{t}_1, \bar{t}_2\}$, under assumptions (5.10) and (5.45) to (5.47). Of course, the relation, $l_1(t) \leqslant l_2(t)$, would not always be true for all $0 \leqslant t \leqslant \max\{\bar{t}_1, \bar{t}_2\}$.

If we assume, instead of (5.47), that

case 4-2:

$$N_i(\hat{t}_i) > 0 \quad \text{for } i = 1, 2, \tag{5.48}$$

then set T_t^i changes discontinuously at time t_i^* which is defined by

$$N_i(t_i^*) = N_i(\hat{t}_i) \quad \text{and} \quad t_i^* < \hat{t}_i, \quad i = 1, 2. \tag{5.49}$$

Hence, the area for buildings of each type might have one spatial-disconnection. Therefore, the set of all the possible patterns of land use at time $\max\{\bar{t}_1, \bar{t}_2\}$ in case 4-2 is obtained by combining two sets of those patterns depicted in Figure 8 (refer to case 2).

6. SPATIAL PATTERNS IN THE CASE OF MANY BUILDING TYPES ($m > 2$)

In this section, we briefly examine spatial patterns of urban growth and contraction in the case of many building types. As in the previous section, we employ (b) of Assumption OPA-3 and Assumption OPA-4 throughout this section.

6.1. CONSTRUCTION PROCESSES

Given demand function $N_i(t)$ for buildings of each type i, occupancy set T_t^i is obtained from (4.9). Then, the optimal utilization plan $\delta_t^i(\tau)$ is defined by (4.11). Consequently, slope $(A_i/k_i)\sigma_i(\delta_t^i(\tau))$ of bid land price line $p_t^i(t \mid \psi_{il}(q_i(\tau), \tau), \delta_t^i(\tau))$ can be calculated for each $t \in [0, \infty)$, $i = 1, 2, \dots, m$, and $l = 1, 2, \dots, g$. And, we can assume, without loss of generality, that

$$\frac{A_1}{k_1}\sigma_1(\delta_0^1(\tau)) \geqslant \frac{A_2}{k_2}\sigma_2(\delta_0^2(\tau)) \geqslant \dots \geqslant \frac{A_m}{k_m}\sigma_m(\delta_0^m(\tau)). \tag{6.1}$$

Next, from (4.8), function $N_i^*(t)$ is obtained for $t \in [0, \infty)$ and $i = 1, 2, \dots, n$. When $N_i^*(t) = 0$, we assume that zero buildings of type i are being constructed in the city at time t; and hence, by employing the same convention which is explained in the beginning of section 5.1, we can consider that construction district $l_i(t)$ exists for each building type i $(i = 1, 2, \ldots, m)$ at any time $t \in [0, \infty)$.

From (6.1) and Property A.27, it must be true that

$$l_1(0) = 1 \leqslant l_2(0) \leqslant l_3(0) \leqslant \ldots \leqslant l_m(0). \tag{6.2}$$

To obtain district $l_2(0)$, consider the following system of simultaneous equations for each l $(l = 1, 2, \ldots, g)$:

$$\frac{A_1}{k_1}\sigma_1(\delta^1_{t_{1l}}(\tau)) = \frac{A_2}{k_2}\sigma_2(\delta^2_{t_{2l}}(\tau)),$$
$$\sum_{i=1}^{2} k_i N_i^*(t_{il}) = \sum_{j=1}^{l} s_j. \tag{6.3}$$

Then, district $l_2(0)$ is obtained as the smallest (positive integer) l for which values of t_{1l} and t_{2l} calculated from (6.3) are both positive.[23] Next, consider the following system of simultaneous equations for each l $(l = 1, 2, \ldots, g)$:

$$\frac{A_1}{k_1}\sigma_1(\delta^1_{t_{1l}}(\tau)) = \frac{A_2}{k_2}\sigma_2(\delta^2_{t_{2l}}(\tau)) = \frac{A_3}{k_3}\sigma_3(\delta^3_{t_{3l}}(\tau)),$$
$$\sum_{i=1}^{3} k_i N_i^*(t_{il}) = \sum_{j=1}^{l} s_j, \tag{6.4}$$

Then, district $l_3(0)$ is given as the smallest l for which values of t_{1l}, t_{2l} and t_{3l} obtained from (6.4) are all positive. Similarly, we can obtain $l_4(0)$, $l_5(0)$, ..., $l_m(0)$.

Next, solving the following equation

$$k_1 N_1^*(t_{1l}) = \sum_{j=1}^{l} s_j,$$

for t_{1l} for each $l = 1, 2, \ldots, l_2(0)-1$, we get construction switching times t_{1l}^* for $l = 1, 2, \ldots, l_2(0)-1$. And, solving (6.3) for (t_{1l}, t_{2l}) for each $l = l_2(0), l_2(0)+1, \ldots, l_3(0)-1$, we obtain construction switching times t_{1l}^* and t_{2l}^* for $l = l_2(0), l_2(0)+1, \ldots, l_3(0)-1$. Similarly, from (6.4), we can obtain t_{1l}^*, t_{2l}^* and t_{3l}^* for $l = l_3(0), l_3(0)+1, \ldots, l_4(0)-1$. Repeating the same procedure we can get construction switching time t_{il}^* for each i $(i = 1, 2, \ldots, m)$ and l $(l = l_i(0), l_i(0)+1, \ldots, \bar{l}-1)$, where fringe district $\bar{l}$ is determined by (4.14).

By using these t_{il}^*, the optimal construction process $u_{il}(t)$ $(i = 1, 2, \ldots, m,$ $l = 1, 2, \ldots, g, 0 \leqslant t < \infty)$ is given by the same equation as (5.7); the number of buildings of type i constructed in district l is by (5.8). And the district of construction $l_i(t)$ is given at each time t by

$$l_i(t) = l \quad \text{for } t_{il-1}^* \leqslant t < t_{il}^*, \quad l = l_i(0),\ l_i(0)+1, \ldots, \bar{l}. \tag{6.5}$$

[23] If we do not have such l, then $l_2(0) = 1$.

6.2. SOME SPECIAL CASES

Let us first examine spatial patterns of urban growth under the following set of assumptions:

case 5:

$$\left.\begin{array}{ll} \dot{N}_i(t) > 0 & \text{for } 0 \leqslant t < \bar{t}_i, \\ \dot{N}_i(t) = 0 & \text{for } t \geqslant \bar{t}_i, \end{array}\right\} \quad i = 1, 2, \ldots, m, \tag{6.6}$$

and

$$\frac{A_1 \alpha_1(t)}{k_1} > \frac{A_2 \alpha_2(t)}{k_2} > \ldots > \frac{A_m \alpha_m(t)}{k_m} \quad \text{for all } t \geqslant 0. \tag{6.7}$$

Under assumption (6.6), $T_t^i = \{\tau |\ \tau \geqslant t\}$ for each $t \in [0, \infty)$ and for each i ($i = 1, 2, \ldots, m$). Hence, under assumption (6.7), we have from (4.20) and (4.21) that

$$\frac{A_1}{k_1}\sigma_1(t) > \frac{A_2}{k_2}\sigma_2(t) > \ldots > \frac{A_m}{k_m}\sigma_m(t) \quad \text{for all } t \geqslant 0, \tag{6.8}$$

and thus, from (4.13),

$$l_1(t) \leqslant l_2(t) \leqslant \ldots \leqslant l_m(t) \quad \text{for all } t \geqslant 0.^{24} \tag{6.9}$$

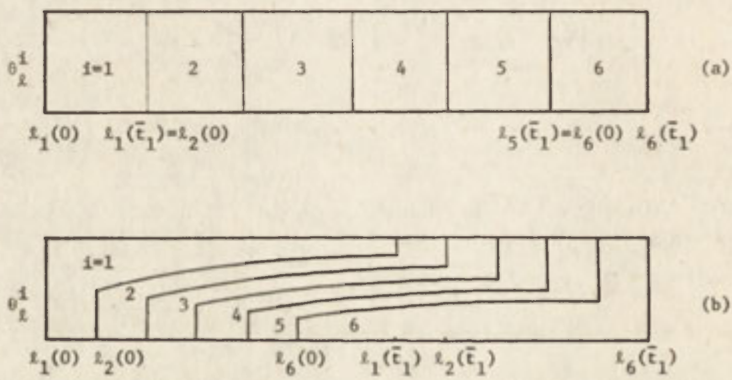


Fig. 14. Examples of spatial patterns of urban land use at the end of the growth of the city in case 5.

Figure 14 depicts two (extreme) examples of spatial patterns of urban land use at the end of the growth of the city in case 5. In pattern (a), there is no mixture of different building types in any district; this pattern can be realized only when the entire city is constructed in a very short time period. In pattern (b), many districts have a mixture of buildings of all the types; this pattern will be realized when the city grows slowly over a long time period.

Next, let us examine the following special case:

case 6:

$$\left.\begin{array}{lll} \dot{N}_i(t) > 0 & \text{for} & t \in [\underline{t}_i, \bar{t}_i) \\ \dot{N}_i(t) = 0 & \text{for} & t \notin [\underline{t}_i, \bar{t}_i) \end{array}\right| \quad i = 1, 2, \ldots, m, \tag{6.10}$$

[24] If each district is sufficiently small, then we have

$$l_1(t) < l_2(t) < \ldots < l_m(t) \quad \text{for all } 0 \leqq t \leqq \max_i \bar{t}_i.$$

and

$$\frac{A_1\alpha_1(t)}{k_1} > \frac{A_2\alpha_2(t)}{k_2} > \ldots > \frac{A_m\alpha_m(t)}{k_m} \quad \text{for all } t \geqslant 0. \tag{6.11}$$

In this case, there are quite a large number of possible patterns (refer to case 2 and Figure 8), and the actual spatial pattern depends on the specific values of the parameters. But, except for a very special case, buildings of many different types will be mixed in many districts.

We close the discussion by investigating the following situation. Given a demand function $N_i(t)$ $(0 \leqslant t < \infty)$ for a building type i, we define $\underline{t}_i$ and $\bar{t}_i$ by

$$\underline{t}_i = \inf\{t|\dot{N}_i^*(t) > 0\}, \tag{6.12}$$

$$\bar{t}_i = \sup\{t|\dot{N}_i^*(t) > 0\}.$$

Then, the construction of buildings of type i begins in district $l_i(0)$ $(= l_i(\underline{t}_i))$ at time $\underline{t}_i$, moves outward with time, and ends in district $l_i(\bar{t}_i)$ at time $\bar{t}_i$. And, a question of interest is whether all the districts in which buildings of type i are constructed are spatially connected or not; if not, how many disconnections there are. It is not difficult to obtain the following answer.

Property A.31. Given a demand function $N_i(t)$ $(0 \leqslant t < \infty)$ for buildings of type i, we define $\underline{t}_i$ and $\bar{t}_i$ by (6.12). Then, the maximum possible number of disconnections of the area on which buildings of type i are constructed is given by

$$n_i + n_i'$$

where

n_i: the number of (different) flat intervals of $N_i^*(t)$-curve between $t = \underline{t}_i$ and $t = \bar{t}_i$,

n_i': the number of flat intervals of $N_i(t)$-curve which are not on $N_i^*(t)$-curve nor on t-axis (i.e., horizontal axis).

For example, in the case of Figure 3, $n_i = 1$ and $n_i' = 1$. In the case of Figure 12, $n_1(= n_2) = 0$ and $n_1'(= n_2') = 1$.

7. CONCLUDING REMARKS

We might consider that this paper has succeeded in characterizing spatial patterns of urban growth and contraction in the framework of Problem A. But, the most serious shortcoming in the framework of Problem A is that the demand for buildings of each type is given exogenously in the problem. Hence the next task would be to study the simultaneous determination of building demands and building locations within a single model. The first step in this direction would be to solve Problem C which is proposed in Fujita (1976b). That is, assuming that there is only one type of household in the city at each time, we study the development pattern of residential area over time.

Mathematically, Problem A and Problem C are very closely related. That is, in the case of Problem C, denote by $y_{il}(t)$ the number of households which occupy residential houses of type i in district l at time t, and by $N(t)$ the total number of households in the city at time t which is exogenously given. Then, if we add the following restriction

$$\sum_i N_i(t) = N(t)$$

to Problem A and if consider that each $N_i(t)$ is a variable, we obtain the formulation of Problem C.

Suppose we have solved Problem C and obtained $N_i(t) = \hat{N}_i(t)$ in the solution. Then, if we solve Problem A with parameters $N_i(t) = \hat{N}_i(t)$ $(i = 1, 2, \ldots, m)$, we should obtain the same solution as the original Problem C. Therefore, we can conclude that the spatial pattern of urban growth in Problem C is just one of the possible patterns from Problem A. But which one? In particular, is it close to Pattern(a) in Figure 14 or Pattern (b) in Figure 14? This question must be answered in the future.

REFERENCES

Arrow, J. K. and M. Kurz, 1970, *Public Investment, the Rate of Return and Optimal Fiscal Policy*, Baltimore, John Hopkins Press.

Boltyanskii, V., 1968, *The Mathematical Method of Optimal* Control, Tokyo, Sogo-Tosho (Japanese translation).

Fujita, M., 1976a, Spatial Patterns of Urban Growth: Optimum and Market, *Journal of Urban Economics*, 3, pp. 209–41.

Fujita, M., 1976b, Toward a Dynamic Theory of Urban Land Use, *Papers of the Regional Science Association*, 37, pp. 133–65.

Halkin, H., 1974, Necessary Conditions for Optimal Control Problems with Infinite Horizons, *Econometrica*, 42, 2 (March), pp. 267–72.

Hestenes, M. R., 1966, *Calculus of Variations and Optimal Control Theory*, New York, John Wiley & Sons.

Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishenko, 1966, *The Mathematical Theory of Optimal Processes*, New York, John Wiley & Sons.

DYNAMIC ORGANIZATION OF SOCIO-ECONOMIC SPACE

RYSZARD DOMAŃSKI

Poznan Academy of Economics, Poznan, Poland

The aim of the present paper is specification of theoretical conceptions, which may form construction elements of the model of dynamic organization of socio-economic space. There is a need for constructing such a model. It results from a trend towards steering of socio-economic processes in space and reshaping of spatial systems of economy and society in order to give them direction corresponding with socio-economic objectives. Steering and reshaping in order to be effective and to lead towards an objective, should be based on knowledge of mechanism of processes and rules according to which spatial systems are organized. Hence, the theory of processes and dynamic organization of socio-economic space is needed.

Attempts to introduce dynamic elements into spatial concepts of economy and society have been taken up by many authors. They have different approaches. One of them is to create a dynamic alternative for Christaller's static theory. The present paper may be regarded as an attempt of this kind. It concerns the introductory stage of research: it selects and juxtaposes former conceptions, which may be helpful in constructing alternative theory. The juxtaposition of these construction elements shows that a lot has already been done and presently the main task is to make a synthesis.

If we attempt at a theory, the character of which should be dynamic it is advisable that its construction from the very beginning should be based on dynamic elements. Movements taking place in spatial systems of economy and society are such elements. One can distinguish the following types of these movements.[1]

1. Induced movements. The main kind of these movements are movements induced by localization of new investments (of industrial, transport, service, academic etc. character). Establishment of administrative authority in a town of previously lower administrative status may also considerably induce such movements. Movements to and from towns, which became capitals of new districts created as a result of territorial administration reform in Poland in 1975, are a distinct example. Localization and realization of large investments induce the following movements: relocation of building enterprises, organization of primary technical and social infrastructure, migration of employees and their families, commuting, flows of in-

[1] Cf. R. Domański (1977).

vestment and production goods, organization of a new production apparatus, inflow innovations in many fields of socio-economic activity, reorganization of local and regional administration, clashing of value systems and culture patterns of immigrated population and local population, signs of social disorganization.

2. Adaptation movements. They depend mainly on adjustment of the third and fourth sector of economy to the needs of the developing first and second sectors as well as on further migration movements, further innovation inflow, gradual for-

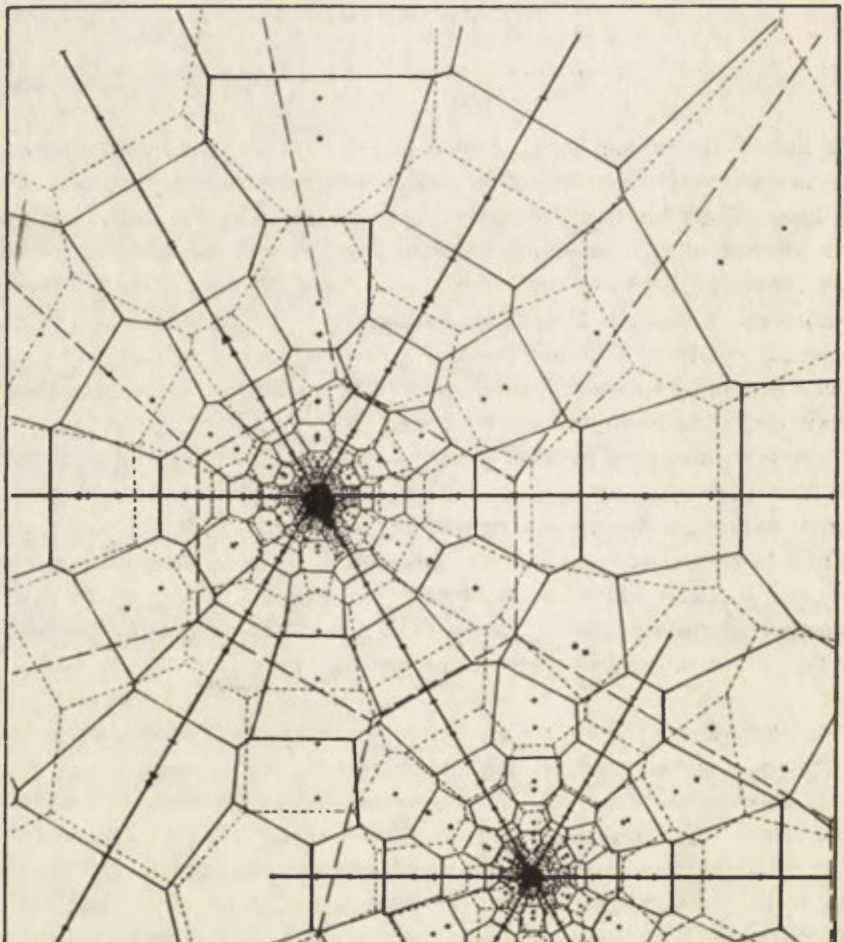


Fig. 1. Economic landscape dominated by central places. After W. Isard (1956, p. 272)

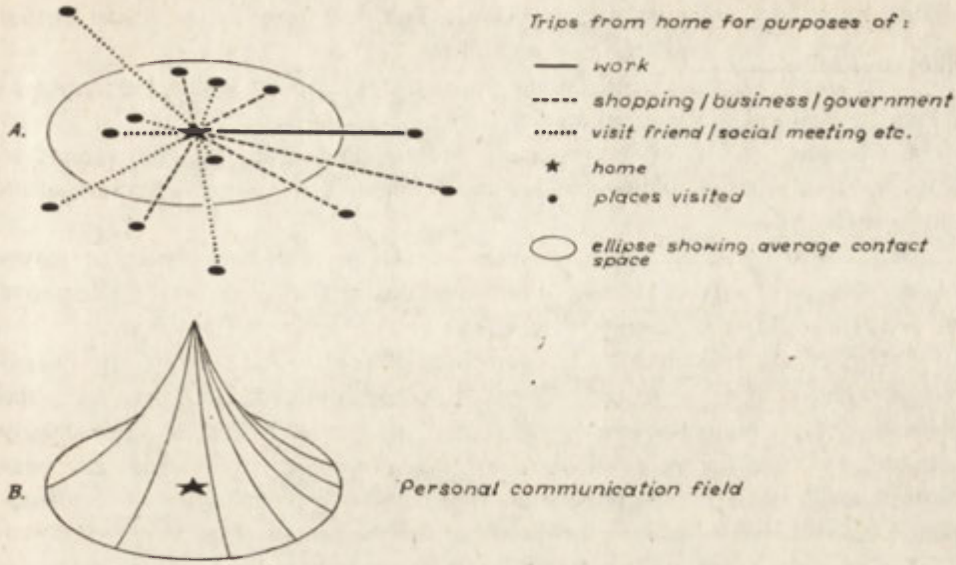


Fig. 2. Daily contact space and personal communication fields.

A person is usually counted for census purposes at his place of residence. But people are not always at home. Part *A* of the diagram represents a typical collection of other places visited at varying frequencies and distances from home. The ellipse shows the average distance to these other locations appropriately weighted by frequency of contact. It is only within this space that a person has contact with other people and with communication devices. Part *B* shows an idealized personal communication field which indicates the range and intensity of potential contact. This idealized cone would vary by culture and by individual. After J. F. Kolars and J. D. Nystuen (1974, p. 126)

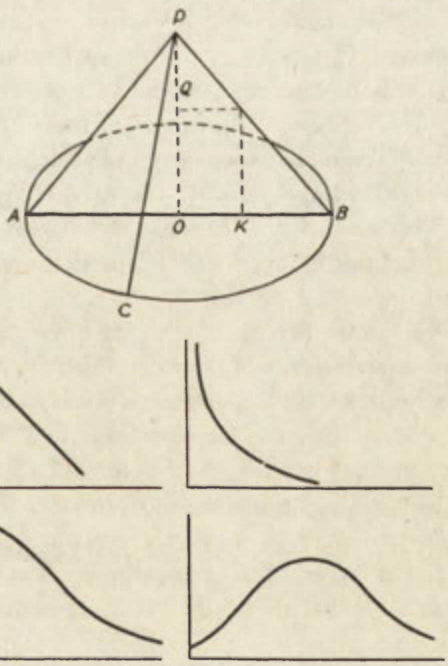


Fig. 3. Functioning of the growth point *O* — above, and some possible deviations of *PA*, *PB*, and *PC*; lines below — (the fourth case, below to the right, is the least possible form). After R. P. Misra (1971)

ming of a new local community, new culture and new forms of social life, further development of local and regional administration.

3. Autonomic development, i.e. development taking place within the system as a result of influence of internal factors.

4. Continuation of former processes. Induced and adaptation movements as a rule overlap processes induced earlier and being at various stages of continuation and extinction.

Question arises: what spatial regularities take place in this variety of movements. Observed regularities may constitute construction elements of alternative theory. The following elements may be useful in attempts of synthesis.

1. Central system. Central system composed of a centre and a subsidiary connected with the centre by means of transportation and communication network is the elementary form of spatial movements formation. Central places have particularly suitable conditions for the development of socio-economic activity. They also concentrate large masses of the population. The density of population and socio-economic activity diminishes with the distance from the centres (Fig. 1).

2. City and regional daily cycle. Many movements on a city and regional scale, particularly population movements, show distinctly a daily cycle (Fig. 2). Particularly it concerns commuters in cities and urban-industrial agglomerations. In the goods flows, the daily cycle can be seen in the supply of shops and trade centres, as well as some industrial enterprises (e.g., dairies). Directions of these movements are relatively stable and easy to define.

3. Inflow of development factors into attraction niches. Poles of growth. Concentration of development factors in attraction niches turns them into poles of growth.[2] From there development impulses spread to both the closer and the farther subsidiaries (Fig. 3). Processes of spreading of these impulses are accompanied by processes of attracting valuable socio-economic elements from subsidiaries to the development centres (backwash effect). The existence of cumulative feedbacks induces cumulative processes which, in non-stable sytems, may cause deepening of socio-economic inequalities between well and underdeveloped regions, in favour of the well developed ones.

4. Locational changes in the process of ripening of new activities. New industries and services in the subsequent stages of their development show characteristic changes in location.[3] They appear first as single enterprises in big cities with high innovation propensities. Along with the development, new technologies stop being experimental and they become routine. As a result their locational dependence on the centres with high innovation propensities decreases. Simultaneously, new products and services win a wider and wider market which induces the necessity of creating further manufacturing and service enterprises. They often find better location conditions outside their mother centres. It is easier for them to gain building

[2] Compare the work of Soviet authors on territorial production complexes. They were started by N.N. Kolosovskii (1947, 1971). The work edited by M. K. Bandman (1976) presents the actual state of research.

[3] E. M. Hoover (1971, pp. 150–151).

land and labour force in medium size and small cities. The number of enterprises increases and location pattern from a unicentral changes into the multicentral one. In mother centres which are big cities difficulties in further development of pioneer enterprises may appear. These centers gradually lose their exceptional position in a given sector of industry and services. As the time passes, the pioneer establishments may be liquidated and production may concentrate in later established enterprises. In this way a city which has given birth to a new sector of industry and services and developed it in the period of childhood may lose it in the period of maturity in favour of smaller cities where the level of innovation is lower. An example here may be New York which lost almost all its sectors of industry established there in the early period of industrialization: milling, steel, meat-processing, textile and tanning industries. Similarly Pittsburgh was at first the pioneer in some sectors of industry in the United States (crude oil refineries, aluminium plants, electrical equipment) but later it lost its superiority.

Repetition of such development cycle is responsible for the fact that big cities in their economic structure have a big share of new sectors, but together with reaching the stage of maturity, the share of these cities in the domestic output of given sectors decreases and sometimes completely declines. Simultaneously, the share of small cities increases. It is, however, an increase based on establishments which are no longer innovative, therefore, less dynamic. As a result, the general economic development of smaller cities is from the very beginning less dynamic. An indispensable condition of granting these cities a bigger dynamics is transferring into them subsequent sectors of industry and services at an early stage of their life cycle, that is when they create many new jobs and require qualified employees, well paid, of a high degree of innovation and a high level of culture.

The presented pattern of development is most suitable for countries of a high economic potential and a high innovation propensities. In medium and small countries one enterprise in a given sector of industry and services (often based on imported technology) is sometimes sufficient. Therefore, no new locations appear and the multicenter location pattern is not formed. In the process of development only the increase in scale and modernization of production in this enterprise take place.

5. Life tracks of persons and families. An individual and his family in the course of his life cycle make characteristic spatial movements. Because of their large scale character they constitute an essential factor of the dynamic organization of space. These movements change in a characteristic way when passing to the subsequent phases of life cycle. Taken together, for all phases, they form life tracks of persons and families. On these tracks there are stations, at which persons and families stop for shorter or longer periods of time. Spatial mobility in childhood is limited, it increases in adolescent and adult age and it weakens in the old age. Variability concerns geographical distance, frequency and purpose of movements and stability of shifts. In adolescent age movements towards schools and academic centres prevail, in adult age to places of employment and successivee centres of employment. In the final years of adult phase movements made for professional reasons weaken. It concerns particularly migrations to successive centres of employment. In the old

age together with increase of affluence, society's movements aimed at recreation, medical tratment and social life become stronger.

6. Principle of gravitation. Spatial movements show a tendency of arranging themselves in a way which can be presented by means of gravitation model. They are directly proportional to volume of cities (regions) and inversely proportional to the distance between them. Volume can be characterized in various ways, e.g. by means of the number of population, output, incomes, etc. Distance can be formulated in geographical categories and expressed in kilometres. In characterizing socio-economic contącts, however, more appropriate is the idea of relative distance comprising the cost, time and effort necessary to cover the geographical distance.

7. Intervening opportunities. Places (regions) which occupy intermediate position exert an influence on movements between more distant places (regions). Intermediate position means position closer than place of destination not necessarily, however, limited to definite and particularly not to the shortest transport line. Migrants looking for new opportunities outside their previous place of living, may not reach centres of very high opportunities, if there are satisfactory opportunities nearer. For example, when a person from Grodzisk Wlkp., who possesses secondary education, considers taking up studies in Poznań or Warsaw academic centres, will most probably choose Poznań centre. The same concerns a person, who has finished secondary vocational school and is looking for a job. Intervening opportunities organize also movements of loads and movements of information.

8. Dependence of organization of spatial movements on institutional structure. Most of all the degree of centralization or decentralization of political, administrative, economic, social and cultural systems exerts an influence on the formation of spatial movements. Size and shape of territorial units making up the system, are the next determinant.

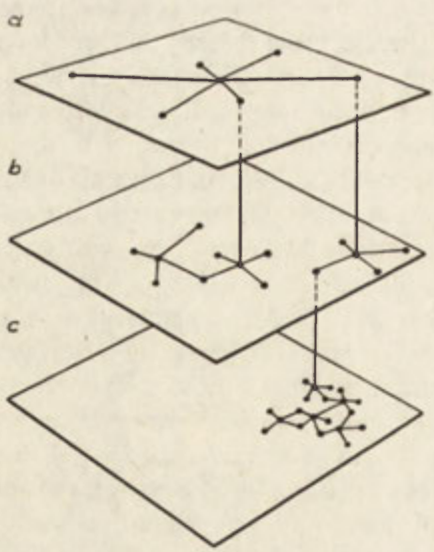


Fig. 4. Information flow in hierarchic system of nodal points *a*, *b*, *c* — levels of gradually decreasing spatial range. T. Hägerstrand after R. P. Misra (1971)

9. Spatial coincidence. It concerns not only the co-existence of movements in the same centres, but also the way of their spreading from the centres as well as relocation of objects and activities between centres. If, for example, in the construction of a big industrial plant three industrial centres (apart from many others) participate to the largest extent, from these centres come investment and production equi-

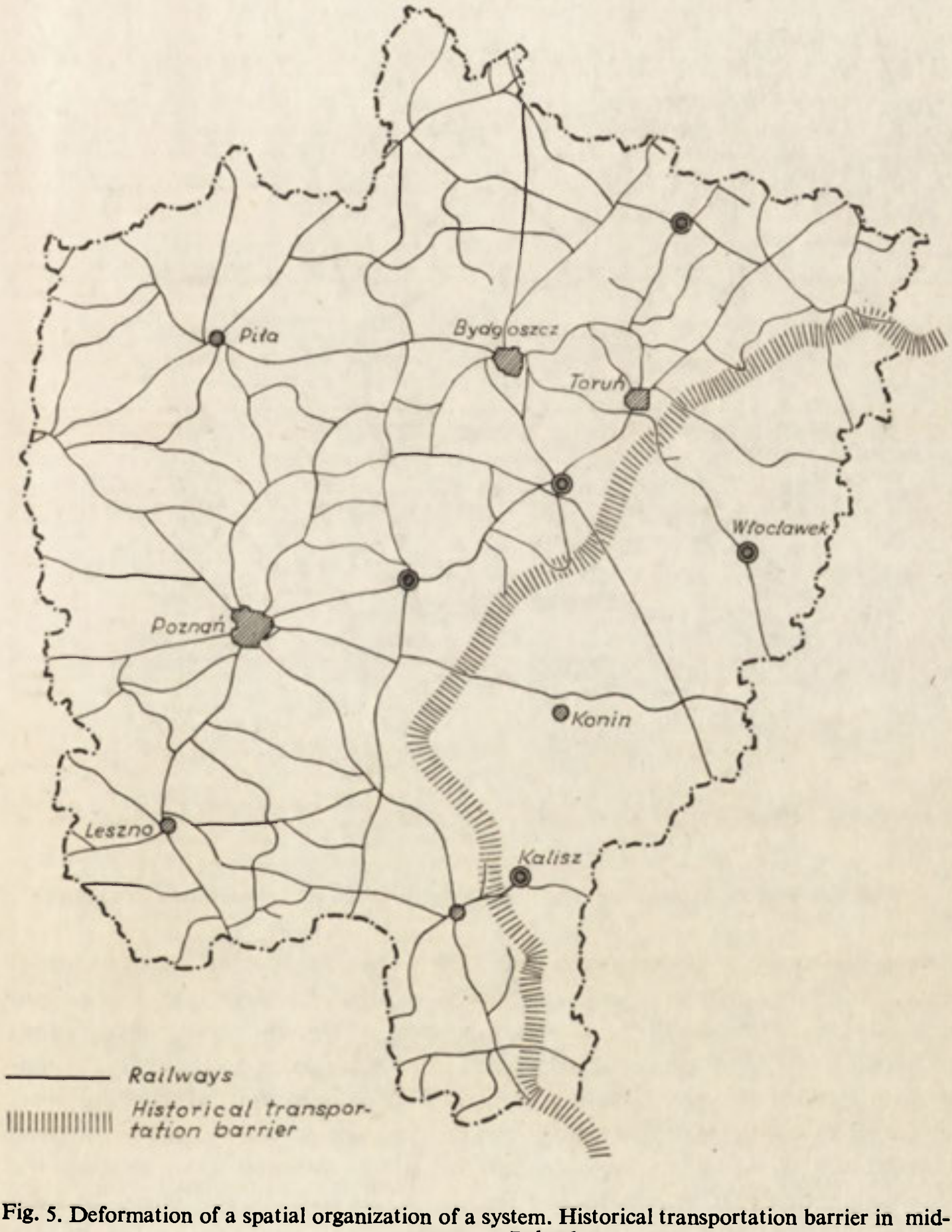


Fig. 5. Deformation of a spatial organization of a system. Historical transportation barrier in mid-western Poland

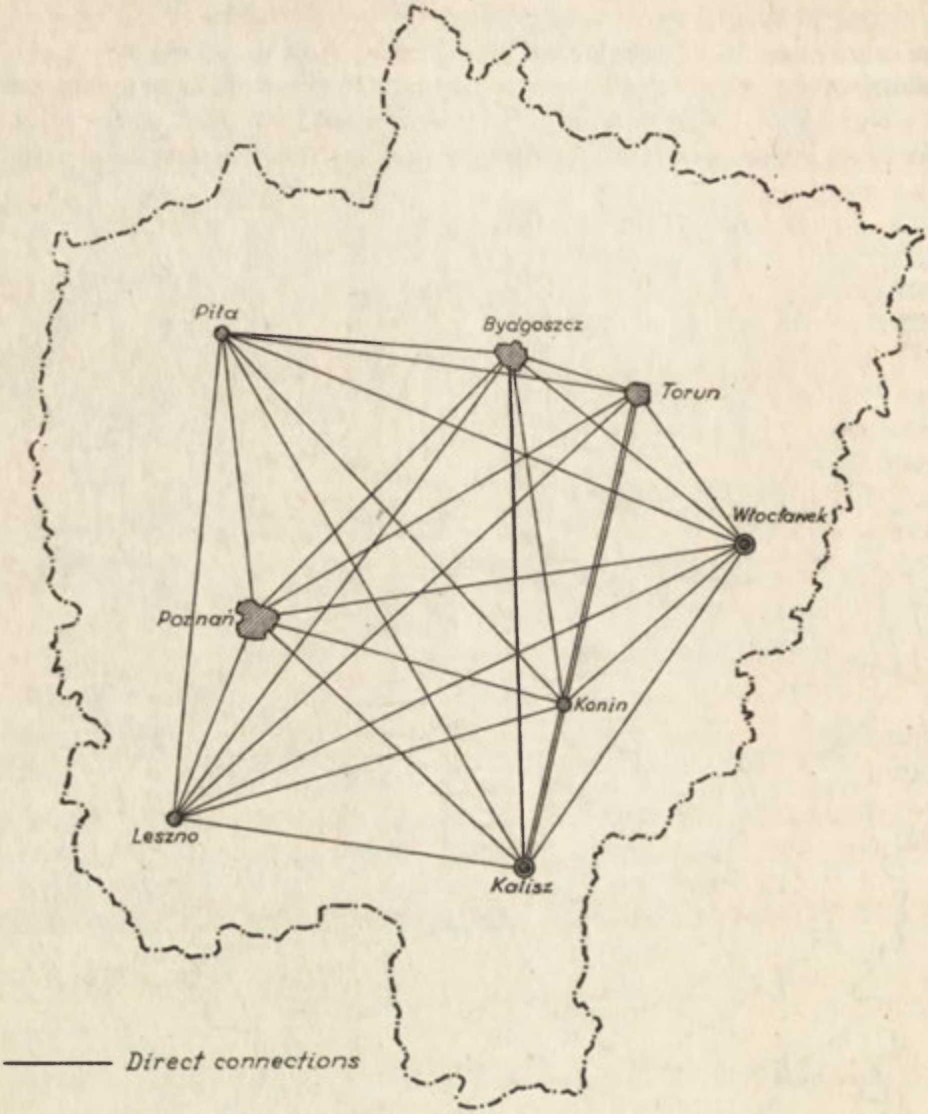


Fig. 6. Complete connectivity of the spatial system. The mid-western part of Poland

pment, engineering and administrative staff, employees' families, models of industrial organization, higher school graduates, the way of building flats, the way of land utilization, systems of values and culture patterns. If the city gains a higher administrative status, it causes reorientation and intensification of movements: commuting, migration, business travels, commodity flows, spreading of building sites, spreading of city transport network, growth of service centres. It is characteristic that some of the movements have the same or similar direction. Coincidence is the sign of interdependence occurring in the process of spatial development between different kinds of socio-economic activity. The term coincidence does not mean here

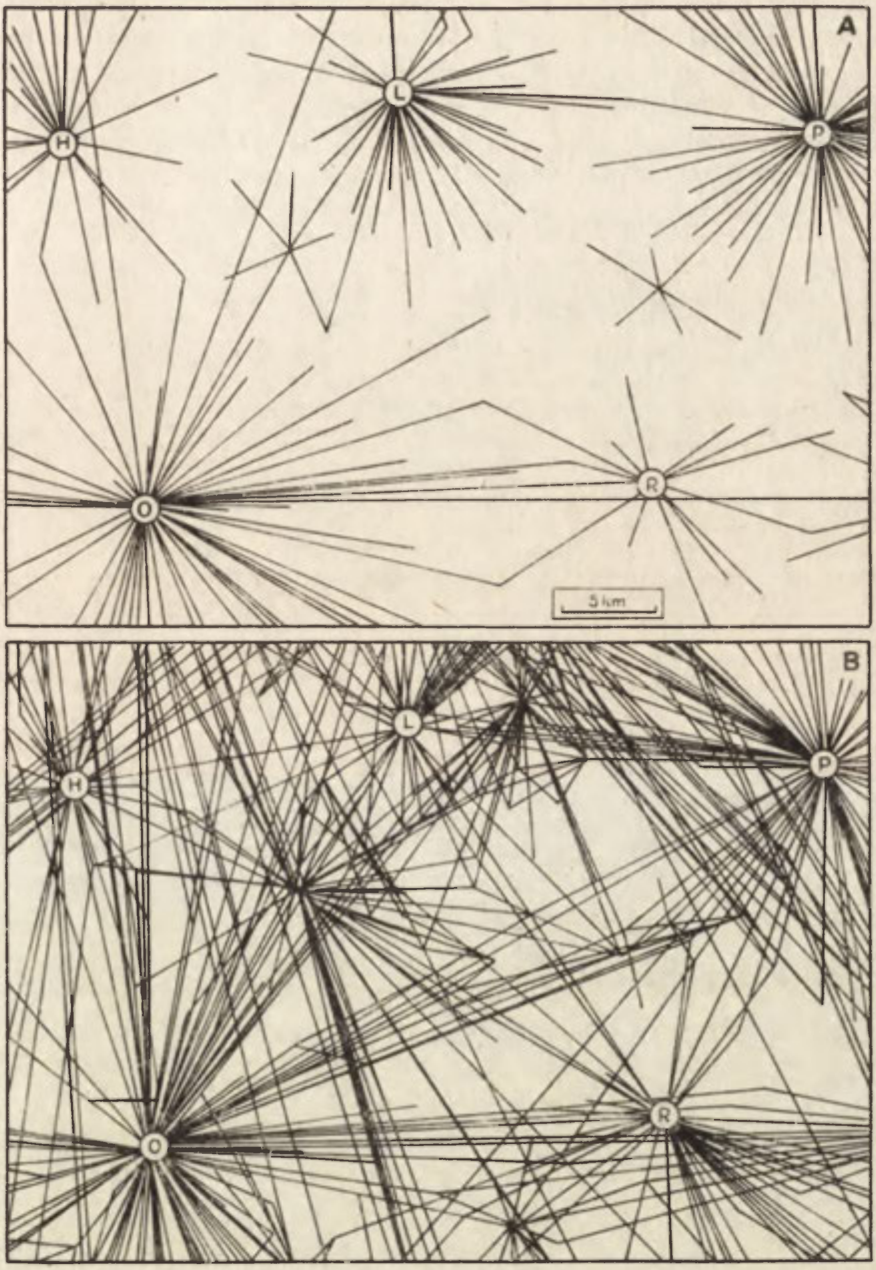


Fig. 7. Development of transport links

A — 1886, B — 1972 (part of Nord and Pas de Calais Departments, France), H — Hazebrouck, L — Lillers, O — St. Omer, P — St. Pol, R — Fruges, After P. Toyne (1974, p. 242)

the exact overlapping of movement directions; it only means their tendency to concur.

10. Hierarchic diffusion of socio-economic phenomena. The real socio-economic space is relative to geographical space. In the relative space distances between big cities linked by means of strong information flows are shorter than geographical distances. Information flowing between big cities skips smaller towns on its way. Information reaches medium size places sooner than smaller places. In this way the process of diffusion becomes hierarchic (Fig. 4). In the hierarchic diffusion the

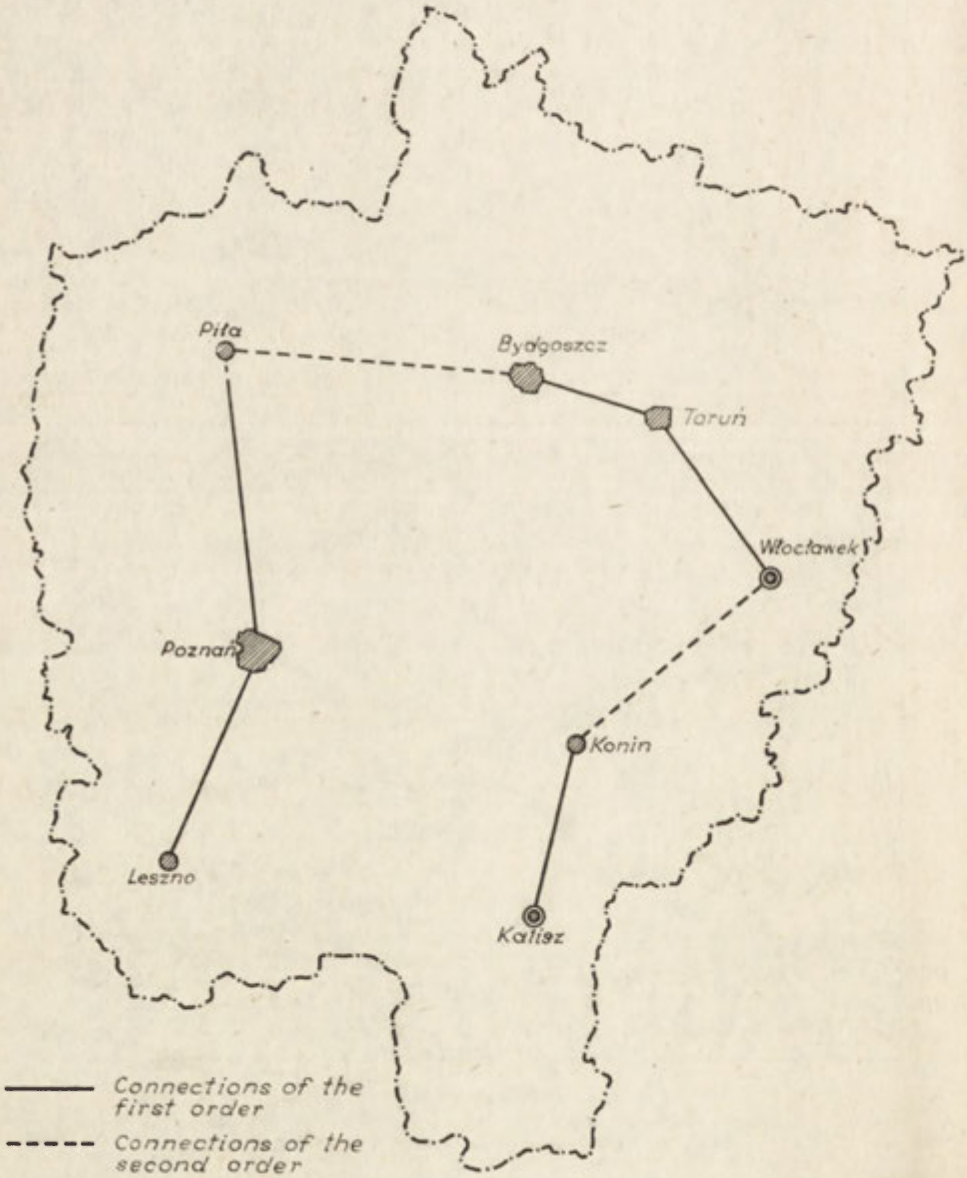


Fig. 8. Minimal connectivity. Graph spread on the capitals of the voivodeships in mid-western Poland

frequency of contacts as well as the time and stock of information received by a decision-maker depend on his location. Location of the decision-maker influences in this way the quality of his decision.

11. Spatial barriers. These barriers cause deformation of spatial movement organization from regular patterns which are optimal and most probable. They may be of different character. The main types of barriers are: natural barriers, inertion of production and social infrastructure, imperfect mobility of population, institutional barriers, historical barriers. Their influence on the organization of spacial movements is also different. For example, natural barriers cause the development and supremacy of prolonged movement patterns and policentric movement concentrations, inertion of infrastructure — concentration of movement, similarly imperfect mobility of population. Institutional barriers, in the case of centralized economic system may induce the prolongation of movement distances while historical barriers — deviation of directions and decrease of movement intensity (Fig. 5).

12. Growth of connectivity. In comparison with the former spatial systems, modern systems show the growth of connectivity. It is shown in the density of connections, in the growth of their intensity as well as in the growth of the number and intensity of direct relations between places (Fig. 6). The growth of direct relations, excluding intermediate stages, is being favoured by the development of automobile transport. The growth of connectivity (Fig. 7) caused disturbances in former hierarchic structure and made it less readable. It gave rise to the wording of statements that hierarchic rules lost in meaning in spatial relations. These, however, are misstatements.

13. Limits of complexity (simplification of organization). The growth of connectivity is not the only trend taking place in the contemporaneous spatial systems.

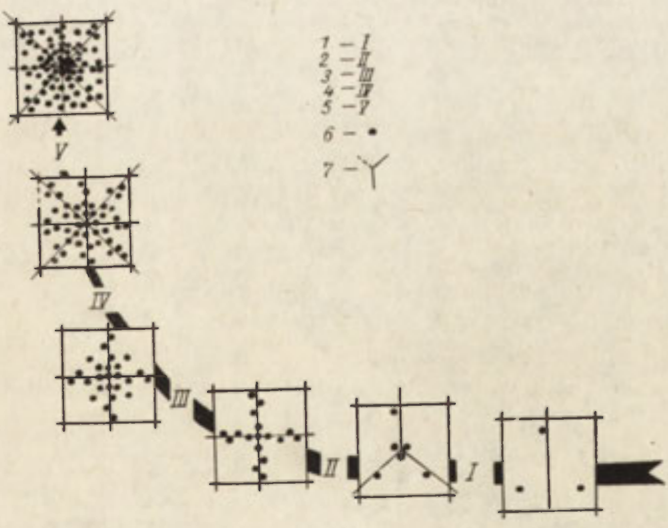


Fig. 9. Development stages of spatial structure of city networks characteristic for different parts of the central economic region of the USSR in the period 1926–1970

1 — first stage, 2 — second stage, 3 — third stage, 4 — fourth stage, 5 — fifth stage, 6 — towns, 7 — main transport routes. After F. M. Listengurt and N. I. Naimark (1973, p. 54)

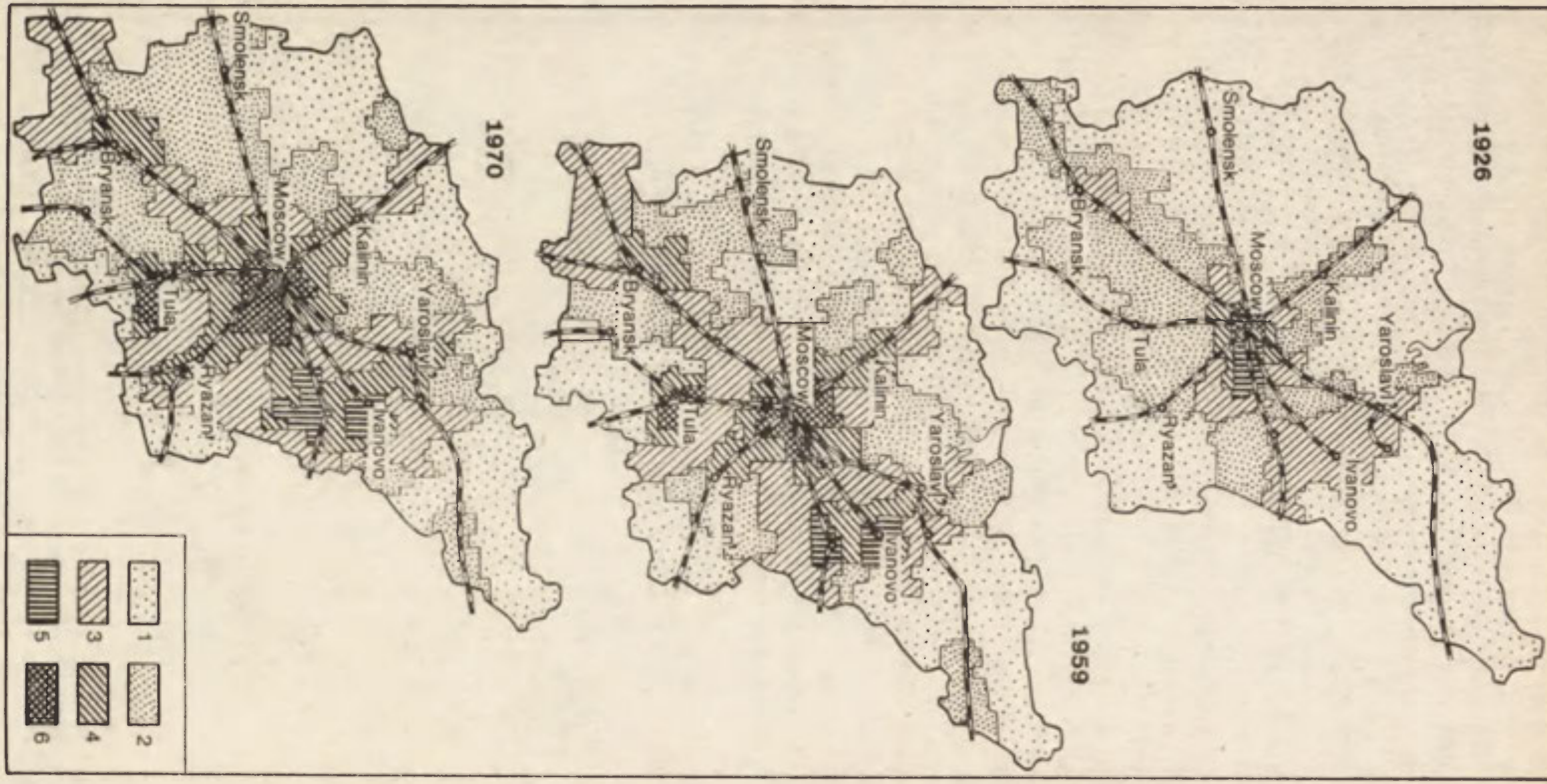


Fig. 10. Development levels of spatial structure of city network of the central economic region in the USSR

1 — very scarce, even network, $U < 0.2$; 2 — scarce, uneven (focal) network, $U = 0.2$–0.5; 3 — group forms of settlements, $U = 0.5$–1.5; 4 — well-developed group forms of settlements on the basis of large cities, $U = 0.5$–1.5; 5 — forming agglomerations, $U = 1.5$–2.5; 6 — formed agglomerations, $U > 2.5$; S — density of city population. For levels 1–3 — S is below 350–400 thousands of inhabitants/10 thousands sq. kms; for levels 4–6 — S is above 350–400 thousands of inhabitants/10 thousands sq. kms. After F. M. Listengurt and N. I. Naimark (1973, p. 56)

Maybe it is the easiest to visualize and that is why it was best recognized and described. Parallel with it, however, another less visible process takes place. It is the simplification process of connection network (Fig. 8). No system can exist or progress properly unless it simplifies, together with the growth of connectivity, its structure. An unrestricted growth of connectivity would decrease the socio-economic efficiency of the system and it would lead to spatial chaos. On the basis of organizational simplification the conclusion can be drawn that what contemporaneously occurs in spatial relations is not a removal of hierarchy but a change of its character. Now it is difficult to distinguish new shapes of hierarchy. We are now in the period of breakdown of the former spatial hierarchy and at the beginning of the reintegration period of spatial connections. However, three reintegration principles can already be distinguished: agglomeration, unitization and incomplete decomposition.

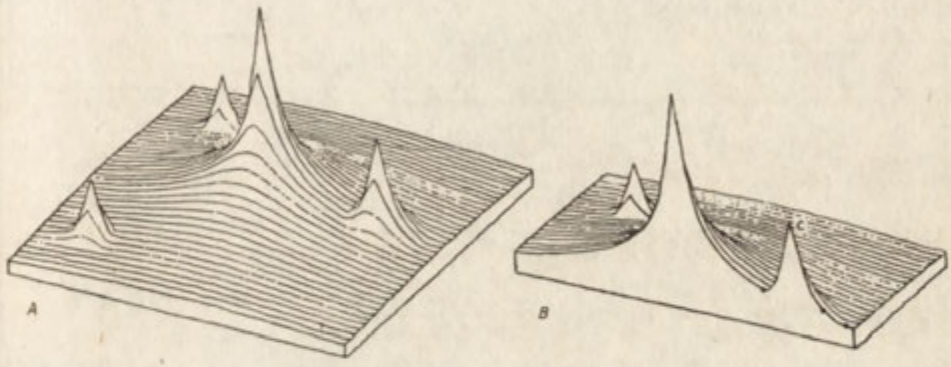


Fig. 11. A — generalized population density surface of a metropolitan area, B — cross-section of metropolitan density surface

This model is a computer-calculated and computer-drawn density surface of a hypothetical metropolitan area. The shape shown could be thought of as the rent surface or density of interaction in a metropolitan area with a central city and three satellite towns. R. S. Yuill after J. F. Kolars and J. D. Nystuen (1974, p. 62)

14. Agglomeration. The scale and importance of urban-industrial agglomeration grows in the modern spatial systems (Figs. 9–11). They are a new form of settlement[4] which cannot be defined within the urban-rural or urban-suburb divisions. The development of motor transport and improvement of urban transport contributed to the birth of agglomerations. Modern transport aroused the urban-suburb migration wave and created conditions for the so-called spatial explosion of cities (Fig. 12). Agglomerations bring a lot of advantages to people, enterprises and social institutions. The so-called large scale economies, accessibility to goods and services, opportunities of establishing contacts belong to them. These advantages attract people, enterprises and institutions. Sometimes the process of agglomeration growth exceeds the optimum limits creating a series of negative phenomena. Everyday journeys to work define the spatial scope of agglomerations. The differentiated structure of land use is developed within these limits. In this structure, besides the land used by industry, trade, services, institutions, transport and housing, more and more land

[4] B. S. Khorev (1975) and O. A. Konstantinov (1976).

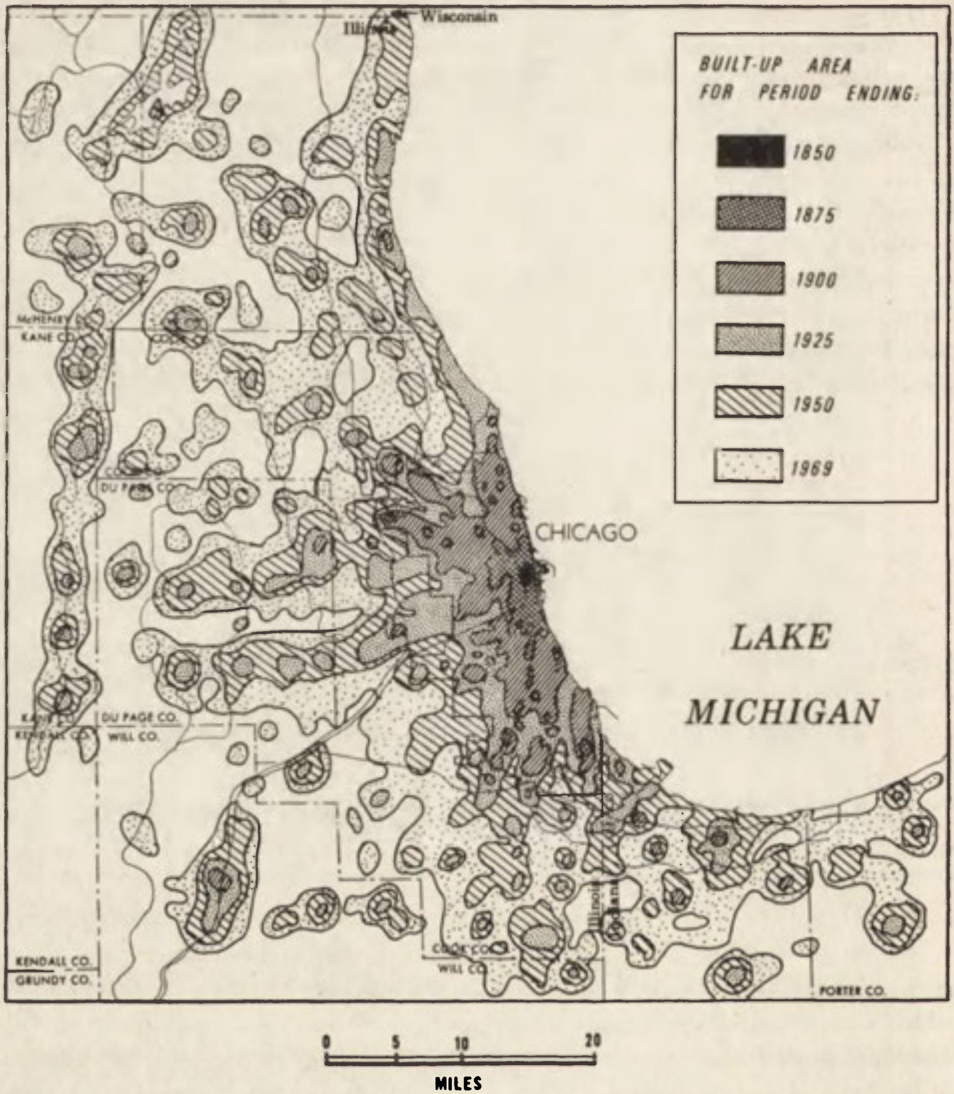


Fig. 12. Chicago area growth patterns. After B. J. L. Berry and F. E. Horton (1970, p. 445)

is being occupied by recreation facilities such as parks, entertainment facilities, playing fields, swimming pools, motels, bungalows.

15. Unitization. The process of unitization[5] takes place in systems of a high level of complexity. It depends on separation, from the whole system, of a subsystem of particularly strong relations. It manifests itself in traffic concentrations on big trunk-lines, expressways, rivers, canals, ocean and airlines. Big urban-industrial agglomerations are the main junctions of these traffic arteries. The conception of node-tract system of development of Poland is the concretization of the rule of uni-

[5] Cf. H. H. Pattee (1973).

Fig. 13. Poland 2000. Node-tract system. Introductory conception

1 — agglomeration range in 1966, 2 — agglomeration range in 2000, 3 — new agglomerations, 4 — city centres of 50–100 thousands of inhabitants, 5 — city centres above 100 thousands of inhabitants, 6 — sets of settlements units, 7 — industrial centres of high dynamics, 8 — centres of new resource districts, 9 — most important transportation lines, 10 — important transportation lines, 11 — directions of spatial integration between agglomerations, 12 — main frontier crossings. S. Leszczycki, P. Eberhardt, S. Heřman after B. Malisz (1974, p. 102)

tization. Beside the conception of moderate policentric concentration it constitutes the fundamental structural element of the perspective plan of spatial development of Poland (Fig. 13).

16. Incomplete decomposition. In the analysis of hierarchic systems and construction of their organizational models it is assumed that distinct hierarchic levels exist among which relations occur. Such assumption means that first we perform a full decomposition and next the isolated levels in the mental conception of systems are joined. This is an acceptable and most frequently a useful approach. However, it does not fit too well in the problem situation considered by us. The modern spatial systems show a property which we call incomplete decomposition. This property shows the integrity of the system, inseparability of the particular elements, even

those of which relations with the rest of the system are weak, as well as the existence of direct relations between elements of the lower level and elements of higher levels, the upper included, excluding the intermediate level.

The idea of incomplete decomposition is in agreement with many ideas of spatial sciences. Among others spatial accessibility, attractiveness of service centres, the overlap of service areas of these centres, the partition of demand, created by the common service area, among different service centres belong here.

What are the effects of the property of incomplete decomposition on the hierarchic structure of modern spatial systems? Following the effects can be made easier by comparing the hierarchic structure of modern spatial systems with the regular hierarchic structure, e.g. Christaller's structure. There are at least three differences: a) less distinct delimitation of the hierarchic levels, b) the new role of the intermediate level, c) reintegration of the basic level not only in relation with the immediately higher level but also, to some extent, in the context with a still higher level.

The idea of formation of a lower level in context with higher levels is worth noticing. Such a context does not occur in the Christaller's model. The particular levels are formed independently from one another and are only joined together in the next step of model formation. Argumentation, in some ways, is even developed in the opposite direction than here, as Christaller first considers central goods and central places of the lower level, and then passes to higher levels. The structure of central places of higher level, to some extent, is formed by central places of lower level, and not vice versa.

The idea of the context with higher levels is more convincing because what can be effectively produced at the lower level and sold outside the local market can be better defined at the higher levels which are the consumers of goods of the lower level.

In the modern spatial systems the disintegration process of the previous local and regional levels as well as the reintegration of these levels in a changed shape take place. The structure of production and services becomes different, the market ranges change and different are the relations between both levels and between them and the national level. Specialization and growing scale of production and services, development and improvement of transport, increase of market range, easier accessibility to centres of higher levels, increasing spatial mobility of population are the reasons of this process.

Territorial connections between production and consumption diminish or disappear at the local level. Connections of this type still exist but to a limited degree within the domain of trade and services for agriculture and rural population. The diminishing range of local connections can lead to a critical threshold of organizational complexity. Organizational complexity of the system has two extremes: the upper one, after the crossing of which the system is in the state of a destructive chaos, and the lower one, below which the organization of the local level cannot work effectively. If the range of local connections shrinks below the threshold of complexity, the local spatial organization undergoes disintegration and its hitherto components establish contacts with the upper organizational level, i.e., with the regional level. Expansion of this process can cause liquidation of the local level and reduction of the number of organizational levels of the spatial system. Then the level of organizational crystallization shifts up to the regional stage. I presume that this consideration may be included in the set of arguments justifying the reform of territorial administration in Poland in 1975 and the liquidation of the *poviat* level. In the transitional period a partial disintegration is to be considered in which the local organization still exists but its components already enter into wider and wider contacts with the regional level.

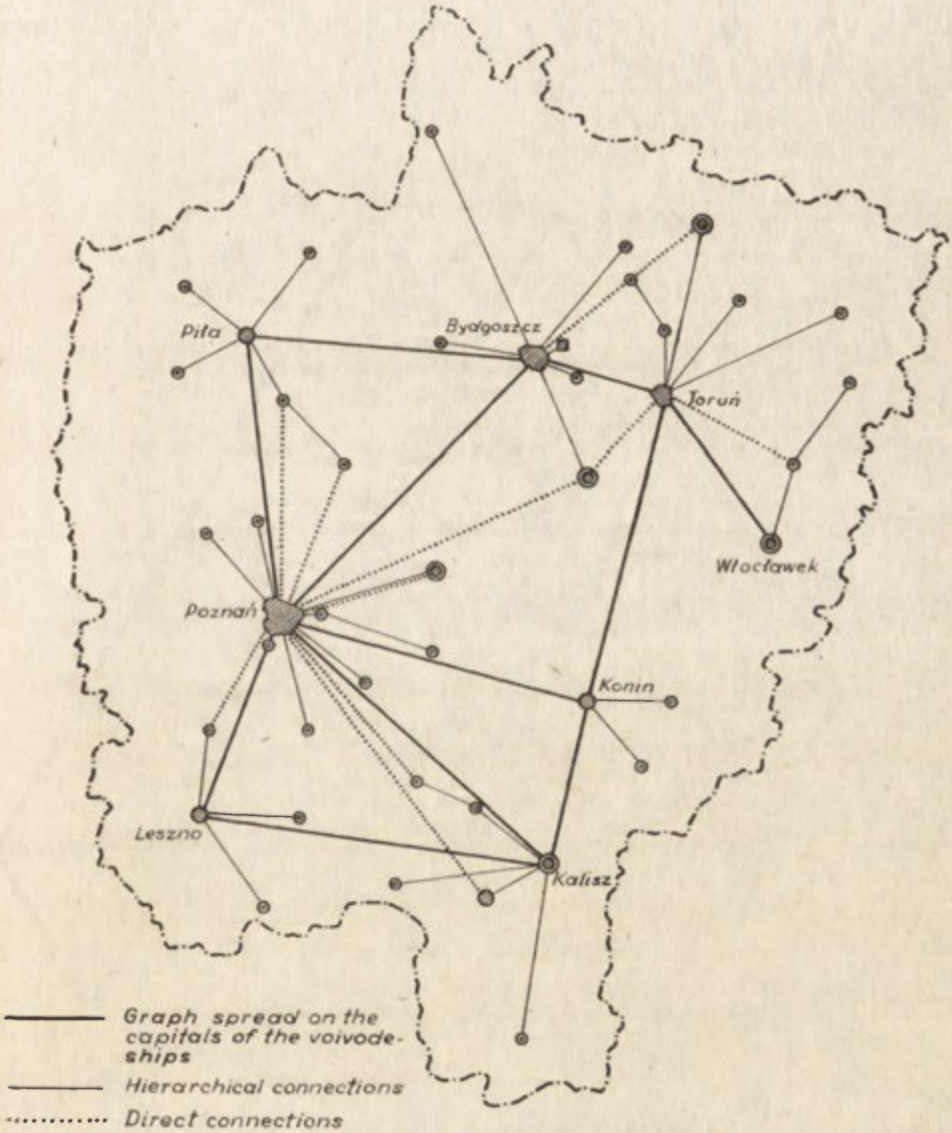


Fig. 14. Incomplete decomposition of a spatial system. Extended accessibility in relation to a regular hierarchical organization

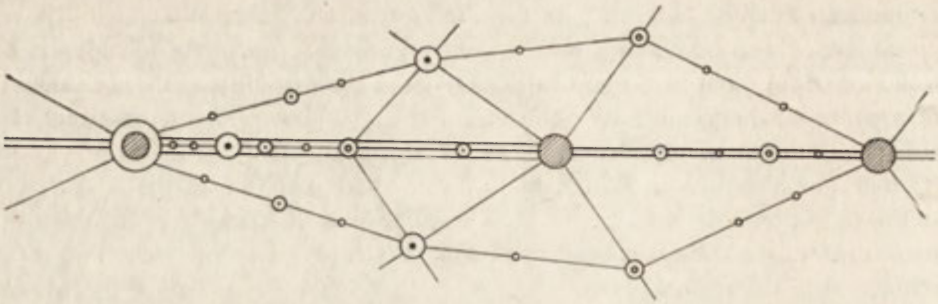


Fig. 15. Anisotropy of socio-economic space. Deviations of settlement and transportation system

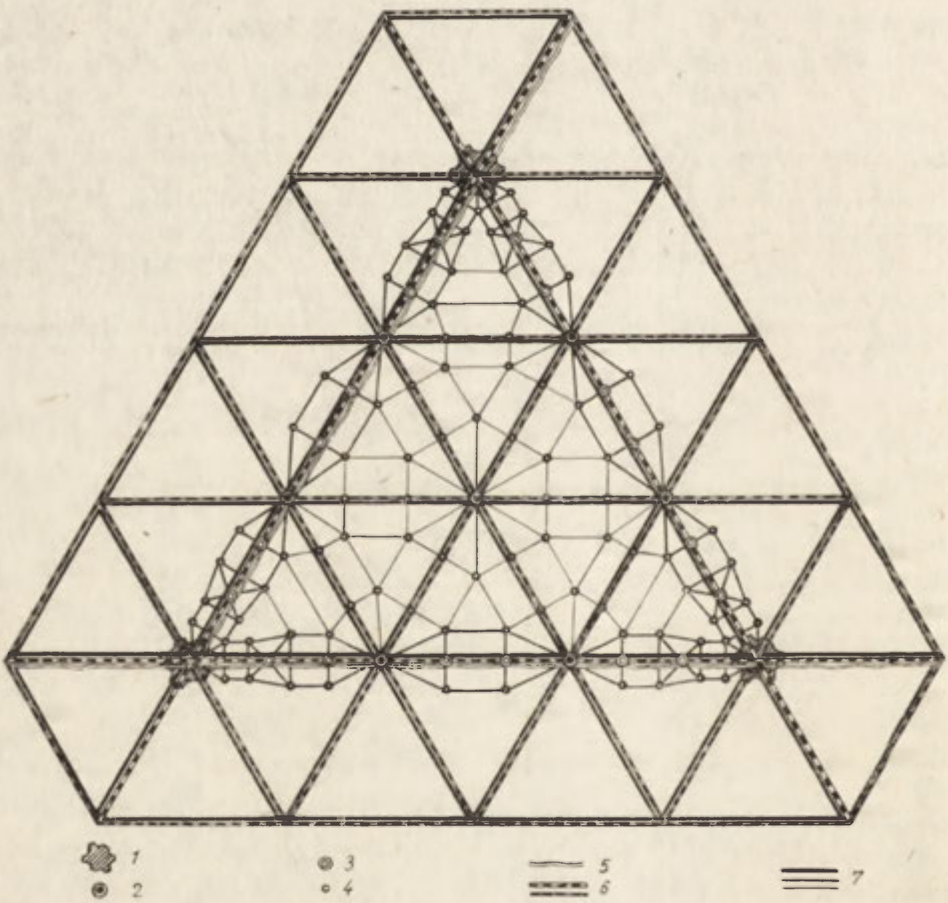


Fig. 16. Anisotropic model of transportation network

1 — nodes of I class, 2 — nodes of II class, 3 — nodes of III class, 4 — nodes of IV class, 5 — rivers, 6 — railroads, 7 — roads

A change of territorial connections also occurs at the regional level but the changed structure does not undergo a general shrinkage, on the contrary, it expands. Similarly as at the local level, territorial connections between production and consumption decrease, while connections in other fields and forms such as journeys to work in regional centre, shopping in regional business centre, utilization of social infrastructure and services of the regional centre, shaping the regional environment, develop. Some regional centres, though having their administrative status of a regional city, may be too weak to fulfill all these functions within their own capacity satisfactorily. In these cases a part of these functions are taken over by multiregional centres.

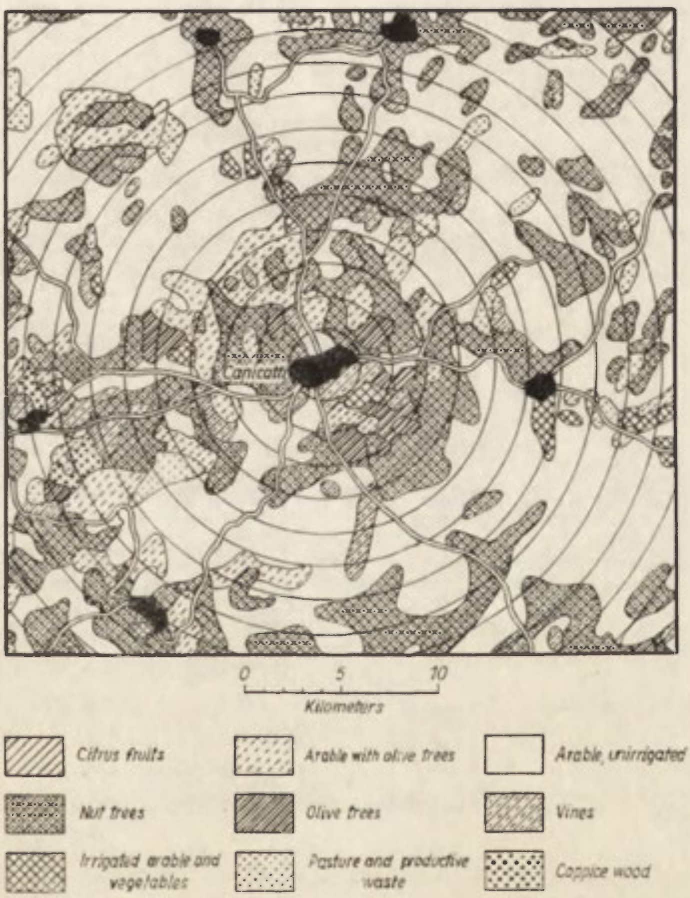


Fig. 17. Land use pattern around Canicatti, Sicily. After J. F. Kolars and J. D. Nystuen (1974, p. 205)

A further development of the scale of spatial connections takes place at the national level. Agglomeration and unitization processes lead to a relative growth of importance of the regional and multiregional centres as well as connections among those centres and among them and the country's central city. Yet, more and more often direct connections of the local and country level occur, by-passing intermediate

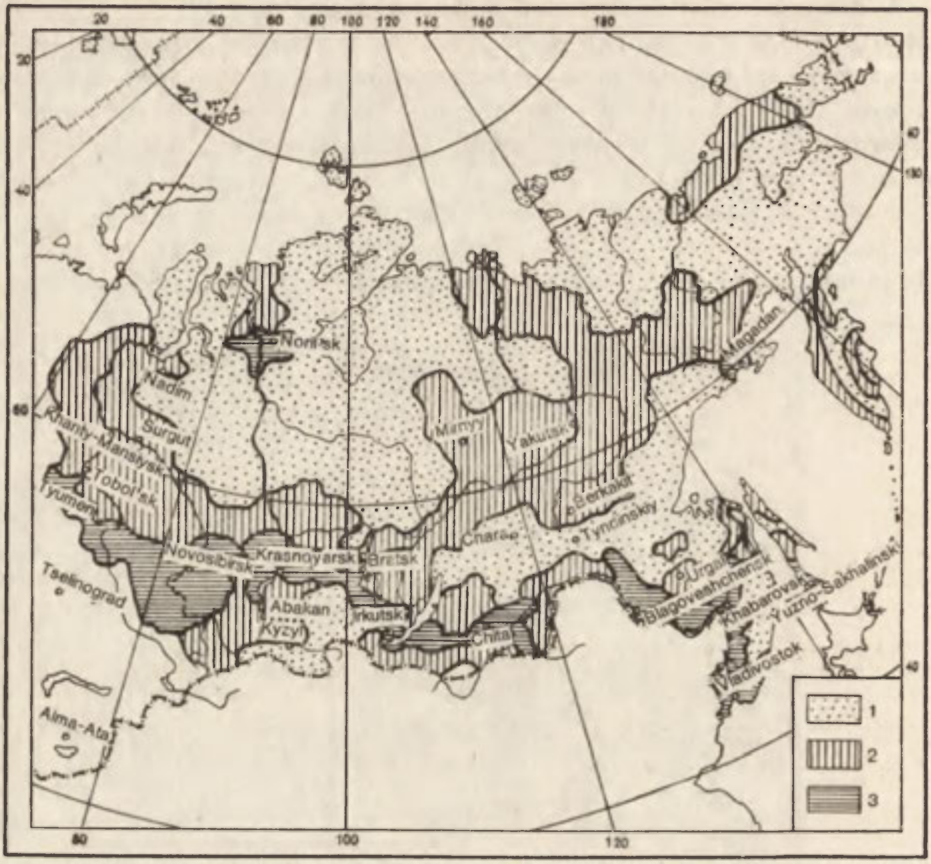


Fig. 18. Development process of Siberia and the Far East of the USSR
Zones of development intensity: 1 — scarce, insular development, 2 — partial, selective. 3 — intensive, continuous. After M. I. Pomus (1975, p. 55)

levels that, however, does not decrease the importance of the intermediate level, the role of which is shaped under the influence of the dominating and constant trend of the system to simplify its organizational structure (Fig. 14).

Now we can add two additional rules to the above-mentioned ones.

17. Reintegration of the given level of spatial organization in the context with higher levels.

18. Indispensability of the intermediate level, i.e. regional and multiregional level.

19. Anisotropic deviations. Typical deviations of transport and settlement system from the regular system are connected with the unitization processes. We call those deviations anisotropy of socio-economic space.[6] They depend on the elongation and relatively more intensive development of transport and settlement system along the main transport arteries (Figs. 15 and 16).

[6] Cf. R. Domanski (1963).

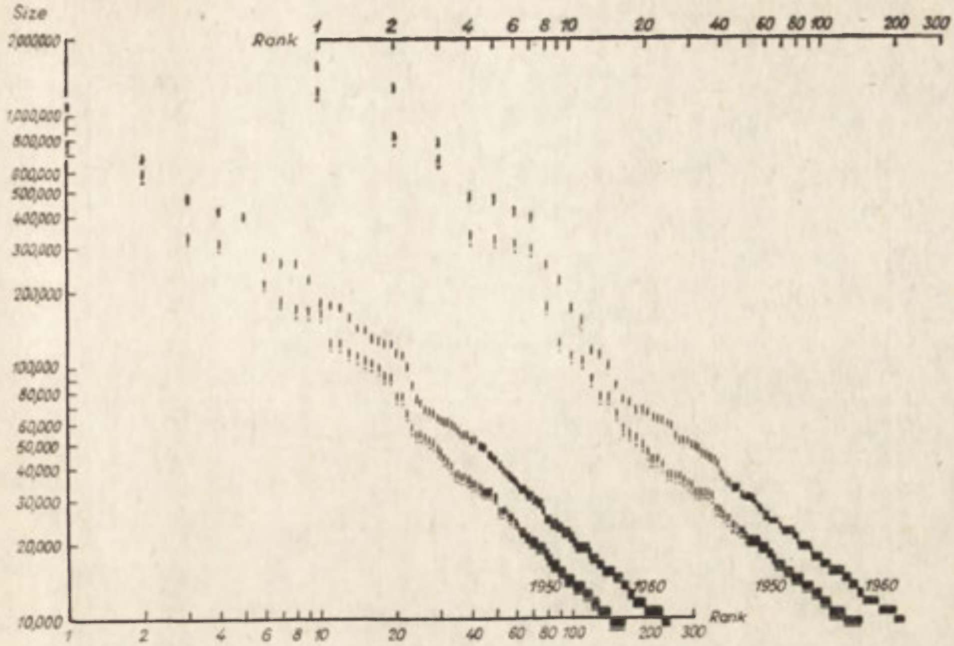


Fig. 19. Classification of Polish towns by rank and size, 1950 and 1960. Curves on the left represent data ordered by administrative units, those on the right data summed together for each metropolitan area or conurbation disregarding their administrative division. After K. Dziewoński (1964, p. 43)

20. Equifinalism. Spatial systems, with different initial conditions and different ways of development, may achieve in the process of development the same final state. For example, the same level of affluence may be achieved by two regions, one of which has an underdeveloped agrarian economy and is being developed due to construction of a new industrial plant while the second region has valuable environmental qualities and develops tourism and health resorts. This property of systems is called equifinalism. It allows also for achieving of the assumed objectives by the same system, despite disturbances in the course of development and despite the change in means of action.

The above-mentioned tendencies in arrangement of movements find their spatial expression in changes of the two synthetic features of the systems: spatial gradients of land use (Figs. 17 and 18) and spatial hierarchy of regions (Fig. 19). Hence, changes in a city's daily cycle, caused by construction of a new trade-service centre on the edge of the city, find their expression in changes of land use depending on the growth of land taken up by trade and services, roads and streets, car-parks. Probably, the share of arable land will decrease and the share and location of residential areas will be changed. The inflow of development factors into attraction niches, which will result in the creation of a new growth pole, will raise the position of a given region in the regional hierarchy, etc.

Therefore, the dynamic organization of space may be regarded as a reaction of the spatial parameters of socio-economic system to impulses induced by decisions

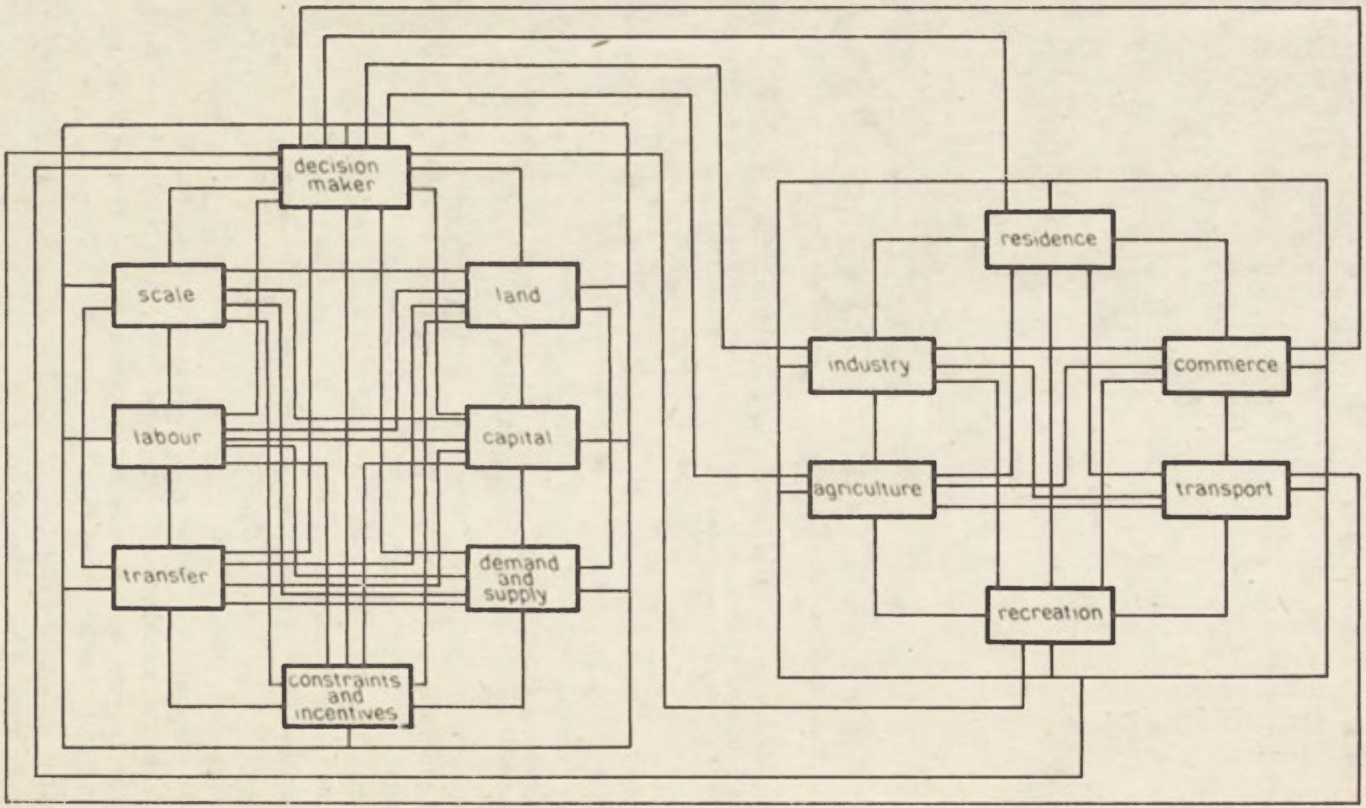


Fig. 20. The system of landscape organization. After P. Toyne (1974, p. 259)

made at different stages of management, starting from personal decisions of individuals. A complicated system of relations and spreading of impulses occures between an organization and decisions (Fig. 20).

Spatial movements are shaped by a tendency to minimalize their costs (expenditures, time, efforts). The same tendency shapes ways of arrangement of movements of different kinds and different spatial scale. The principle of economic effectiveness by this tendency in spatial organization of processes of socio-economic development works itself a way. Positive principles of dynamic organization of space, expressing the general regularities of movements are therefore indications for effective activity in space.

REFERENCES

Bandman, M. K. (ed.), 1976, *Modelirovaniye formirovaniya territorialno-proizvodstvennykh kompleksov* (Modelling of Territorial-production Complexes Formation), Novosibirsk, Izdat. Nauka.

Berry, B. J. L., and F. E. Horton, 1970, *Geographic Perspectives on Urban Systems. With Integrated Readings*, Englewood Cliffs, N. J., Prentice-Hall, Inc.

Khorev, B. S., 1975, *Problemy gorodov. Urbanizatsiya i yedinaya sistema rasseleniya v SSSR* (Problems of Cities. Urbanization and a Unified System of Settlement), 2nd ed., Moskva, Izdat. Mysl.

Domański, R., 1963, *Zespoły sieci komunikacyjnych* (Sum.: Complexes of Transport Networks), Prace Geograficzne, Instytut Geografii PAN, 41, Warszawa, PWN.

Domański, R., 1977, Dynamika systemów przestrzennych. Model procesów przestrzennych (Sum.: Dynamics of Spatial Systems. A Model of Spatial Processes), *Przegląd Geograficzny*, 49, 3, 401–436.

Dziewoński, K., 1964, Urbanization in Contemporary Poland, *Geographia Polonica*, 3, 37–56.

Hoover, E. M., 1971, *An Introduction to Regional Economics*, New York, Alfred A. Knopf.

Isard, W., 1956, *Location and Space-economy*, New York, J. Wiley and Sons, Inc.

Kolars J. F., J. D. Nystuen, 1974, *Geography. The Study of Location, Culture and Environment*, New York, McGraw-Hill Book Company.

Kolosovskii, N. N. 1947, Proizvodstvenno-territorialnoe sochetanye v sovetskoy geografii (Territorial-production Complex in Soviet Geography), in *Geografiya khoziaystva SSSR 1917–1947*, Moskva, Geografgiz.

Kolosovskii, N. N., 1971, *Problemy territorialnoy organizatsii proizvoditelnykh sil Sibirii* (Problems of Territorial Organization of Production Forces in Sibiria), Novosibirsk.

Konstantinov, O. A., 1976, Tipy urbanizirovannosti v SSSR (The Types of Urbanization in the USSR), *Izvestia Akademii Nauk SSSR, Ser. Geograficheskaya*, 5, 72–83.

Listengurt, F. M., and N. I., Naimark, 1973, Regionalnye osobennosti territorialnoy struktury promyshlennogo razvitiya i seti gorodskikh poselenii (Regional Specific Features of the Spatial-structural and Industrial Development of the Urban Settlement System), *Izvestia Akademii Nauk SSSR, Ser. Geograficheskaya*, 3, 51–62.

Malisz, B., 1974, *Problematyka przestrzennego zagospodarowania kraju* (Problems of Spatial Development of the Country), Warszawa, PWN.

Misra, R. P., 1971, The Diffusion of Information in the Context of Development Pląnning, in T. Hagerstrand and A. Kukliński (eds.), *Information Systems for Regional Development. A Seminar*, Lund Studies in Geography, Lund.

Pattee, H. H. (ed.), 1973, *Hierarchy Theory. The Chalenge of Complex Systems*, New York, G. Braziller.

Pomus, M. I., 1975, Konstruktivnoye napravleniye v ekonomicheskoy geografii i novye problemy Sibirii (A Constructive Trend in Economic Geography and Modern Problems of Sibiria), *Izvestia Akademii Nauk SSSR, Ser. Geograficheskaya*, 4, 51–63.

Toyne, P., 1974, *Organization, Location and Behaviour. Decision-making in Economic Geography*, London, The Macmillan Press, Ltd.

GEOGRAPHIA POLONICA 42, 1979

STABILITY OF THE DEMAND FUNCTION WITH SPATIAL DIFFUSION

Oscar Fisch

Department of City and Regional Planning, The Ohio State University, Columbus, Ohio, USA

I. INTRODUCTION

Since Hagerstrand's classical work*, the diffusion of innovations as a spatial process has made significant inroads into geographical analysis. Fed by cultural anthropology and by information and graph theory, the geographical conceptualization of innovation adoption stresses the spatial configuration of the social communication network, which is composed of individuals playing the double role of senders and receivers with their spatial location and their degree of connectivity. With the addition of the concept of noise in the communication channels, a spatial pattern of information flows emerges as the backbone of innovation diffusion. This systems engineering approach altogether fails to bring into focus human behavior. It reduces the learning process to a minimum or threshold of message 'hits', after which the adoption of an innovation will take place. The problems rests on that information theory — in Shannon's best tradition — suffices for the examination of precision questions in the transmission of information in a channel. But the conceptual core of the innovation adoption problem is not only the transmission of information, but also the processing of that information. But the theory of information processing needs to, at least, include the concepts of the meaning and value of information. Man's brain receives a very large amout of diverse stimuli, and if he were to react to all these stimuli, the response would be an erratic sequence of chaotic actions. Instead, man makes a selection of the information hitting his sensors according tq its meaning, i.e., its correspondence with the biological, physical and social environment he interacts with, and according to its value, i.e., the utility of the information for his own actions in these environments.

As one segment of these actions are economic in nature (a subset of which is related to consumer behavior), it is rather surprising to find that there is not a unified approach between economic theory of consumer behavior and the geographical approach to spatial diffusion of information. In spite of the growing concensus among economists that if the analysis of consumer behavior eschews interpersonal comparisons affecting the shape of individual demand functions then the scope of

* Hägerstrand, T., 1953, *Innovationsförloppet ur korologisk synpunkt*, Lund.

economic theory is indeed being limited to a great extent, economic analysis is still not far away from the conceptual ambiguity of 'consumer sovereignty'.

It is the main purpose of this paper to depart from the notion that the consumer is a 'Robinson Crusoe', and analyze consumer behavior as perturbed by his social and spatial environment, i.e., by the presence of a network of social communication and spatial flow of information.

2. MARKET SIGNALLING AND INDIVIDUAL DEMAND FUNCTION

We submit the notion of a continuum of consumers on a geographical two-dimensional space Z. We submit the second notion that the individual's demand function is being driven by two dialectically interconnected behavioral components: the autonomous part and the heteronomous (environmentally perturbed) part

The mapping, a continuum of consumers "i", states that there is a relation g such that each domain Z is paired with exactly one range element i. The i which is thus uniquely determined by g and z raises the mapping

$$i = g(Z) = g(z_1, z_2), \tag{2.1}$$

where z_1 and z_2 are the cartesian coordinates of the space Z. The domain of z_i is the closed metric interval [0, 1]. Thus, the demand of individual i is transformed to a density demand function by geographical location

$$q(i) = q(z_1, z_2), \tag{2.2}$$

where a bundle of goods q is a point in the non-negative orthant Ω of R^n. The dimensionality n of the space R represents the number of different goods being demanded. Similarly, the individual's income is mapped as a density function of our geographical space

$$y(i) = y(z_1, z_2). \tag{2.3}$$

Abstracting from the intricacies of the interconnection between the two components, we will assume an additive relationship in the demand function. Given an n-dimensional vector of price p, the demand function for every individual in Z, is given by

$$q(z_1, z_2) = \Psi[y(z_1, z_2), p]+\Delta[y(z_1, z_2), p], \tag{2.4}$$

where Ψ and Δ are vector functions, the first being the autonomous part, the second being the heteronomous part. In relation to (2.4), the following vector products holds:

$$\langle\Psi, p\rangle = y, \tag{2.5}$$

$$\langle\Delta, p\rangle = 0. \tag{2.6}$$

The demand for goods should vary continuously with price in the sense that for a given vector p, there is a unique response of Ψ and Δ. Fixing p arbitrarily, then (2.4) becomes for $\Psi[y(z_1, z_2)] = f(z_1, z_2)$

$$q(z_1, z_2) = f(z_1, z_2)+\Delta(z_1, z_2). \tag{2.7}$$

We assume that the function f is Riemann integrable. The integral of a vector function is to be taken, as usual, as the vector of integrals of the components.

The vector function $\Delta(z_1, z_2)$ reflects the impact on individual i located at (z_1, z_2) of a set of individuals $j \neq i$ at different locations $(x_1, x_2) \in Z$ with a given demand function $q(x_1, x_2)$.

We define the impact of j's demand on i's demand of good l as $\Delta_l(z_1, z_2; x_1, x_2)$; then it follows that

$$\Delta_l(z_1, z_2) = \int_0^1 \int_0^1 \Delta_l(z_1, z_2; x_1, x_2)\, dx_2\, dx_1 . \tag{2.8}$$

In specifying the function $\Delta_i(z_1, z_2; x_1, x_2)$, we find two kinds of problems. First, there is the information transmission problem. Individual i may have, without cost, from complete to nill information about any component of $q(j)$. This information flow can be a function of the spatial or social distance between i and j, or a function of the spatial or social location of i, or both. The elucidation of this social structural problem is not our direct concern here, but our approach should support broadly its inclusion in our analysis. We define j's demand of good l, as perceived by individual i, as $q_l(i,j)$ with $0 < q_l(i,j) < q_l(j)$.

Second, there is the information processing problem. The result of this processing can have a positive impact on i's demand — social emulation — or a negative one — social differentiation. Furthermore, j's whole demand vector may have an impact on i's demand of good l, i.e., if there is some degree of complementarity or substitution between any pair of goods. In general, the perturbation function will be

$$\Delta_l(z_1, z_2; x_1, x_2) = \int_{x_1-\delta}^{x_1+\delta} \int_{x_2-\varepsilon}^{x_2+\varepsilon} \Delta_l[(z_1, z_2; q_1(z_1, z_2; x_1, x_2) \ldots q_l(z_1, z_2; x_1, x_2) \ldots q_n(z_1, z_2; x_1, x_2)]\, dx_2\, dx_1 . \tag{2.9}$$

The processing content of the function (2.9) can be summarized as follows:

i's behavior in relation to j's demand	$\dfrac{\delta\Delta_l(z_1, z_2; x_1, x_2)}{\delta q_l(z_1, z_2; x_1, x_2)}$	$\dfrac{\delta\Delta_l(z_1, z_2; x_1, x_2)}{\delta q_m(z_1, z_2; x_1, x_2)}$
Social emulation	> 0	< 0 Gross Substitution
		> 0 Complementarity
Social differentiation	< 0	> Gross Substitution
		< 0 Complementarity

We can see in the above table that the adoption of an innovation is not the only possible result. Not only can we emulate through the demand of some good (positive impact), but we can also choose to differentiate ourselves from the 'crowd' through the demand of another good (negative impact). Or, in processing the information of simultaneous demand for one single good, we could emulate one segment of the society and at the same time try to differentiate from another segment of the society with an end impact that can be either way in our demand of that good.

3. THE LINEAR CASE

Let us assume that the information received by i is linear on $q(j)$, i.e.,

$$q_l(z_1, z_2; x_1, x_2) = I_l(z_1, z_2; x_1, x_2)\, q_l(x_1, x_2), \tag{3.1}$$

with $I_l \in [0, 1]$. $I(z_1, z_2; x_1, x_2)$ describes the spatial configuration of the social communication network and the level of noise in the channels. Let us approximate linearly the response function (2.9), i.e.,

$$\Delta_l(z_1, z_2; x_1, x_2) = \lambda \int\limits_{x_1-\delta}^{x_1+\delta} \int\limits_{x_2-\varepsilon}^{x_2+\varepsilon} \sum_{m-1}^{n} E_{lm}(z_1, z_2, ; x_1, x_2) q_m(z_1, z_2; x_1, x_2) dx_1\, dx_2, \tag{3.2}$$

where the signs of the information processing function E corresponds to the behavioral characteristics of the precceding table, the parameter λ being sufficiently small so as to keep the vector $q(i)$ within Ω. Then from (2.7), (2.8), (3.1) and (3.2), we have that the function of i's total perturbation by its social environment is given by

$$\Delta(z_1, z_2) = \lambda \int\limits_0^1 \int\limits_0^1 N(z_1, z_2; x_1, x_2)\, q(x_1, x_2) dx_2 dx_1, \tag{3.3}$$

where N is a matrix function with components

$$N_{lm}(z_1, z_2; x_1, x_2) = E_{lm}(z_1, z_2; x_1, x_2) I_m(z_1, z_2; x_1, x_2), \tag{3.4}$$

with I as a real, non-negative and bounded vector function, ans E as a real and bounded matrix function. It should be noted that whereas the function I is non-negative in the whole geographical space, the function E has no sign restriction and may change sign more than once in the same space. Figure 1 exemplifies one hypothetical case. $N(z_1, z_2; x_1, x_2)$ portrays, in a very differentiated way, the role of individual j at location (x_1, x_2) as an emitter of consumption signals, individual i's role at location (z_1, z_2) as a receptor of these signals the level of noise in the communication channel between both locations, and individual i's processing and response to the signals received.

In relation to $N(z, x)$, we assume that the Riemann integrals

$$\int\limits_0^1 \int\limits_0^1 N(z_1, z_2; x_1, x_2) dz_2 dz_1 \quad \text{and} \quad \int\limits_0^1 \int\limits_0^1 N(z_1, z_2; x_1, x_2) dx_2 dx_1$$

exist: the first for every fixed location (x_1, x_2), the second for every fixed location (z_1, z_2). As $N(z, x)$ is bounded, there exists a number k such that

$$|N(z_1, z_2; x_1, x_2)| \leqslant k.$$

From (2.7) and (2.9), it follows that the unknown spatial vector demand function, for all the locations in our geography, is given by

$$q(z_1, z_2) = f(z_1, z_2) + \lambda \int\limits_0^1 \int\limits_0^1 N(z_1, z_2; x_1, x_2)\, q(x_1, x_2)\, dx_2 dx_1. \tag{3.5}$$

If the following condition holds,

$$n^2|\lambda|k < 1 \tag{3.6}$$

we can state the following.

Theorem I: There exists an integrable solution of the spatial demand function $q(z_1, z_2)$ for all locations in he geographical map and it is unique and bounded. The heteronomous component of the solution is a spatially weighted integral of the autonomous component of the demand function for the whole space.

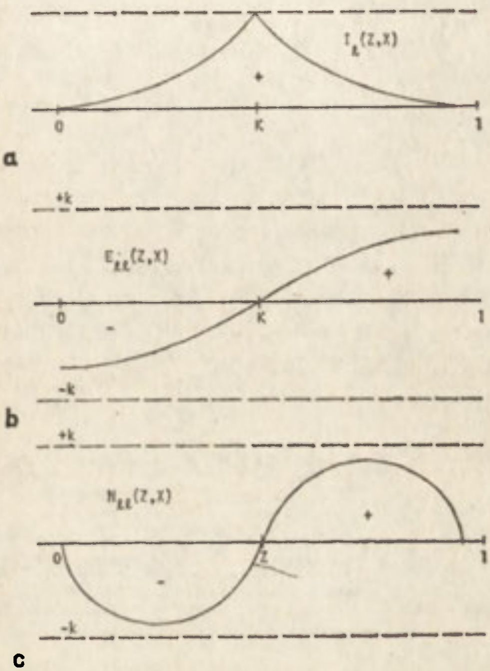


Fig. 1. a — information function, b — emulation — differentiation function, c — spatial net emulation differentiation function

Proof: The system of n linear integral equations implicitly given in (3.5) can be reduced to one linear integral equation.

Equation (3.5) is really a linear system, i.e.,

$$q_l(z_1, z_2) = f_l(z_1, z_2) + \lambda \int_0^1 \int_0^1 \sum_{m=1}^{n} N_{lm}(z_1, z_2; x_1, x_2) q_m(x_1, x_2) dx_2 dx_1,$$

$$l = 1, 2, \ldots, n,$$

where $q_1(z_1, z_2), \ldots, q_n(z_1, z_2)$ are the unknown functions in the region Z.

This system can be reduced to one equation by considering regions $Z_2, Z_3, \ldots, Z_n$ obtained by parallel displacement of the region $Z_1 = Z$ with non-overlapping points. Let $F(z_1, z_2)$ denote a function in the domain $Z_1 UZ_2U \ldots UZ_n$, defined by the relations

$$F(z_1^l, z_2^l) = f_l(z_1^1, z_2^1),$$

where (z_1^l, z_2^l) is a point of z_l corresponding to a point (z_1^1, z_2^1) of $Z_1 = Z$.

Let us define $N(z_1, z_2; x_1, x_2)$ as a function of two variables in $Z_1 U Z_2 U \ldots U Z_n$ with the relations

$$N(z_1^l, z_2^l; x_1^m, x_2^m) = N_{lm}(z_1^1, z_2^1; x_1^1, x_2^1), \; l, m = 1, 2, \ldots, n,$$

where $(z_1^l, z_2^l) \in Z_l$ and $(x_1^m, x_2^m) \in Z_m$. It follows that if the functions $q_1(z_1, z_2), \ldots, q_n(z_1, z_2)$ satisfy the linear system (3.5), then the function $Q(z_1, z_2)$ defined in $z_1 \cup z_2 \cup \ldots \cup z_n$ by the relations

$$Q(z_1^l, z_2^l) = q_l(z_1^1, z_2^1), \; (z_1^l, z_2^l) \in Z_l, \; l = 1, 2, \ldots, n$$

satisfies the equation

$$Q(z_1, z_2) = F(z_1, z_2) + \lambda \int_0^n \int_0^n N(z_1, z_2; x_1, x_2)\, Q(x_1, x_2) dx_2\, dx_1 \tag{3.5'}$$

and conversely, if $Q(z_1, z_2)$ is a solution of (3.5′), then the value of this function in $Z_1, Z_2, \ldots, Z_n$ defines a system of n functions in Z, which constitute a solution to the system (3.5). Thus, the one unknown function $Q(z_1, z_2)$ is equivalent to the n unknown functions $q_l(z_1, z_2)$, $l = 1, 2, \ldots, n$. The proof that follows corresponds to one integral equation.

The first iteration of (3.5)

$$q(z_1, z_2) = f(z_1, z_2) + \lambda \int_0^n \int_0^n N(z_1, z_2; x_1, x_2) \Big[f(x_1, x_2)$$

$$+ \lambda \int_0^n \int_0^n N(x_1, x_2; s_1, s_2)\, q(s_1, s_2) ds_2\, ds_1 \Big].$$

If we define

$$N_2(z_1, z_2; x_1, x_2) = \int_0^n \int_0^n N(z_1, z_2; s_1, s_2)\, N(s_1, s_2; x_1, x_2) ds_2 ds_1,$$

then

$$q(z_1, z_2) = f(z_1, z_2) + \lambda \int_0^n \int_0^n N(z_1, z_2; x_1, x_2) f(x_1, x_2) dx_2 dx_1$$

$$+ \lambda^2 \int_0^n \int_0^n N_2(z_1, z_2 : x_1, x_2) q(x_1, x_2) dx_2\, dx_1.$$

Then, the mth iteration is given by

$$q(z_1, z_2) = f(z_1, z_2) + \sum_{i=1}^{m} \lambda^i \int_0^n \int_0^n N_i(z_1, z_2; x_1, x_2) f(x_1, x_2) dx_2\, dx_1$$

$$+ \lambda^{m+1} \int_0^n \int_0^n N_{m+1}(z_1, z_2; x_1, x_2)\, q(x_1, x_2) dx_2\, dx_1,$$

where

$$N_{m+1}(z_1, z_2; x_1, x_2) = \int_0^n\int_0^n N_m(z_1, z_2; s_1, s_2)\, N(s_1, s_2; x_1, x_2)\, ds_2\, ds_1 .$$

Then, at the limit

$$q(z_1, z_2) = f(z_1, z_2) + \sum_{i=1}^{\infty} \lambda^i \int_0^n\int_0^n N_i(z_1, z_2; x_1, x_2) f(x_1, x_2)\, dx_2\, dx_1, \qquad (3.6)$$

with

$$\left|\lambda^i \int_0^n\int_0^n N_i(z_1, z_2; x_1, x_2) f(x_1, x_2)\, dx_2\, dx_1\right| < (n^2|\lambda|k)^i .$$

Then, the solution of (3.5) can be written as

$$q(z_1, z_2) = f(z_1, z_2) + \lambda \int_0^n\int_0^n R(z_1, z_2; x_1, x_2; \lambda) f(x_1, x_2)\, dx_2\, dx_1, \qquad (3.7)$$

where the spatial weight function R is given by

$$R(z_1, z_2; x_1, x_2; \lambda) = \sum_{i=1}^{\infty} \lambda^{i+1} N_i(z_1, z_2; x_1, x_2). \qquad (3.8)$$

R is the sum of a series that is absolutely and uniformly convergent, provided that the parameter λ satisfies condition (3.6).

4. THE NON-LINEAR CASE: HOUSEHOLD PRODUCTION FUNCTION

If we introduce the notion of the household's production function (Lancaster, 1966), then social emulation — differentiation will occur via the vector of commodities r produced by the household j as they impact i's vector demand of goods q.

Let us assume that the household production function of individual i is given by

$$r(i, p) = r(i, q(i), p), \qquad (4.1)$$

where the commodity vector r is m-dimensional, and the production is a convex function with arguments given by i's technology, the n-dimensional vector of goods q and prices p. Again, we assume that it varies continuously with the prices in the sense that for a given set of prices, there is a unique response r. Fixing arbitrarily p, and locating the household production process in the geographical space, we can analyse (4.1) in the following way

$$r(z_1, z_2) = r\big(z_1, z_2; q(z_1, z_2)\big). \qquad (4.2)$$

We call the commodities produced at (x_1, x_2) and perceived at (z_1, z_2) $r(z_1, z_2; x_1, x_2)$ and it follows that $0 < r(z_1, z_2; x_1, x_2) < r(x_1, x_2)$. Then the perturbation on i's demand of goods will be given by

$$\Delta_l(z_1, z_2; x_1, x_2) = \int_{x_1-\delta}^{x_1+\delta}\int_{x_2-\delta}^{x_2+\varepsilon} \Delta_l(z_1, z_2; r_1(z_1, z_2; x_1, x_2) \ldots$$

$$\ldots r_m(z_1, z_2 : x_1, x_2))\, dx_2\, dx_1 . \qquad (4.3)$$

Let us approximate linearly the function (4.3), i.e.,

$$\Delta_l(z_1, z_2; x_1, x_2)$$
$$= \lambda \int_{x_1-\delta}^{x_1+\delta} \int_{x_2-\varepsilon}^{x_2+\varepsilon} \sum_{v=1}^{m} E_{lv}(z_1, z_2; x_1, x_2) r_v(z_1, z_2; x_1, x_2) dx_2 dx_1 \tag{4.4}$$

where E not only corresponds to behavioral characteristics of the consumer located at (z_1, z_2), but it also maps the impacts on the goods space R^n from the commodities space R^m. Additionally, if the information flow between the two locations is still linear, i.e.,

$$r_l(z_1, z_2; x_1, x_2) = I_l(z_1, z_2; x_1, x_2) r_l(x_1, x_2), \tag{4.5}$$

then the total perturbation at (z_1, z_2) is given by

$$\Delta(z_1, z_2) = \lambda \int_0^1 \int_0^1 N(z_1, z_2; x_1, x_2) r(x_1, x_2) dx_2 dx_1. \tag{4.6}$$

N still portrays almost the same role as in the preceeding section. The difference with the first case is that signals are not related to the input (goods) space, but to the output (commodities) space. From (2.7) and (2.9), (4.2) and (4.6), it follows that the unknown demad at each location is given by

$$q(z_1, z_2) = f(z_1, z_2) + \lambda \int_0^1 \int_0^1 N(z_1, z_2; x_1, x_2) r\big(x_1, x_2; q(x_1, x_2)\big) dx_2 dx_1. \tag{4.7}$$

Given that the household production function is convex, there exists a number L such that

$$||r(z_1, z_2; q) - r(z_1, z_2; q)|| \leqslant L||q-q||, \; L > 0. \tag{4.8}$$

If the following condition holds

$$|L \cdot K \cdot \lambda| < 1,$$
$$\int_0^1 \int_0^1 |N(z_1, z_2; x_2, x_2)| \, dx_2 dx_1 \leqslant K, \tag{4.9}$$
$$r(z_1, z_2; 0) = 0,$$

we can state the following.

Theorem II: There exists a unique and bounded demand function $q(z_1, z_2)$ for all locations on the geographical map, which is a solution of equation (4.7).

Proof: We will use the method of successive approximations, constructing a sequence of $q_r(z_1, z_2)$ functions with $q_0(z_1, z_2) = f(z_1, x_2)$ and

$$q_n(z_1, z_2) = f(z_1, z_2) + \lambda \int_0^1 \int_0^1 N(z_1, z_2; x_1, x_2) r(x_1, x_2; q_{n-1}(x, x_2) dx_2 dx_1,$$

$$|q_n(z_1, z_2)| \leqslant f(z_1, z_2) + \lambda K \sup_{z_i \in [0,1]} r\big(z_1, z_2; q_{n-1}(z_1, z_2)\big),$$

then, as

$$\big|r\big(z_1, z_2; q_{n-1}(z_1, z_2)\big)\big| \leqslant L(|q_{n-1}(z_1, z_2)| + |r(z_1, z_2; 0)|$$

this shows that, indeed, the $q_n(z_1, z_2)$ is a sequence of bounded functions on all the geographical space.

For the $n+1$ approximatiom, we have that

$$\sup_{z_i \in [0,1]} |q_{n+1}(z_1, z_2) - q_n(z_1, z_2)| \leqslant \lambda KL \sup_{z_i \in [0,1]} |q_n(z_1, z_2) - q_{n+1}(z_1, z_2)|.$$

For $n = 1$, we have

$$\sup_{z_i \in [0,1]} |q_1(z_1, z_2) - q_0(z_1, z_2)| \leqslant \lambda KS,$$

where

$$S = \sup_{z_i \in [0,1]} \left|r(z_1, z_2; f(z_1, z_2))\right|.$$

S is finite, because

$$|r(z_1, z_2; f(z_1, z_2))| \leqslant L|f(z_1, z_2)| + r(z_1, z_2; 0).$$

It follows that

$$\sup_{z_i \in [0,1]} |q_{n+1}(z_1, z_2) - q_n(z_1, z_2)| \leqslant SL^n(\lambda K)^{n+1}$$

and at the limit, the series are given by

$$\sum_{n-1}^{\infty} [q_n(z_1, z_2) - q_{n-1}(z_1, z_2)] \leqslant S\lambda K \sum_{n=1}^{\infty} (\lambda KL)^n.$$

But condition (4.9) guarantees the convergence of the last series, implying the uniform convergence of the sequence $\{q_n(z_1, z_2)\}$. If we denote $q(z_1, z_2) = \lim_{n\to\infty} q_n(z_1, z_2)$, then $q(z_1, q_2)$ is the unique solution of equation (4.7) and is continuous and bounded on $(z_1, z_2) \in [0, 1]$.

5. THE NON-LINEAR CASE: THE GENERAL PROBLEM

If the information received by i on j's output of commodities is not linear, then

$$r_l(i, j) = r_l(i, j; q(j)) \tag{5.1}$$

and the perturbation at i's location on demand of good l is

$$\Delta_l(z_1, z_2; x_1, x_2) = \int_{x_1-\delta}^{x_1+\delta} \int_{x_2-\varepsilon}^{x_2+\varepsilon} \Delta_l[z_1, z_2; r_1(z_1, z_2; x_1, x_2; q(x_1, x_2)) \dots r_m(z_1, z_2; x_1, x_2; q(x_1, x_2))] dx_2 dx_1, \tag{5.2}$$

with $\Delta_l(i, j) = 0$ if $r(i, j; q(j)) = 0$ for $i = 1, 2, \dots, n$.

In equation (5.2) the vector function under the integral sign can be expressed as

$$\Delta[z_1, z_2; r(z_1, z_2; x_1, x_2; q(x_1, x_2))] = \lambda N[z_1, z_2; x_1, x_2; q(x_1, x_2)]. \tag{5.3}$$

As the household production function $r(z_1, z_2)$ is convex, it follows that N is also convex, i.e.,

$$||N(z_1, z_2; x_1, x_2; q) - N(z_1, z_2; x_1, x_2; q)|| \leqslant L||q - q||. \tag{5.4}$$

We assume that the function N is bounded, i.e., that for all locations (z_1, z_2)

$$\int_0^1\int_0^1 |N_l(z_1, z_2; x_1, x_2; q(x_1, x_2))|\, dx_2 dx_1 \leqslant K. \tag{5.5}$$

The parameter λ keeps the demand function in Ω. The unknown spatial demand function in all the geographical space, is the solution of the following non-linear equation

$$q(z_1, z_2) = f(z_1, z_2) + \lambda \int_0^1\int_0^1 N(z_1, z_2; x_1, x_2; q(x_1, x_2))\, dx_2 dx_1. \tag{5.6}$$

If the following condition holds

$$|\lambda \cdot K \cdot L \cdot n^2| \leqslant 1 \tag{5.7}$$

we can state

Theorem III: There exists a unique spatial demand function $q(z_1, z_2)$ for all locations (z_1, z_2) in the geography under consideration, which is a solution of the non-linear system of equation (5.6).

Proof: The system (5.6) has n non-linear equations

$$q_l(z_1, z_2) = f_l(z_1, z_2) + \lambda \int_0^1\int_0^1 N_l[z_1, z_2; x_1, x_2; q_1(x_1, x_2) \dots q_n(x_1, x_2)]\, dx_2 dx_1$$

$f_l(z_1, z_2)$ are positive continuous and bounded functions, such that

$$\underline{v}_l \leqslant f_l(z_1, z_2) \leqslant \bar{v}_l,$$

$q_l(z_1, z_2)$ are non-negative bounded functions such that

$$0 \leqslant q_l(z_1, z_2) \leqslant w_l = y(z_1, z_2)/p_l$$

with

$$0 < \underline{v}_l \leqslant \bar{v}_l < w_l.$$

To solve the above system by successive approximations, we construct α functional sequences $\{q_l^\alpha(z_1, z_2)\}$ by the recursive relation

$$q_l^{\alpha+1}(z_1, z_2) = f_l(z_1, z_2) + \lambda \int_0^1\int_0^1 N_l[z_1, z_2; x_1, x_2; q_1^\alpha(x_1, x_2) \dots q_n^\alpha(x_1, x_2)]\, dx_2 dx_1. \tag{5.8}$$

Given and arbitrary continuous function

$$0 \leqslant q_l^0(z_1, z_2) \leqslant w_l.$$

Then, if

$$|\lambda| \leqslant \min \frac{\underline{v}_l}{K_l}, \frac{w_l - \bar{v}_l}{K_l},$$

all the functions of (5.8) satisfies for all α

$$0 \leqslant q_l^\alpha(z_1, z_2) \leqslant w_l.$$

To prove the convergence of the series, we will consider the differences

$$q_l^{\alpha+1}(z_1, z_2) - q_l^{\alpha}(z_1, z_2) = \lambda \int_0^1\int_0^1 \{N[z_1, z_2; x_1, x_2; q^{\alpha}(x_1, x_2)] - $$
$$- N[z_1, z_2; x_1, x_2; q^{\alpha-1}(x_1, x_2)]\} \, dx_1 dx_1 .$$

From the convex condition

$$q_l^{\alpha+1}(z_1, z_2) - q_l^{\alpha}(z_1, z_2) \leqslant L|\lambda| \int_0^1\int_0^1 \sum_{l=1}^{n} |q_l^{\alpha}(x_1, x_2) - q_l^{\alpha+1}(x_1, x_2)| \, dx_2 dx_1 .$$

This inequality implies

$$|q_l^2(z_1, z_2) - q_l^1(z_1, z_2)| \leqslant L|\lambda| K w n^2 ,$$

where $w = \max \{w_l\}$. By induction, we have that

$$|q_l^{\alpha+1}(z_1, z_2) - q_l^{\alpha}(z_1, z_2)| < w(L|\lambda|Kn^2).$$

Given condition (5.7), the last inequality implies the absolute and uniform convergence of the series involved.

6. CONCLUSIONS

It is difficult to reduce the output of the spatial difusion of innovations as a function of the quantity of messages received. In our paper, different alternative views were conceptually stressed. First, the economic nature of the adoption of innovations implies that, given the income contraint of the individual, the adoption of a particular good means the reduction (or total abandoning) of the consumption of another good. This reduction is also broadcast in the communication network and may reinforce the innovation impact.

Second, given the technological link among the goods consumed (complementarity or gross substitution), the adoption of one good means the adoption of another good (complementarity) or the abandoning of another one (gross substitution).

Third, given the intricacies of human behavior in relation to its social environment the adoption of innovations (social emulation) is not the only possible response. An individual's rejection of social trends (social differentiation) is not an untenable result. Also, a combination of both opposite responses at the same time by the same individual is possible in a society structurally differentiated.

GEOGRAPHIA POLONICA 42, 1979

DEMAND IN THE SPATIAL ECONOMY: I. HOMO DETERMINISTICUS

LESLIE CURRY

Department of Geography, University of Toronto, Toronto, Canada

1. INTRODUCTION

While there is little doubt that economic geography is the most advanced field of the discipline, its organisation is fairly chaotic being little more than a series of *ad hoc* notions brought forward as the occasion warrants so that analyses are never articulated and our understanding only piecemeal. The larger task then is to provide comprehensive theory which will allow a synoptic view. The narrower exercise is to phrase the demand side of the economy in geographical terms in order first, to see how concepts available for all economic agents work out for consumers and second and most important to build up a geographical theory of demand. The yet more limited, but still large subject of this essay is the geography of demand conceived deterministically. The approach we adopt is novel and allows the concepts developed here to be used for a probabilistically stated world and for other sectors of the economy. However, the treatment is self-contained and the results remain valid in a random economy whenever expectations are viable descriptions.

There appear to be two main classes of theory in economic geography. The first is more immediately applicable to the real world because its design is aimed at numerical substantiation. It tends to be more general in approach by excluding hypotheses about relationships which are implicitly catered for empirically. Good examples are regional interaction matrices or autoregressive processes. They tend to be only moderately interesting from a conceptual viewpoint. The second class makes much greater appeal to intuition although falling down in application. It needs special hypotheses in order to provide understanding at a general level but is usually not specific enough to be empirically viable. This is to be compared with the first class which avoids specific hypotheses and does not care for general understanding but can be applied to many real-life situations. Our approach is of the second type. A strong effort has been made to avoid technical issues. However, the type of theory written is intimately bound up with the technical apparatus available. We base our reconciliation and integration of apparently disparate explanations on particular mathematical structures which must be outlined, especially potential theory in its various manifestations.

2. GRAPHS AND POTENTIAL

Draw a graph representing an individual's preferences with an object at each vertex and a directed arc from a to b when he prefers a to b (Berge and Ghouila-Houri, 1965). Associate a degree of preference $\theta_{ab} > 0$ with each arc (a, b). For convenience $\theta_{ba} = -\theta_{ab}$ and for a consistent individual,

$$\theta_{ab}+\theta_{bc} = \theta_{ac},$$

The vector $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$ is a potential difference when there is an associated function $t(a)$ such that for every arc $i = (a, b)$ we have

$$\theta_i = t(b)-t(a).$$

For a cycle $\mu = (i_1, i_2, \ldots, i_k)$ with successive vertices $a, b, c, \ldots, z$ we have

$$\begin{aligned} \mu_{i(1)}\theta_{i(1)} &= t(b)-t(a), \\ \mu_{i(2)}\theta_{i(2)} &= t(c)-t(b), \\ &\ldots\ldots\ldots\ldots \\ \mu_{i(k)}\theta_{i(k)} &= t(a)-t(z). \end{aligned}$$

Summing,

$$\sum_{i\in\mu^+}\theta_i - \sum_{i\in\mu^-}\theta_i = 0,$$

where μ^+ and μ^- denote the arcs of the cycle oriented in a given sense and in the opposite sense, respectively.

Label an arbitrary vertex with the coefficient $t(a_0) = 0$.

Given an arc

$$i = (a, b),\ t(b) = t(a)+\theta_i.$$

Similarly for

$$i = (b, a),\ t(b) = t(a)-\theta_i.$$

Not only can the potentials of all vertices be defined in this way but are uniquely defined.

For a graph with vertices $a_1, a_2, \ldots, a_n$ and arcs $i = 1, 2, \ldots, m$, the incidence matrix $S = ((s_i^j))$ is constructed with

$$s_i^j = \begin{cases} +1 \text{ if the arc } i\in\omega^+(a_j), \\ -1 \text{ if the arc } i\in\omega^-(a_j), \\ \ \ 0 \text{ if the arc } i\notin\omega(a_j). \end{cases}$$

Each arc has $+1$ at its origin, -1 at its destination and forms a column of i of S. $\omega(a_j)$ is a coboundary.

From above,

$$\theta_i = t(b)-t(a).$$

Hence,

$$\theta_i = -\sum_{j=1}^{n} s_i^j t(j).$$

Thus, letting S^* be the transpose of $S(s^{*i}_j = s_j^i)$, a necessary and sufficient condition that a vector θ be a potential difference is that for the vector $t = (t_1, t_2, \ldots, t_n)$,

$$\theta = -S^*t.$$

3. PREFERENCE STRUCTURES

INDIFFERENCE CONTOURS

Let an individual set out the combination of amounts x, y, z, ... of goods X, Y, Z, he will consume in a given time. In a metaphorical experiment, he must alter the amounts x, y, of X, Y, in all possible ways without changing the quantities of other commodities. Any one change in combination can be regarded as a shift in coordi-

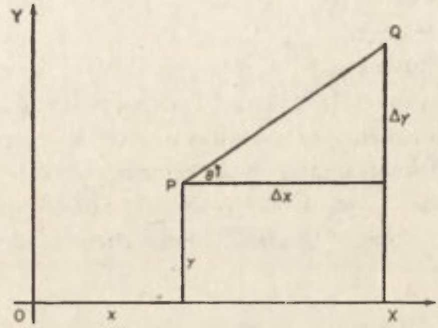


Fig. 1

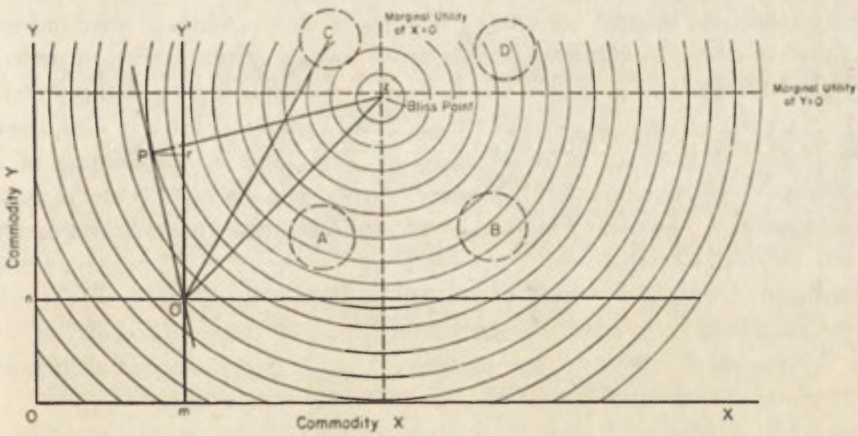


Fig. 2

nates as in Fig. 1 (Allen, 1932) from P to Q. The direction of PQ, which is all that is necessary for analysing static situations, is obtained from $\Delta y/\Delta x$ or tan θ. The individual will prefer some ratios to others, being willing to give up some of X in exchange for Y or vice versa. However, there will be a small range of values where he will be indifferent to the direction of the hypothetical exchange. Assuming continuity and, passing to the limit, the range of indifferent ratios assumes a unique value $-dx/dy = \varphi_y$ and here $dx+\varphi_y dy = 0$. For values > 0, trade-offs are preferred and for < 0 values, trade-offs are not wanted. The magnitude of PQ is also defined, i.e. $(\Delta x^2+\Delta y^2)^{0 \cdot 5}$ but we leave this for now.

Fig. 2 (Lerner, 1934) shows the indifference surface obtained for all the combinations of X and Y where Y is varied, X remaining the same. The table explaining the various quadrants is self-explanatory.

marg. utils.

A $X < 0$, $Y < 0$ diminishing marg. util. of Y; more X gained by trading Y
B $X > 0$, $Y < 0$ additional X obnoxious; accept more X only if more Y given
C $X < 0$, $Y > 0$ additional Y obnoxious; accept more Y only if more X given
D $X > 0$, $Y > 0$ additional X or Y obnoxious; accept more X or Y only if some obnoxious Y or X removed

From any point O', a curve can be drawn which shows the trade-off the individual will make at any market price. It is the locus of all points from which tangents $O'P$ to the indifference curves radius ZP pass through O'. It is circular because the locus of the apexes of right angled triangles on the same hypotenuse is a circle.

If the line of exchange is represented by the slope of $O'P$ the individual will move along it to P where the slope of the indifference curve is identical, giving up Pr of X for $O'r$ to Y.

TRADE-OFFS

Consider now the direction at P which is at right angles to the indifference surface $\varphi_x dx + \varphi_y dy = 0$. The separate components dx, dy of an infinitesmal movement in the direction of the normal are related as $dx/\varphi_x = dy/\varphi_y$, when positive movement is to a preferred region. These can be integrated giving a system of lines of preference which pass through all points in space. The components φ_x, φ_y of the preference direction are the marginal utilities of the goods X, Y, at the point x, y. The marginal utility of a good X to an individual is mainly dependent on the amount of X he possesses, then on the amounts of other goods he possesses and finally on the amounts of goods possessed by other individuals. The individual can make a choice locally between very small changes from any combination; he need not judge his relative preferences for widely separated combinations. While there is no need to make assumptions about the existence to total utility or about the measurability of utility, if the differential equation of the indifference loci is integrable it is then possible to find an index of utility which provides a measure of total utility.

Allen points out that the integrals of the individual's preference directions, the individual total utility, may be regarded as potential. The economic field of the individual can be regarded as a system of lines of preference and an orthogonal system of indifference surfaces. We now take this matter up more closely. It is important to note that we are working within the totally arbitrary Euclidean framework where relationships are linear. This is in fact a hypothesis about behaviour which we shall retain throughout with various warnings.

POTENTIAL

Let the x and y coordinates in the first quadrant represent quantities of two goods which the consumer desires in positive amounts. Each point (x, y) represents a combination of the two goods. Associate with each basket combination (x, y)

a scalar value $\varphi(x, y)$ where $(x > 0, y > 0)$. Lines of constant φ can be drawn and the gradient of φ can be derived.

$$\text{Grad } \varphi = \nabla\varphi = (\partial\Phi/\partial x)i + (\partial\varphi/\partial y)j,$$

where

$$\nabla = (\partial/\partial x)i + (\partial/\partial y)j$$

and i, j are unit component orthogonal vectors

Interpreting φ as a utility index, the contours are indifference curves. The set over which φ is defined is the consumer choice set. Potential φ is a function establishing a preference ordering and defines a potential field (Hum, 1972).

The line integral of the field between two points is path independent $\varphi_{21} = \int_{p_1}^{p_2} E\, ds$. With p_1 fixed, φ_2 is a scalar function of the position of p_2. This is the integrability condition of utility theory and defines rationality. The value of a line integral along any closed curve in the field (i.e. cirulation or curl) is zero.

Let $\vec{V}$ be a vector field in the (x, y) plane resulting from the potential field.

$$\vec{V} = v_1 i + v_2 j$$

and differentiating, we define its divergence

$$\nabla \cdot \vec{V} = (\partial v_1/\partial x) + (\partial v_2/\partial y).$$

If $\nabla \cdot \vec{V}$ is such that $\vec{V} = \nabla\varphi$ we have $\nabla \cdot \nabla\varphi = \nabla^2\varphi$ which is called div grad φ or the Laplacian.

If $\nabla \cdot \vec{V} = 0$, this is the continuity equation and when it is satisfied goods do not disappear. Such a field is called conservative, analogous to Kirchoff's law in network theory,

$$\text{Curl } \vec{V} = \nabla \times \vec{V} = \big((\partial v^2/\partial x) - (\partial v/\partial y)\big)k = 0.$$

Preferences cannot be circular and indeed this is why potential can be used to define preference structures.

4. ELASTICITIES

ELASTICITY OF DEMAND

Consider the indifference curves *I*, *II* in Fig. 3 (Allen, 1962). As the price of *X* changes from that indicated by the price line *AB* to that on *AB'*, the consumer moves from *I* to *II* and the quantity he consumes rises from *TR* to *TS* while his expenditure remains constant at *TO*. The slopes of *I* and *AB* at *R* and *II* and *AB'* at *S* happen to be such that their points of tangency at *R* and *S* are in a straight line defining a fixed $ outlay for *X*. Thus the elasticity of demand is equal to unity. Arc elasticity

$$E = \frac{(Ox - Ox')}{(Ox + Ox')} \cdot \frac{(OA/OB) + (OA/OB')}{(OA/OB) - (OA/OB')} = (-XX'/(O_x + O_x')) \cdot (OB' + OB)/BB'$$

$$= (-BB'/(OB' + OB)) \cdot (OB' + OB)/BB' = 1.$$

If *TS* had had a negative slope, this change in the amout demanded would have been greater for the same change in price and elasticity of demand would have been greater than one. With *TS* having a positive slope, elasticity of demand would have been less than one.

Clearly a whole demand curve can be obtained if *RS* is made to trace put a whole series of points defining a price-consumption curve. The amounts of X consumed are graphed against their corresponding prices on the vertical axis to give the familiar curve. If an existing price is marked off parallel to the horizontal axis against

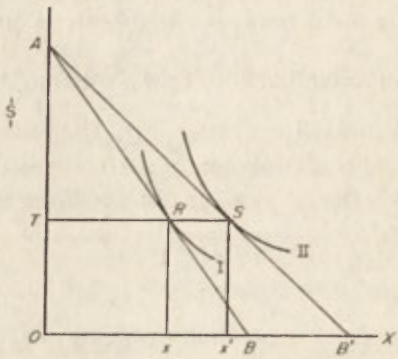


Fig. 3

all amounts consumed it cuts the demand curve which is downward sloping to the right. The area between the two lines and above the price line is called the consumers surplus as it represents this difference between the maximum amount a consumer is willing to pay and the amount he actually pays. One must be extremely wary of carrying through notions like consumer surplus when distance is the 'price' argument. Thus in outdoor recreation demand, it is asserted that consumer surplus occurs when those able to enjoy a recreational facility do not incur the full expense of the most distant visitors. But demand schedules will differ for various distance groupings and many people may not even use the facility (Denman and Prodano, 1972).

UTILITY MAXIMISATION

The contour method does not allow several independent variables to be handled at once. We shall use here the method of Lagrangian multipliers (Henn, 1961) but the use of linear programming should not be forgotten and in particular the use this makes of potential notions (Kantorovich, 1965). To express the preferences of consumption policy we have to be able to say that of any feasible combinations, $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{x}' = (x'_1, \ldots, x'_n)$, the target indices being x_v and x'_v, at least one of the two statements $\boldsymbol{x} \gtreqless \boldsymbol{x}'$ or $\boldsymbol{x}' \gtreqless \boldsymbol{x}$ be correct, where $\geqslant$ means 'is preferred or is equivalent to' and is transitive. With n competing aims represented by a system of $(n-1)$-dimensional indifference hypersurfaces, let the objective function ϱ assign a real number to each combination such that $\varrho(\boldsymbol{x}) \geqslant \varrho(\boldsymbol{x}')$ when $\boldsymbol{x} \geqslant \boldsymbol{x}'$. Combinations belonging to the same indifference curve have equal objective function values.

Starting from $\mathbf{x}'$, the set of situations which can be reached forms a transformation hypersurface $f(x_1, ..., x_n) = 0$. An optimal policy is one which maximises ϱ subject to the transfomation surface. The necessary conditions are

$$(\partial\varrho/\partial x_v) = \lambda(\partial f/\partial x_v) = 0, \; v = 1, 2, ..., n,$$

where λ is a Lagrange multiplier and $-f(x_1, ..., x_n) = 0$ or, equivalently,

$$\partial\varrho/\partial x_1 : \partial\varrho/\partial x_2 : ... : \partial\varrho/\partial x_n = \partial f/\partial x_1 : \partial f/\partial x_2 : ... : \partial f/\partial x_n$$

and $f(x_1, ..., x_n) = 0$

In the optimim situation the marginal utilities of any two policy aims are related to each as the rate of substitution is related to unity. In geometric terms, the optimal combination is at the point of contact between an indifference hypersurface and a transformation hypersurface.

Take a simple numerical example with objective function

$$\varrho = 100-9x_1-x_2^2$$

giving

$$\partial\varrho/\partial x_1 = -9 \text{ and } \partial\varrho/\partial x_2 = -2x_2.$$

A transformation curve is given externally as $(x_1+6)x_2-48 = 0$.
We have $\partial f/dx_1 = x_2$ and $\partial f/\partial x_2 = x_1+6$.
The optimum conditions are $-9/-2x_2 = x_2/(x_1+6)$ giving $x_2 = 6, x_1 = 2$.

Maximising a utility function is a standard tool of economic analysis and there is little point in mentioning papers which repeat it. It is only rarely that the distance relationships which result from the models provide insights. Evidently it is the content of analytic models rather than the formal grammar they employ which is decisive: of course this does not deny their value in numerical applications.

DEMAND FUNCTIONS

The table relating standard spatial demand curves, aggregate demands assuming a uniform distribution of consumers and elasticity of aggregate demand is provided from Moran (1966).

Demand functions	Demand in i from residents in j: S_{ij}, e.g.
1. Linear	$-a+b(\omega_i-\theta_{ij})$
2. Positive Exponential $S'' = (S')^2/S$	$a\exp(b(\omega_i-\theta_{ij}))$
3. More convex than (2) $S'' > (S')^2/S$	$a\exp((\omega_i-\theta_{ij})^b)$
4. Less convex than (2) $S'' < (S')^2/S$	$a\exp(\log(\omega_i-\theta_{ij}))$
5. Logarithmic	$-a+b\log(\omega_i-\theta_{ij})$

ω_i — utilities offered from a point i
θ_{ij} — costs incurred in projecting demand from j to i

Global demand at i from all points in space $S_i = \sum_{j=1}^{n} s_{ij}$	Elasticity of demand, $\varepsilon(S_i\omega_i)$
1. $-na+nb\omega_i-b\sum_{j=1}^{n}\theta_{ij}$	$nb\omega_i S_i$
2. $a\exp(b\omega_i)\sum_{j=1}^{n}\exp(-b\theta_{ij})$	$b\omega_i$ / $\varepsilon(s_{ij}b\omega_i)$ = elasticity of a simple link
3. $a\sum_{j=1}^{n}\exp(\omega_i-\theta_{ij})^b$	$b(\omega_i-\theta_{ij})^{b-1}\omega_i$
4. $a\sum_{j=1}^{n}\exp(\log(\omega_i-\theta_{ij}))$	$\omega_i/(\omega_i-\theta_{ij})$
5. $-na+b\sum_{j=1}^{n}\log(\omega_i-\theta_{ij})$	$b(\omega_i/(\omega_i-\theta_{ij})\cdot(a+b\log(\omega_i-\theta_{ij}))^{-1})$

5. PREFERENCES AND DISTANCE

ASPATIAL AND SPATIAL POTENTIAL

So far we have kept our potential theory formulation aspatial or rather the space is not explicit. Potential is simply a functional manner of representing a vector field, i.e., a collection of vectors, and has nothing inherently geographical about it. How far preference fields have an inherently geographical component is something which is argued somewhat obliquely by contrasting models and of which we shall have more to say. In short term analysis, it is reasonable to regard an individual as having a daily round among endogenously given locations and consumption preferences are formed relative to that round. Any distance factors involved are implicitly discounting the goods being evaluated. It is in this sense that a theory of individual demand divorced from its locational context is viable and potential. It is a strange quirk of intellectual history that a number of the early mathematical economists, Edgeworth (1881), Fisher (1892), Allen (1932) referred to potential theory in this context but then the idea died out, save for Hum (1972).

However, such an approach is inadequate for looking at an individual in relation to his environment, particularly to its spatial arrangement. We need theory which explicitly incorporates the two dimensional array of people and things with which the individual interacts and in which his preferences are formed and are expressed. Without concerning ourselves about the 'mental' mapping by which the physical landscape is recorded and evaluated in the brain, there must be some function relating the aspatial indifference surface to the real world. One's ideas of the world and the world itself must exhibit some degree of identity, however bizarre the transformations involved, simply in order for life to proceed. It is the virtue of the potential surface that it can be used both to depict locational structures and their concomitant preference structures. That potential methods also nicely straddle such complementary but difficult to handle pairs as configurations and flows or opportunities and interaction is a further advantage which will be brought out later.

DISTANCE AND UTILITY

Distance is normally regarded as a negatively valued quality aspect, the more of it there is the worse it is, so that a negative price elasticity will be evident. Thus in constructing an indifference curve for the distance from a delivery point to a consumer for, say, bread and the quantity of bread it is clear that large distances are grouped with large quantities and short distances with small amounts. If one wanted to deal in positive quantities alone distance could be replaced by 'place utility' to use a happy phrase of Wolpert's.

It is when we approach travel behaviour that distance is most directly involved. One approach to modelling its utility theory is to assert an individual's trip preferences are determined solely by the non-spatial properties of opportunities with distance merely limiting the range of feasible travel patterns. Another is to take account of the interrelations of opportunity interactions and their associated trips. For example, for each spatial opportunity i perceived by an individual, there exists some utility of interaction with i and some disutility of travel to i, depending on distance, with the difference being the net utility.

Rushton (1969a, b) falls within this group in seeking to separate preference and opportunity empirically, his remarks on method having theoretical significance. Paired comparison of locations provides a subjective ranking and allows investigation of transitivity in the preference matrix. The space preference function is to be independent of the particular opportunity map so that choices would be predictable given any such map. Size of town is deemed to stand as a measure for many urban attributes and conventional indifference curves can be drawn relating size of town and distance to town.

DISTANCE DISCOUNTING

Another example provided by Smith (1974) is for a trip maker to discount his anticipated opportunity interactions t_i in terms of the travel distance to these opportunities d_i by a factor $\alpha(d_i) \leqslant 1$ when α is his spatial discount function. Alternative patterns of interaction levels to any collection of spatial opportunities $i = 1, \ldots, n$ are evaluated in terms of their associated discounted interaction levels $(t_1 \alpha(d_1), \ldots, \ldots, t_n \alpha(d_n))$. The discount function is assumed to be subjective and an implicit, imputed chracteristic of an individual observable only indirectly in travel behaviour.

All possible travel configurations $c = (t, d)$ of an actor a are contained in a set C: he can compare all c's and order them transitively so that there exist travel preference structures for all possible weak orderings by a.

If c is modified by changing d_i then a can find some new level of t_i which will offset this change. Other things being equal, a prefers higher levels of interaction with opportunities (to lower) and shorter distances to opportunities (to longer). If d_i is changed, then the percentage change in t_i required to offset it is invariant both with respect to the particular opportunity involved and with respect to all other c's. A more special axiom which provides an exponential spatial discuonting function is incremental proportionality. Whenever an increase of Δ is made in d_i it may be offset by a percentage change in t_i which depends only on the magnitude of Δ.

POTENTIAL AND FLOWS

Earlier we treated marginal utility as potential so that the indifference curves were contours of constant marginal utility. From these we can obtain demand elasticity by comparing a rate of exchange slope with a preference slope (cf. Chap. 8). However, it is usually more convenient to treat prices as potentials, particularly when we are transferring our theoretical apparatus into the spatial domain. This can be done most readily in the present argument by assuming a uniform distribution of consumers/producers and no transport costs so that the elasticity of substitution of the good in question for dollars determines demand and thus price. Instead of remaining with elasticities we may now deal with slopes or price differences with flows of commodities and factors moving from low to high price locations. These flows are responding to the elasticities. Configuration and flow thus become the two faces of one coin.

Let V be a vector field in the (x, y) plane resulting from an assumed inverse first power law of attraction at a point, i.e. interactions are linear (Beckenbach, 1956). Its magnitude $|V| = 2km/R$ where $R = (x^2+y^2)^{0\cdot 5}$ and m is the value of the attraction at (x_0, y_0). The components of V are $V_x = -2kmx/R^2$ and $V_y = -2kmy/R^2$. Integrating, to obtain the potential $\varphi(x, y)$

$$\int V \cdot dR = \int_{(0,1)}^{(x,y)} V_x dx + V_y dy = km \log (1/R).$$

The partial derivatives are the corresponding components of V

$$\partial\varphi/\partial x = V_x,\ \partial\varphi/\partial y = V_y,$$

when

$$(\partial V_x/\partial x)+(\partial V_y/\partial y) = 0,$$

i.e.

$$\text{div } V = \nabla V = 0.$$

The density of the flow induced by the attraction does not change so that divergence is zero. In economic geography terms, goods coming into an area where there are no sources or sinks will leave again.

Similarly, let

$$(\partial V_x/\partial x)-(\partial V_y/\partial y) = 0,$$

i.e.

$$\text{curl } V = \nabla x V = 0.$$

Now the flow has no return tracks.

From a purely logical viewpoint we cannot have goods being sent off on wild goose chases and end up back where they started. Goods which do make out and back trips rationally, unwanted goods which are being returned, for example, cannot be explained in terms of potential.

V is the gradient of $\varphi(x, y)$ so that the lines where φ = constant are contours of equipotential and their orthogonal trajectories Q = constant are the lines of flow.

Let $f(x, y)$ be some continuous differentiable function of the coordinates. Construct in the plane a vector whose x, y components are equal to the respective partial derivatives. This vector is called the gradient of f and is written grad$f$ or ∇f_i

$$\nabla f = \hat{x}(\partial f/\partial x)+\hat{y}(\partial f/\partial y).$$

The relation of the scalar function f to the vector function ∇f is the same as the relation of the potential φ to the field V (Purcell, 1965).

$$d\varphi = (\partial\varphi/\partial x)dx+(\partial\varphi/\partial y)dy.$$

But $d\varphi = -V \cdot ds$ and since the infinitesmal vector displacement ds is just the gradient, then V is the gradient of φ.

We may note here that potential can be considered as a geographical multiplier by which structure at j, x_j, is the result of the direct effect of structure at i on events at j, x_{ij}, times the potential operator, φ_{ij}, which mirrors all the round-about effects of x_i feeding back to j

$$x_i = \varphi_{ij}x_{ij}.$$

PURCHASES AND DISTANCE

Beckmann (1971) has contributed to the analysis by treating the allocation of purchases of an individual among competing suppliers. *Ceteris paribus*, relative sales at a given location by two non-adjacent firms offering similar products

$$S_1/S_2 = \varphi(d(r_2)-d(r_1)+a_{12}), \tag{5.1}$$

where $d(r)$ is a measure of economic distance and a_{12} is a difference in attractiveness, e.g. price or quality difference.
To be able to write $a_{ij} = a_i-a_j$ it is necessary to have

$$a_{ij} = -a_{ji}, \quad a_{ij}+a_{jk}=a_{ik}.$$

Note that $S_2/S_1 = \phi(d(r_1)-d(r_2)+a_{21}) = \phi(d_1-d_2+a_{21})$.
To obtain a utility function consistent within the sales ratio above, let the c.i.f. price $p_i = d_i-a_i$ where a_i is here the degree of 'repulsion' (i.e. negative attraction) from the source so that the negative of f.o.b. cost and transport costs are being added. Eq. (5.1) becomes $S_1/S_2 = \phi(p_2-p_1)$.

Let expenditures on these commodities be small so that changes will not affect the marginal utility of money. In terms of utilities, we have

$$\partial u(S_1, S_2, \ldots)/\partial S_i = p_i.$$

Assuming $u = \sum_i(-\int_1^S i \log S\, dS)$

$$\begin{aligned}\log S_i &= -p_i,\\ \log(S_1/S_2) &= p_2-p_1,\\ S_i/S_2 &= \exp(p_2-p_1)\\ &= \exp(a_1-d_1)-(a_2-d_2)\\ &= \phi(d_2-d_1+a_1-a_2),\ \text{where } \varphi \text{ is exponential.}\end{aligned}$$

Market share

$$m_i = S_i/(S_1+ \ldots +S_i+ \ldots +S_n)$$
$$= 1/\phi(d_i-d_1+a_1-a_i)+ \ldots +\phi(d_i-d_n+a_n-a_i).$$

Assuming the exponential function again, $\varphi(x) = e^{-\lambda x}$

$$m = \exp(\lambda a_i - \lambda d_i)/\sum_j \exp(\lambda a_j - \lambda d_j).$$

Writing A_i, A_j for $e^{\lambda a_i}$, $e^{\lambda a_j}$ to improve recognition

$$m_i = A_i \exp(-\lambda d_i)/\sum_j A_j \exp(-\lambda d_j).$$

Since the denominator is the locational potential, a firm is patronised according to its share of the location potential.
Substituting r^α where $\alpha \geqslant 1$ for exp (λd)

$$m_i = (A_i r_i^{-\alpha})/\sum_j (A_j/A_i)(r_i/r_j)^\alpha.$$

In the case of equal attractiveness $A_j = 1$ so that

$$m_i = 1/\Big(1+\sum_{j\neq i} (r_i/r_j)^\alpha\Big).$$

In the region where $r_i < r_j$ for all $j \neq i$, if α is large, market share is near unity in the region and near zero outside. α is related to the degree of substitutability of commodities: indeed, if the commodity's mill price is zero so that c.i.f. price is proportional to distance, then α is the cross elasticity of demand. It is unfortunate that the potential which Beckmann obtains is so directly a function of the logarithmic utility assumption. Given the linear direct interactions postulated, the total indirect interaction should appear as logarithmic as given earlier.

6. STATIC SPATIAL EQUILIBRIUM

It is clear that without external changes, the laws of motion referred to will cause all temporal change to work itself out and a stationary state will result. In this equilibrium case, Laplace's equation holds

$$\nabla \cdot \nabla\phi = \nabla^2\phi = 0.$$

$\operatorname{div} F$ is the flux out of an infinitesimal area

$$\operatorname{div} F = (\partial Fx/\partial x)+(\partial Fy/\partial y).$$

Let $E = -\operatorname{grad}\phi = -(\hat{x}(\partial\phi/\partial x)+\hat{y}(\partial\phi/\partial y)$,

$$\operatorname{div} E = (\partial Ex/\partial x)+(\partial Ey/\partial y), \tag{6.1}$$

i.e. $Ex = (-\partial\phi/\partial x)$. Substituting in (6.1)

$$\operatorname{div} E = (\partial^2\phi/\partial x^2)+(\partial^2\phi/\partial y^2) = -\operatorname{div}\operatorname{grad}\phi,$$
$$\nabla^2\phi = (\partial^2\phi/\partial x^2)+(\partial^2\varphi/\partial y^2) = 0.$$

To have equilibrium, i.e. no net flows in or out of a region, the second derivative of the potential difference is zero.

Corresponding to the 2D potential equation $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 = 0$, the partial finite difference equation is $(\Delta_m^2 + \Delta_n^2)\ u(m, n) = 0$ where $\Delta_n f(n) = f(n+0.5) - f(n-0.5)$ and hence

$$\Delta_n^2 f(n) = f(n+1) - 2f(n) + f(n-1).$$

This can be written

$$u(m+1, n) + u(m-1, n) + u(m, n+1) + u(m, n-1) - 4u(m, n) = 0.$$

If the values of $u(m, n)$ are plotted in a square lattice, the value at any point is the arithmetic mean of the four surrounding points. The potential field is highly smoothed as occurs on a monthly calender (van der Pol, 1959). Flows could still occur from high to low potential but there are no local sources or sinks.

The Laplace equation is linear homogeneous so that any linear combination of potentials which individually satisfy the Laplace equation also satisfies the equation and hence must be a potential. Thus we add potential fields to obtain a resultant field. Preferences for high wages and for urban living, for example,

$$\nabla\phi_a + \nabla\phi_b = \nabla(\phi_a + \phi_b)$$

or for high wages and dislike of high living costs

$$\nabla\phi_a - \nabla\phi_b = \nabla(\phi_a - \phi_b)$$

are easily combined.

Component potential fields may be assigned different weights and if these latter are not constant over the whole field, the way in which they vary with position must be specified. Laplace's differential equation has been studied exhaustively precisely because it represents situations which interest us, the 'equilibrium' conditions on a two (or 0, 1, 3, n) dimensional surface when some initial conditions have been imposed and the equations of motion have taken over so that now the effect of the initial conditions has passed away and present conditions are independent of time.

As long as a man is regarded as independent of other men using his resources optimally to satisfy given preferences, little insight can be gained into the question of changing utilities. Strictly speaking, in a deterministic universe all of the future is known so that a change can be no more than a materialisation of foreknowledge. The sort of thing which can alter preferences, consumption experience for example, implies the gaining of knowledge and therefore a previous time when the world was incompletely specified. Again, the spread of new tastes implies an interaction among peoples' preferences which is debarred by definition.

7. BID PRICES FOR LAND

A bid price curve is the set of prices for land the individual could pay at various distances while deriving a constant level of satisfaction (Alonso, 1964). If the price of land were to vary with distance as in the bid price curve, the individual would be indifferent among locations. A family of bid price curves exists for the individual corresponding to different levels of satisfaction. We wish to find the bid prices p_0 for a quantity of land q at location t_0 for an individual such that a level of satisfaction u_0 is achieved. His income y, his commuting costs function $k(t)$ and the price of the composite good remainder p_z are given. His utility function appears as

$$u_0 = u(z, q, t_0). \tag{7.1}$$

The budget balance equation describes his locus of opportunities as

$$y = p_z z + p_0 q + k(t_0). \tag{7.2}$$

He is free to vary the quantities q and z to maximise u_0 for given t_0 at price p_0. Differentiating the above equations,

$$du_0 = u_z dz + u_q dq = 0,$$

$$dy = p_z dz + p_0 dq = 0,$$

giving

$$u_q/u_z = p_0/p_z. \tag{7.3}$$

Using equations 7.1, 7.2, 7.3, z, q and p_0 may be obtained.

To find bid price $p_i(t)$ as a function of distance (t), t is substituted for t_0 and $p_i(t)$ for p_0, and one of the variables, say t, as a parameter in terms of which the other three variables may be expressed. The slope of this curve is obtained by differentiating the amended eqs. 7.1, 7.2, 7.3, while holding q constant to obtain

$$du_0 = u_z dz + u_t dt + 0,$$

$$dy = p_z dz + q(dp_i(t)/dt)dt + (dk/dt)dt = 0.$$

Combining and rearranging,

$$\frac{dp_i}{dt} = \frac{p_z}{q}\frac{u_t}{u_z} - \frac{1}{q}\frac{dk}{dt}.$$

The negative slope of the bid price curve is contributed from both terms on the right hand side, i.e. the disutility of distance ($u_t < 0$) and commuting costs ($dk/dt > 0$)

Alternatively,

$$u_t/u_z = ((q dp_i(t)/(dt) + (dk/dt))/p_z.$$

The ratio of the marginal utilities on the left is the marginal rate of substitution of distance for other goods. The ratio on the right is of marginal costs with the numerator as movement away from the centre. This marginal cost must represent a saving in order for the bid price curve to maintain a given level of satisfaction as outward movement produces disutility. This saving must be in land costs, i.e. quantity of land times change in land price. The bid price definition thus allows the substituting of land and the composite good for accessibility to maintain a given level of satisfaction. The equilibrium location of the household occurs where the rent gradient is tangential to the lowest achievable bidprice curve. Households with the steepest bid price curves will locate nearest the centre.

The bid price curve should be contrasted with the indifference curve. The latter describes combinations of goods among which the individual is indifferent, without regard to prices. The former is a set of combinations of land prices and distances among which he is indifferent. Evans (1973) asserts that "In the study of residential location, the demand curve is a concept which is of little use. Other tools of analysis have to be developed — hence the concept of the bid price curve". However, whether it is adequate for its task is another matter. Harris (1972) argues "It is quite obvious that the Alonso model and its basic modifications cannot deal with the externalities of social preferences in an equilibrium framework. The bid rents depend on the locational pattern and the locational pattern depends on the bid rents." While Alonso could not be expected to handle the externalities problem, it is difficult to see how land prices and distances could be treated as independent.

8. COMPLEMENTARITY AND SUBSTITUTABILITY

INTERACTION AND INDIFFERENCE CURVES

It is clear that the trade-off between goods will depend on the extent to which they compete with or complement one another. Fig. 4 (Fisher, 1892) shows an indifference surface of competition with the individual being indifferent to the combination of X and Y he uses. If articles are competitive the indifference curves are

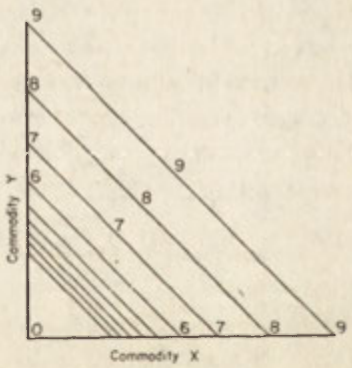


Fig. 4

elliptical with the long axis sloping down from left to right; in the case of perfect substitutes these reduce to parallel straight lines whose intercepts in the Y and X axes are inversely proportional to the constant ratio of their marginal utilities. For complementary articles, the elliptical indifference curves have their long axis sloping down from right to left and for perfect complementarity the family of curves reduces to a straight line passing through the origin. The intuitive notion of the complementarity of two commodities is clear: increased consumption of k will increase the marginal utility of i and vice versa. Mathematically we use the second order cross partial derivatives to provide a cardinal definition, $\partial^2 u/\partial x_i \partial x_k > 0$. Similarly substitutability implies a negative value with the dividing line of zero implying independence.

HICKSIAN DEFINITIONS

The sign of the off-diagonal elements of the substitution matrix determines whether two commodities are substitutes or complements according to Hicksian definitions (Phlips, 1974). Subtraction of the 'general' substitution effect from these elements gives the specific substitution effect which provides information on the structure of preferences in terms of the second order partial derivatives of the utility function.

The Hicksian ordinal definitions are taken from the Slutsky equation

$$\partial x_i/\partial p_j = k_{ij} - x_j(\partial x_i/\partial y), \quad k_{ij} = dx_j/dp_i.$$

The uncompensated price rise on the left causes a reduction in purchasing power of a consumer on the far right so that his general demand goes down — the income effect — and he switches in the central term some of his demand to relatively lower priced substitutes — the substitution effect. k_{ij} is the cross substitution effect and indicates complementarity where $k_{ij} < 0$. Only when it is positive and equals the compensated income effect on the right which is usually negative so that the cross price elasticities are zero can a demand equation of a good stand independently, i.e. $\partial x_i/\partial p_j = 0$.

The notions of complementarity and substitutability of goods can be considerably embellished. Substitutable goods are those serving similar functions but usually having a great variety of 'styles'; they are thus represented by 'shopping goods'. Complementary goods have linked but different functions and are typified by 'convenience' goods. The former are thus of higher order in the seettlement hierarchy. They have a spur to locate together since they cannot afford not to include themselves in an inventory located for convenient searching. Complementary goods are more locationally independent, any stimulus to contiguity being more for convenience than for economy.

SUBSTITUTABILITY AND INCOME

The straight line AB represents a consumption combination of Y and X purchasable for a fixed sum; as such, it is a partial income line (Fisher, 1892). The individual will select that combination which will maximise his utility, at the point I where AB is tangent to an indifference curve. That arrangement which maximises utility and therefore potential will be chosen (cf. Chap. 4 for a mathematical treatment). The line AB describing quantities is inversely proportional to the corresponding prices. If prices remain the same but income is increased AB simply moves further away from the origin while remaining parallel. While this will generally increase the use of both commodities, these are unlikely to be in a constant ratio. For the point of tangency to follow a straight line from the origin as income rises, homogeneity in the first degree is necessary; not only are ratios fixed but the rise in 'altitude' is also so that there are no 'scale economies' for the pair of goods.

If relative prices change the greatest change in relative quantities occurs with substitute goods. Fig. 5 is Fisher's representation of two grades of a commodity, a superior one on the horizontal B axis and an inferior one on the vertical A axis with the bliss point on the *B* axis. It can be seen that the income line for a poor person is near the *A* axis as he can only afford inferior goods, but as income gets larger for richer and richer people more and more of the superior good is bought. It may be seen that the slope of *AB* determined by relative prices, is such that it can remain tangential to the indifference curves of the poorer persons and the cheaper good thus stays on the market. A slope of unity, and the inferior good would almost go out of use. Fisher makes the interesting point that if rich and poor are geographically separated so that two distinct markets exist, the price lines will follow the general trend of the contours, flatter in poor districts and steeper in rich. In a rich market a slight difference in qualities of goods will cause a large difference in their prices and many qualities will be distinguished. In poorer areas different grades may not exist and will sell for nearly the same price.

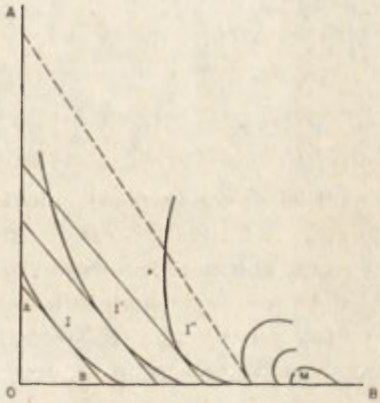


Fig. 5

The use of ordinal rather than cardinal definitions of utility, while preferable in many ways, did produce contradictions because these definitions were not invariant to a monotonic increasing transformation of the utility function. The Hicksian definition is biased in favour of substitutability and gives contradictory results. This stimulated development of the ideas of separable utility.

LEXICOGRAPHIC ORDERING

An individual is assumed to order his preferences into orthogonal dimensions and then rank these dimensions in order of importance (Taylor, 1973). Of say two alternative goods, he prefers the one which is closer on the most important dimension to his optimum point — and so on through less salient dimensions if earlier ones are tied. He can be thought of as having a utility function for each dimension but he does not, and cannot, combine these utilities. There are no indifference curves associated with an individual's preferences because, although there is a complete quasi-ordering based on indifferences, they cannot be represented by a single real valued function. The name arises because of the resemblance to a dictionary: the most important (or initial) letter is used first, then the second and so on, each letter representing an independent dimension. The 'distance' of an alternative from an individual's optimum point within a dimension is not represented by this term.

The concept arose because of a criticism of the von Neumann-Morgenstern 'continuity' postulate that "if f is preferred to g and g to h then there is a probability p such that g is indifferent to a gamble with outcome pf and $(1\text{-}p)h$". If f is two pins, g is one pin and h is death, then p would not normally exist. The lexicographic principle would handle this nicely by placing a subjective value on life and death which is not comparable with that arising from ownership of pins.

SEPARABLE UTILITY

Partition the consumption set into subsets which comprise goods that are closer substitutes or complements to each other than to members of other sets. Let n commodities be partitioned into m groups with $n_r (r = 1, ..., m)$ commodities in each group $\left(n = \sum_{r-1}^{m} n_r\right)$

$$u = f(x_{11}, ..., x_{1n_1}, ..., x_{2}, ..., x_{rn_r}, ..., x_{m1}, ..., x_{mn_m})$$
$$= F\big(f_1(x_1), f_2(x_2), ..., f_r(x_r), ..., f_m(x_m)\big). \qquad (8.1)$$

Note that additivity of the subsets is not necessary, only that the left hand side and right hand side of the expression are equal. Evidently the notion of subsets of consumption goods is similar to the notion of factors in production theory or indeed, to regions in trade theory. For a utility function to be separable, the marginal rate of substitution between any two goods within the same group must be independent of the utility of any good in any other group (Phlips, 1974)

$$(\partial/\partial x_{qk})(\partial f/\partial x_{ri})/(\partial f/\partial x_{rj}) = 0.$$

Here good i and j belong to the same group r while good k belongs to a different group q. This condition is known as 'weak separability', and allows a household to first allocate its income among the m broad consumption groups, $y_r(\sum_r y_r = y)$, and then purchases can be made within each branch with no further reference to purchases in other branches.

From (8.1),

$$\partial u/\partial x_i = (\partial u/\partial f_r)(\partial f_r/\partial x_i),$$

$$\frac{\partial^2 u}{\partial x_i \partial x_k} = \frac{\partial^2 u}{\partial f_r \partial f_q} \frac{\partial f_r \partial f_q}{\partial x_i \partial x_k} = \left(\frac{\partial^2 u}{\partial f_r \partial f_q} \Big/ \frac{\partial u}{\partial f_r} \frac{\partial u}{\partial f_q}\right) \frac{\partial u}{\partial x_i} \frac{\partial u}{\partial x_k}.$$

At equilibrium,

$$\partial u/\partial x_i = \lambda p_i,$$
$$\partial^2 u/(\partial x_i \partial x_k) = T_{rq} p_i p_k, \qquad (8.2)$$

where

$$T_{rq} = \lambda^2 \partial^2 u/(\partial f_r \partial f_q)/(\partial u/\partial f_q).$$

T_{rq} is the interaction coefficient between groups r and q: the sign of the difference between two of these coefficients is invariant to any monotonic increasing transformation of the utility function. A pair of independent groups is used to calibrate all other coefficients. A negative difference means substitution, and is reversed for complementarity. It may be seen from (8.2) that the intuitive definition of interaction between commodities stemming from cardinal concepts now squares with ordinal notions. It would appear that we maywork with cardinal utilities and have a translation available to move into ordinal ideas if desired.

9. INDIRECT INTERACTIONS IN UTILITIES

FEEDBACK IN UTILITY INTERACTIONS

A consumer describes his welfare position in real quantities and he has to maximise his utility function $V(x_1, \ldots, x_n)$ subject to $p'x = M$, i.e. unit price × quantity = income.

The vector of first-order derivatives is $\boldsymbol{v}(x)$ and the matrix of second-order derivatives is $\boldsymbol{V}$. The off-diagonal elements of $\boldsymbol{V}$

$$v_{1,2}(x) = (\partial^2/\partial x_1 \partial x_2)V(x_1, x_2|x_3, \ldots, x_n)$$

are measures of direct complementarity between x_1 and x_2 given a certain position in $(x_3, \ldots, x_n)$ and the larger the diagonal elements v_{ii}

$$v_{1,1}(x) = (\partial^2/\partial x_1{}^2)V(x_1|x_2, \ldots, x_n)$$

the more rapid the decrease in the value of marginal units.

For a change in equilibrium caused by a change in price $\boldsymbol{\Delta p}$, if an individual can be compensated by an income increase ΔM to neutralise the monetary effect, we have

$$\lambda(M_0+\Delta M, \boldsymbol{p}_0+\boldsymbol{\Delta p}) = \lambda(M_0, \boldsymbol{p}_0).$$

The change in equilibrium is given by

$$\boldsymbol{V\Delta x} = \lambda\boldsymbol{\Delta p} = \boldsymbol{v}(\boldsymbol{x}_0+\boldsymbol{\Delta x})-\boldsymbol{v}(\boldsymbol{x}_0)$$

leading to a solution

$$\boldsymbol{\Delta x} = \lambda \boldsymbol{V}^{-1}\boldsymbol{\Delta p}.$$

While it is possible that an individual would know how far a change in consumption $\boldsymbol{\Delta x}$, would change the marginal evaluation $v_1(x), \ldots v_n(x)$, it is highly unlikely that he could know or guess the total complementary relations reflected by $\boldsymbol{V}^{-1}$ (Van Praag, 1968).

Setting $\lambda°$ to unity for convenience, he could perhaps calculate

$$v_{11}\Delta x_{11} = \Delta p_1.$$

With a new budget $(x_1+\Delta x_{11}, x_2, \ldots, x_n)$ he proceeds to Δx_{12} where he assumes he has already bought Δx_{11}

$$v_{21}\Delta x_{11}+v_{22}\Delta x_{12} = \Delta p_2.$$

He thus finds a first step $\boldsymbol{\Delta x}^{(1)} = (\Delta x_{11}, \ldots, \Delta x_{1n})$.
Starting from $(\boldsymbol{x}+\boldsymbol{\Delta x}^{(1)})$ he proceeds to find step $\boldsymbol{\Delta x}^{(2)}$ etc. With convergence, this method leads to a unique solution of $\boldsymbol{V\Delta x} = \boldsymbol{\Delta p} = \boldsymbol{V}^{-1}\boldsymbol{\Delta p}$. Convergence occurs

if the marginal utility of money decreases as income increases, a traditional assumption. V^{-1} is the total complementarity effect built up from several rounds of direct complementarity effects between pairs of goods.

If income is not compensated, the substitution effect becomes $\lambda(I-E)V^{-1}p$ when $(I-E)$ is a correction factor due to the monetary restriction.

If an individual plans a redistribution of his expenditures within a certain commodity category only, the maximisation problem is

$$\max U(x_1, \ldots, x_k | x_{k+1}, \ldots, x_n)$$

subject to $p_1', x_1 = M_1, x_1 \geqslant 0$.

The first group is that of decision variables and the second is the set of conditions: x_1 stands for the k vector and x_2 for the n-k vector. The price vector p is split up similarly into p_1 and p_2 as is his income, M_1 and M_2. The matrix of direct complementarity effects may be partitioned as

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix},$$

where

$$v_{ij} = \frac{\partial^2}{\partial x_i \partial x_j} V(x).$$

The matrix of conditional direct complementary effects is the matrix V_{11} and the marginal effects of a change in (p_1, M_1) is calculated as previously but using V_{11} instead of V.

The conditional income effect

$$\partial x_1 / \delta M_i = (p_1' V_{11}^{-1} p_1)^{-1} V_{11}^{-1} p_1$$

and the pure substitution effect

$$\partial x_1 / \partial p_1 = \lambda(I - E_1) V_{11}^{-1},$$

where

$$E_1 = \partial x_1 p_1' / \delta M_1.$$

In general the effects of a change in consumption will be greater since only a partial redistribution of expenditures is possible.

THE POTENTIAL MULTIPLIER

We have introduced the feedback effects in utility which result when a preference pattern is disturbed but treated the calculation of the total complementary relations from the direct effects as an exercise in matrix manipulation. It is much better understood in terms of potential theory.

Let $p_{ij}^{(1)}$ represent the amount of direct interaction between i and j or in matrix form P. But there will also be interaction via another good k as first-order substitution effects are looked at (Sheppard, 1976).

$$p_{ij}^{(2)} = \sum_k p_{ik} p_{kj} = P^2$$

The kth order effects are obtained as indirect interactions work themselves out

$$p_{ij}^{(k)} = P^k.$$

To obtain the total effect each of these subsidiary effects needs to be added in. Thus

$$m_{ij} = \sum_{k=0}^{\infty} p_{ij}^{(k)},$$

or

$$M = I+P+P^2+P^3+ \dots +P^k+ \dots$$
$$= (I-P)^{-1}.$$

This multiplier is a potential comparable to the integral of the linear interaction discussed earlier.

10. INTERACTING PEOPLE

AGGREGATION OF PREFERENCES

Preferences belong to the individual so that, while a community demand curve can exist as a semi-empirical fact, its analysis is not possible. It is disconcerting to read (Phlips, 1974) "The attitude of most applied econometricians... is simply to ignore this aggregation problem and adopt a third approach by formulating aggregate relationships directly from the theory of the individual consumer" and "of all errors likely to be made in demand analysis the aggregation error is the least troublesome". Presumably interacting individual demands are excluded from these statements. So we are left with that fictional character, the representative individual who stands for the group and somehow reflects its averaged behaviour.

On another aspect of aggregation, demand theory refers basically to the individual commodity yet it is doubtful if calculation in terms of individual goods is more than occasional. There is a tremendous degree of interaction between a person's preferences and the linkages mirror life styles, the existence of which cut down greatly on both the degree of differentiation of individuals and on their hypothetical bookkeeping in utilities. Things seem to be very much simpler than the incredibly complex possibilities that theory allows for. The environment of the person does not need to provide for a myriad of possibly conflicting needs. Preference sub-systems have developed which can be catered for by environmental subsystems which need to satisfy only a few objective functions.

Aggregation of individual demands is easier to provide for in the case of probabilistic preferences but understanding its rationale requires a background in that subject.

EXTERNALITIES INTRODUCED

The previous treatment of individual spatial preferences and locational economic environment as independent follows the pattern of economics regarding independent utilities. The 'classical' approach to the economic landscape is designed to separate the individual from his environment so that allocation can be understood as a set of decisions. Resources, technology and preferences are given and are related by behavioural rules using prices as parameters of the environment. The economist has devised the notion of perfect competition for this purpose. The agent is self-contained reacting only to prices, competitive strategies do not exist and are not needed; markets exist without effort, firms need no policies for marketing or choice of technology; site selection is a matter of transport costs and rents. External economies are allowed to impinge on the individual but since they are reflected in prices no new principle is involved. They represent the savings from agglomeration and problems do not arise if there is no concern with the mechanism of establishment. Shackle (1972) insists that economists prefer the question 'What will things be like?' allowing a unique solution and a prescription for action for all individuals rather than 'How will things happen?' which demands concentration on the routes to equilibrium and a variety of answers.

Externalities are to be treated in their narrow sense of allowing side payments which maintain the use of rational decisions within groups of dependent individuals. Much of the work in economics is concerned with the adjustments necessary to reach Pareto optimal conditions and bargaining within small groups, usually an unprofitable mode of analysis in terms of generality of insight. One aspect of inter dependent utilities on which little has been said concerns the relative position an individual prefers to have in the social, income or other hierarchies. Obviously, Pareto optima have no meaning since any arrangement will be impossible to improve on (Mishan, 1969; Haavelmo, 1970).

Beyond this there are spatial externalities in the broad sense: individuals are now dependent on each other and micro-situations are partly functions of macro-situations and vice versa. The economic problem cannot be conceived as a set of individual decisions to satisfy preference criteria in manipulating the environment because preferences are partly a function of that environment. But the treatment of this subject, however inadequately, requires a separate paper in its own right.

NECESSARY CONDITIONS

Externalities in the narrow sense generally have as necessary conditions of their existence some of the basic attributes of society: law and its enforcement and the geographical contiguity of people. There are of course, other more specific reasons for the failure of the market to achieve efficiency due to one man's consumption entering into another man's utility. Among the former conditions is the problem of excluding persons in an area from benefiting from services they have not paid for. While there are some private exclusion costs involved, it is clear that this is mainly a question of law and order costs and that these are biased in their impact on societal groups. Exclusion is often not possible and the problem is to collect payments for services provided. Tiebout (1961) states the ideal case for a public facility, the market size of which is to be decided. Residents are asked to reveal their *true* preferences, i.e. the amount a person is willing to pay for a certain sized facility at a certain distance. To choose the optimum size of precinct, pick the combination which yields the largest surplus in proportion to the cost per resident. Taxes could either be (1) in proportion to share of total benefits; then the surplus received will be proportional to tax paid, or (2) so that each person's surplus is equal. However, for people to reveal their true preferences would be irrational and many of them would prefer to take a free ride. Shibata (1972) remarks that the law plays a much more significant role in negotiations relating to consumption interdependence than in production externalities. The location of final settlement and the distribution of welfare is less flexible and more arbitrary because tastes are apparently less substitutable than productive factors.

Much, if not most of the externality effect is based on the closeness of people in space. The hazards of the wayward cricket ball on neighbouring houses is a recent British case in point. One less obvious aspect is that the number of neighbours is usually small so that negotiations over externalities is among small numbers. There are no free entry conditions — the monopolistic effects of space à la Losch — to produce competition and thus an equilibrium point.

CLASSIFICATION

In initially classifying external effects on utility, Leibenstein (1968) used popular categories. First, the bandwagon effect described the extent to which the demand for a commodity is increased because others are also consuming it. Second, the snob effect was the reverse of the previous effect. Last was the Veblen effect, conspicuous consumption in which demand is increased at higher prices. A more complete and more suggestive classification in Bish and Nourse (1975) is listed below.

(+, +)	mutual attraction	high complementarity
(−, −)	mutual dislike	high substitutability
(0, 0)	pure indifference	independence
(+, 0)	attraction and indifference	medium complementarity
(+, −)	attraction and dislike	separable sets
(0, −)	indifference and dislike	medium substitutability

The column on the right exploits the similarity between the substitution terminology for two commodities in one person's utility function and that for one commodity in two different person's utility functions. But while the concepts are much the same — and are thus linked to potential — the actual goods are different. The means for smoking and drinking may be complementary in one person's mind but either product may produce avoidance between two persons. These terms could be translated into covariance measures for spatial externalities, negative correlation for substitutability and positive for complementarity and used to describe areas. We shall first summarise an analysis of Buchanan and Stubblebine (1968) as introduction to more formal analysis of externalities.

INTERACTING UTILITIES

Consider the desire for interaction between two persons *A* and *B* as well as their feelings of privacy as affected by a barrier between them of various possible sizes and costs. An externality exists when the utility of a person A depends not only on his own activities $X_1, X_2, \ldots, X_m$ but also on the activity of another person B, Y_1

$$u^A = u^A(X_1, X_2, \ldots, X_m, Y_1).$$

The marginal externality $\partial u^A/\partial Y_1$, assumes that the variation in Y_1 is evaluated relative to 'equilibrium' values for the X's adjusted to the externally given value of Y_1. Fig. 6(a) shows the indifference contours: it is assumed that B's utility increases with difficulty of interaction up to a reasonably high level, while A's first increases to ensure privacy then decreases reflecting a desire to communicate and finally levels

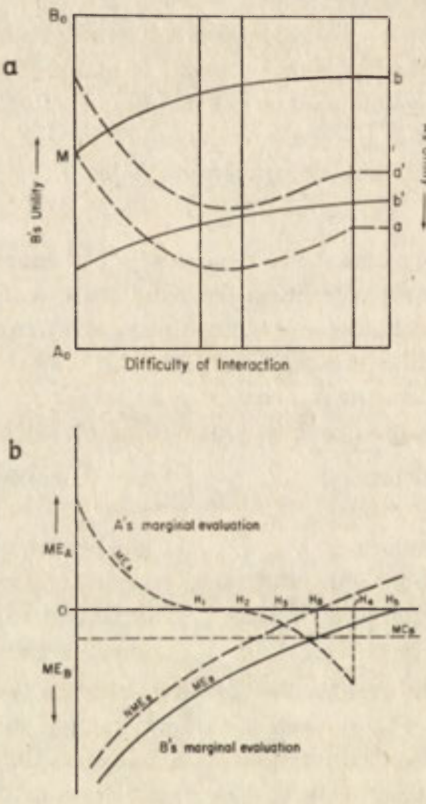


Fig. 6

off, since changes in the very small interactions are irrelevant to him. Fig. 6(b) shows the marginal evaluations of the extent of the barrier to interaction which it is B's task to pay for and control. The marginal cost of this, i.e. the disutility of goods foregone, is added to B's marginal utility (the negatives of the slopes of successive money-activity indifference curves for increased barriers). It is assumed that marginal costs are constant, shown by the curve MC. B's position of private utility maximising equilibrium is at H_B where $MC_B = ME_B$. This is shown as within the range of marginal external diseconomies for A so that B's behaviour is Pareto relevant in that A could compensate B for reducing his barrier. Pareto equilibrium is at H_3

where A's marginal evaluation ME_A equals the 'net' marginal evaluation of B, NME_B.
(1) In the range OH_1 a marginal external economy exists because a small change in B's activity will change A's utility in the same direction, i.e.

$$\partial u^A / \partial Y_1 > 0.$$

It is potentially relevant when the activity generates a desire on the part of the externally benefitted party A to modify the behaviour of B through persuasion etc. A wants B to increase his resources devoted to the activity when

$$\partial u^A / \partial Y_1 |_{Y_1 = \bar{Y}_i} > 0,$$

where $\bar{Y}_i$ represents the equilibrium value of Y_1 in which the marginal cost to B equals marginal utility in terms of money.
(2) Next in the range $H_1 H_2$, an infra-marginal external economy exists implying that while incremental changes in B's activity Y_1 have no effect on A's utility, the total effect of B's action will have increased A's utility, i.e.

$$\partial u^A / \partial Y_1 = 0$$

and

$$\int_0^{Y_1} (du^A / dY_1) dY_1 > 0.$$

However, these are irrelevant since no change in B's activity would increase A's utility. In this case A has achieved an absolute maximum with respect to changes over Y_1.
(3) Over $H_2 H_4$ marginal external diseconomies are exerted by B's activity

$$\partial u^A / \partial Y_1 < 0$$

and these are potentially relevant to A

$$\partial u^A / \partial Y_1 |_{Y_1 = \bar{Y}_i} < 0$$

so that he wishes B to reduce his activity.
(4) Beyond H_4 infra-marginal external economies or diseconomies exist and are potentially relevant

$$\partial u^A / \partial Y_1 |_{Y_1 = \bar{Y}_i} \neq 0.$$

The direction of the effect is dependent on the ratio between the total utility derived from privacy and the total reduction in utility derived from obstructed interaction. Shibata (1972) makes the point that when external economies in consumption occur it is unlikely that two agents in negotiation will reach a collective solution which also maximises their independent optima.

AREAL HOMOGENEITY

The first example is of (+, +) type interactions. Tiebout (1956) takes off from the position that for goods and services involving considerable externalities so that pricing will be arbitrary, nevertheless in a spatial economy a consumer cannot avoid revealing his preferences. In the case of public goods, at the local level, a consumervoter will pick the community which best satisfies his preferences. Given assumptions on independence of communities and mobility of knowledgeable consumers, each locality has an aggregate of demands that will approximate the true preferences of the consumers. Should the costs of a service rise causing a rise in taxes, some people may leave causing a fall in these services. Mobility is the mechanism of adjusting demand, whether it be walking to a community or walking to market.

Fitch (1972) objects to this famous 'voting with your feet' exposition since the homogeneity of preferences by communities which results is income related with only the wealthy groups having a vote. He seems to prefer communities offering somewhere between $(+, 0)$ and $(0, -)$. Local facilities are better thought of as multi-purpose packages; the greater the number of products, the more likely the package is to be adopted since voters are willing to endorse the whole thing rather than lose a favourite item in it. It is evident that more rigorous analysis is required with postulates having some claim to realism.

The final geographical example is from Bish and Nourse (1975) and depicts the $(+, -)$ category. They quote an illustration by Bailey of eight parallel streets of identical houses occupied by two income groups, high and low, of equal size. All persons like to live surrounded by high income houses. In these circumstances, *A*, *B* and *C* streets, say, could be occupied by low income families and priced at P_l and *F*, *G*, *H* by high incomes at price P_h with *E* selling for less among high income families at P_{h-d} and *D* for more than other low income properties, price P_{l+p}. If prices of *ABC* were the same as *FGH*, then *D* at $P_{l+p} > P_{h-d}$. So long as $(P+p)-(P_h-d)$ was greater than the cost *C* of converting high income property to low income use, owners of property in *E* would convert. However, if the houses were not owned individually, should *E* convert as above, properties on *D* would lose their premium as the border low income street; at the same time *F* would then become the bordering high income street and its prices would drop to $P_h - d$. The profits for the single owner would be $((P_l+p)-(P_h-d)-C)-p-d = P_l-P_h-C$. When $P_l = P_h$ and $C = O$, then no conversion will take place. It may be seen then that conversion will go further in the individual ownership case because no account is taken of the impact of preferences on surrounding property. Property prices will not normally lead to efficient allocation if interdependencies among land uses exist.

This is a very puzzling aspect of spatial externalities. If the quality of each unit of the environment, say a house, adapts to the quality of the environment, i.e. to neighbouring units, what will the quality be? Housing growing old in the inner area of a city may head into slum conditions or be 'white-painted', 'gentrified' etc. Indeed this is a very common feature of geographical specialisation and appeal to the historical narrative unsatisfactory explanation.

11. CONCLUSION

There is a great simplification in the demonstration that both geographical space and mental ordering can be conceived as vector fields and as graphs and treated in the same terms. The concept of potential does systematise apparently disparate notions: that the off-diagonal elements of the matrix of second order derivatives of the utility function is integrally related to the gravity model is certainly not obvious. The complementarity of map and flow, while more widely accepted, is only treated adequately by potential theory. The abstractions we make may be highly unreal and the situations we treat purely static but it does appear necessary to traverse this route before getting fancier.

Obviously the next step is to the random economy by studying individual probabilistic preferences with measures such as covariances and mean-variance or entropy-expected loss ratios. Choice becomes more meaningful in probability terms and aggregation a possibility. With Markovian preferences, features such as autocorrelation, transfer function, income multiplier and discounting receive a new interpretation as potential operators. The rationale of price as a random walk in time and space and the diffusion of price change can be provided in the same terms.

Beyond this, we leave behind geographical arrangement as efficient allocation and mechanistic change and encounter externalities in the wide sense. The search for and dissemination of information in the making of a market while related to uncertainty and risk can be conveniently handled now. The establishment of tastes via the individual can be contrasted with group selection. Spatial behaviour whether in competitive processes in migration or in exploiting the environment is basic here. Then we come to the case where individuals are partly a function of their environment and only global probabilities have meaning. Finally there is the development of the economy to be studied rather than components of change but by then an integrated viewpoint should be established.

REFERENCES

Allen, C. L., 1962, *Elementary Mathematics of Price Theory*, Wadsworth Pub., Co., Belmont, Calif.

Allen, R. G. D., 1932, The Foundations of a Mathematical Theory of Exchange, *Economica*, 12, 197–226.

Alonso, W., 1964, *Location and Land Use*, Harvard U. P., Cambridge.

Beckenbach, E. F., 1956, Conformal Mapping Methods, in E. F. Beckenbach (Ed.), *Modern Mathematics for the Engineer*, McGraw Hill, N. Y.

Beckmann, M. J., 1971, Market Share, Distance and Potential, *Regional and Urban Economics*, 1, 1, 3–18.

Berge, C. and A. Ghouila-Houri, 1965. *Programming, Games and Transportation Networks*, Eng. trans., John Wiley & Sons, N. Y.

Bish, R. L. and H. O. Nourse, 1975, *Urban Economics and Policy Analysis*, McGraw Hill, N.Y.

Buchanan, J. M., and W. C. Stubblebine, 1968, Externality, in W. Brent and H. M. Hochman (Eds.), *Readings in Microeconomics*; Holt, Rinehart and Winston, N. Y.

Denman, D. R. and S. Prodano, 1972, *Land Use: An Introduction to Proprietary Land Use Analysis*, George Allen & Unwin, London.

Edgeworth, F. Y., 1881, *Mathematical Psychics*, Reprinted A. M. Kelley, N. Y., 1967.

Evans, A. W., 1973, *The Economics of Residential Location*, MacMillan, London.

Fisher, I., 1892, *Mathematical Investigations in the Theory of Value and Prices*, reprinted Yale U. P., New Haven, 1925.

Fitch, L. C., 1972, Scale and effectiveness of urban government: a commentary, in R. M. Bird and J. G. Head (Eds.), *Modern Fiscal Issues*, Univ. of Toronto Press.

Haavelmo, T., 1970, Some Observations on Welfare and Economic Growth, in W. A. Eltis et

al. (Eds.), *Induction, Growth and Trade*, Clarendon Press, Oxford.
Harris, B., 1972, Externalities and Urban Decision-Making, in M. Perlman et al. (Eds.), *Spatial, Regional and Population Economics*, Gordon and Breach, N. Y.
Henn, R., 1961, Mathematical Appendix to Johr, W. A., Full Employment and Monetary Stability: An Application of Indifference Curves to the Solution of Economic Policy Problems, in Hegeland, H. (Ed.), *Money, Growth and Methodology*, C. W. K. Gleerup, Lund.
Hum, D., 1972, *Models of Optimum Currency Areas*; *Critique and Synthesis Employing Potential Theory Concepts*, Unpub. Ph. D. thesis, Univ. of Toronto.
Kantorovich, L. V., 1965, *The Best Use of Economic Resources*, Harvard U. P., Cambridge, Mass.
Leibenstein, H., 1968, Bandwagon, Snob and Veblen Effects in the Theory of Consumer's Demand, in W. Brent and H. M. Hochman (Eds.), *Readings in Microeconomics*, Holt, Rinehart, & Winston, N. Y.
Lerner, A. P., 1934, The Diagrammatical Representation of Demand Conditions in International Trade, *Economica*, 1, 319–34.
Mishan, E. J., 1969, *Welfare Economics*: *An Assessment*, North-Holland Pub. Co, Amsterdam.
Moran, P., 1966, *L'Analyse Spatiale en Science Economique*, Editions Cujas, Paris.
Phlips, L., 1974, *Applied Consumption Analysis*, Amsterdam.
Purcell, E. M., 1965, *Electricity and Magnetism*, Berkeley Physics Course 2, Educational Services Inc.
Rushton, G., 1969a, The Scaling of Locational Preferences, in K. R. Cox and R. G. Golledge (Eds.), *Behavioural Problems in Geography*: *A Symposium*, Northwestern University Studies in Geography, 17, Evanston.
Rushton, G., 1969b, Analysis of Spatial Behaviour by Revealed Space Preference, *Annals*, *A. A. G.*, 59, 391–400.
Shackle, G. L. S., 1972, *Epistemics and Economics*, Cambridge U. P.
Sheppard, E. S., 1976, *Interaction Feedback Modelling*: *Explorations into Configurations*, *Flows*, *and the Dynamics of Spatial Systems*, Unpublished Ph. D. thesis, Univ. of Toronto.
Shibata, H., 1972, Joint Production, Externality and Public Goods, in R. M. Bird and J. G. Head (Eds.), *Modern Fiscal Issues*, Univ. of Toronto Press.
Smith, T. E., 1974, Spatial Discounting Theory, Discussion Paper, *Regional Science Research Institute*.
Taylor, M., 1973, The Problem of Salience in the Theory of Collective Decision-Making, in H. R. Alker et al. (Eds.), *Mathematical Approaches to Politics*, Elsevier, Amsterdam.
Tiebout, C. M., 1956, A pure theory of local expenditures, *Journal of Political Economy*, 64, 416–24.
Tiebout, C. M., 1961, An Economic Theory of Fiscal Decentralization, in *Public Finance*: *Needs, Sources and Utilization* (New York: National Bureau of Economic Research), pp. 79–96.
van der Pol, B., 1959, The Finite Difference Analogy of the Periodic Wave Equation and the Potential Equation. Appendix 6 in Kac M., *Probability and Related Topics in Physical Sciences*, Interscience Pubs., London.
Van Praag, B. M. S., 1968, *Individual Welfare Functions and Consumer Behaviour*, North-Holland Pub. Co., Amsterdam.

CONTENTS OF VOLUMES

GEOGRAPHIA POLONICA

Vol. 1. 11 papers devoted to the present status of geography in Poland and 3 papers giving the results of research. List of Polish geographers, geographical institutions and geographical periodicals, 262 pp., 20 Figures, 1964 (out-of-print).

Vol. 2. 34 papers prepared by Polish geographers for the XXth International Geographical Congress in London, July 1964, 259 pp., 91 Figures, 1964.

Vol. 3. Problems of Applied Geography II. Proceedings of the Second Anglo-Polish Seminar at Keele — Great Britain, September 9–20 1962, Co-edited by the Institute of British Geographers, 24 papers by British and Polish geographers, 274 pp., 69 Figures, 1964.

Vol. 4. Methods of Economic Regionalization. Materials of the Second General Meeting of the Commission on Methods of Economic Regionalization, International Geographical Union, Jabłonna — Poland, September 9–10, 1963. Reports, communications and discussion, 200 pp., 6 Figures, 1964.

Vol. 5. Land Utilization in East-Central Europe. 17 case studies on land use in Bulgaria, Hungary, Poland and Yugoslavia, 498 pp., 104 Figures, 16 colour maps, 1965.

Vol. 6. 14 papers prepared by Polish geographers for the Seventh World Conference of INQUA in U.S.A., September 1965, 150 pp., 86 Figures, 1965.

Vol. 7. 10 papers on the geography of Poland, mostly dealing with the economic-geographical problems of Poland, 132 pp., 46 Figures, 1965.

Vol. 8. Aims of Economic Regionalization. Materials of the Third General Meeting of the Commission on Methods of Economic Regionalization IGU, London, July 23, 1964. Report and 5 papers, 68 pp., 7 Figures, 1965.

Vol. 9. Colloque de Geomorphologie des Carpathes. Materials of the geomorphological symposium held in Cracow and Bratislava, September 17–26, 1963. Report, 7 papers, 2 summaries, 118 pp., 22 Figures, 1965.

Vol. 10. Geomorphological Problems of Carpathians II. Introduction and 6 papers by Rumanian, Soviet, Polish, Hungarian and Czech geographers, 172 pp., 68 Figures, 1966.

Vol. 11. 11 papers prepared by Polish geographers dealing with the history of Polish geography, Polish studies on foreign countries and different economic-geographical questions concerning Poland, 154 pp., 36 Figures, 1967.

Vol. 12. Formation et l'Amenagement du Reseau Urbain. Proceedings of the French-Polish Seminar in urban geography. Teresin, Poland, September 20–30, 1965. Papers by French and Polish geographers, discussion, 298 pp., 51 Figures, 1967.

Vol. 13. 9 papers embracing different fields of both physical and economic geography, all of which have been devoted to methodological problems and research techniques, 130 pp., 4 Figures, 1968.

Vol. 14. Special issue for the 21st International Geographical Congress in New Delhi, 1968, 43 papers pepared by Polish geographers: 24 dealing with physical and 19 with economic and human geography, 406 pp., 80 Figures, 1968.

Vol. 15. Economic Regionalization and Numerical Methods. The volume contains the final report on the activities of the IGU Commission on Methods of Economic Regionalization, as well as a collection of 8 papers by American, Canadian, Soviet and Polish authors, 240 pp., 54 Figures, 1968.

Vol. 16. 11 papers dealing with research problems and techniques in both economic and physical geography, 136 pp., 27 Figures, 1969.

Vol. 17. Special issue prepared for the 8th Congress of the International Union for Quaternary Research, Paris, 1969, 28 papers by Polish authors, including studies in stratigraphy and neotectonics (6), geomorphology and paleohydrology (10), paleobotany (3), sedimentology (5), archeology (4), 428 pp., 122 Figures, 1969.

Vol. 18. Studies in Geographical Methods. Proceedings of the 3rd Anglo-Polish Geographical Seminar, Baranów Sandomierski, September 1–10, 1967, 260 pp., 54 Figures, 1970.

Vol. 19. Essays on Agricultural Typology and Land Utilization. 20 papers presented at the meeting of the Commission on World Agricultural Typology of the IGU, held 1968 in New Delhi, 290 pp., 97 Figures, 1970.

Vol. 20. 9 papers on various aspects of both physical and economic geography, including urbanization, international trade, changes in rural economy, industrial development, urban physiography and hydrographic mapping, 183 pp., 69 Figures, 1972.

Vol. 21. 10 papers dealing with selected problems of economic growth, transportation, cartographic methods and theory, climatology and geomorphology, 147 pp., 82 Figures, 1972.

Vol. 22. 15 papers prepared for the 22nd International Geographical Congress in Montreal, August 1972, 205 pp., 43 Figures, 1972.

Vol. 23. Present-day Geomorphological Processes. Issue prepared for the 22nd International Geographical Congress by the IGU Commission on Present-day Geomorphological Processes, 180 pp., 82 Figures, 1972.

Vol. 24. Geographical aspects of urban-rural interaction. Proceedings of the 4th Anglo-Polish Geographical Seminar, Nottingham, September 6–12, 1970, 256 pp., 76 Figures, 1972.

Vol. 25. Perspectives on spatial analysis. 7 papers presented at the meeting of the Commission on Quantitative Methods of the IGU, held 1970 in Poznań, Poland, 140 pp., 51 Figures, 1973.
Vol. 26. Scientific results of the Polish geographical expedition to Vatnajökull (Iceland), 311 pp., 253 Figures, 1973.
Vol. 27. 20 papers presented by the Soviet and Polish Geographers at the First Polish-Soviet Geographical Semminar in Warsaw, Szymbark and Cracow, 22nd May to 1st June 1971, 189 pp., 25 Figures, 1973.
Vol. 28. 9 papers embracing different fields of geography. 144 pp., 36 Figures, 1974.
Vol. 29. Modernisation des Campagnes, Actes du IV Colloque Franco-Polonais de Geographie. Augustów, Pologne, Septembre, 1973, 444 pp., 145 Figures, 1974.
Vol. 30. Proceedings of the second Polish-GDR Semminar, Szymbark (Poland), April, 1972, 151 pp., 21 Figures. 1975.
Vol. 31. 11 papers embracing different fields of geography, 235 pp., 61 Figures, 1975.
Vol. 32. 12 papers dealing with physical development of Poland, 143 pp., 22 Figures, 1975. Index to „Geographia Polonica", vols 1–32.
Vol. 33. 26 papers prepared for the 23rd International Geographical Congress in Moscow, July, 28–August, 3, 1976, part 1: Physical geography, 141 pp., 63 Figures, part 2: Economic geography, 197 pp., 49 Figures, 1976.
Vol. 34. Approaches to the study of man-environment interactions. Proceedings of the 5th Anglo-Polish Geographical Seminar, Toruń, September 1974, 290 pp., 99 Figures, 1976.
Vol. 35. Proceedings of the Polish-Soviet Seminar on contemporary problems of developing countries, Warsaw, November 1973, 148 pp., 11 Figures, 1977.
Vol. 36. The collection of studies presented to Professor S. Leszczycki in commemoration of the 50th anniversary of his scientific activity, 237 pp., 27 Figures, 1977.
Vol. 37. Urbanization and settlement. Proceedings of the Second Soviet-Polish Geographical Seminar, Moscow and Leningrad, June 1974, 224 pp., 22 Figures, 1977.
Vol. 38. Rural Landscape and Settlement Evolution in Europe. Proceedings of the Conference, Warsaw, September 1975, 304 pp., 72 Figures, 1978.
Vol. 39. Urbanization and Settlement Systems, 240 pp., 51 Figures, 1978.
Vol. 40. Agricultural Typology. Proceedings of the 8th Meeting of the Commission on Agricultural Typology, International Geographical Union, Odessa, 20–26 July 1976, 260 pp., 78 Figures, 1979.
Vol. 41. Human Impact on the Physico-geographical Processes. Proceedings of the Second Hungarian-Polish Geographical Seminar, Budapest, September 1975, 87 pp., 34 Figures, 1978.

Subscription orders for the GEOGRAPHIA POLONICA should be placed with
FOREIGN TRADE ENTERPRISE ARS POLONA — RUCH

Warszawa, Krakowskie Przedmieście 7, Poland
Cables, ARSPOLONA, Warszawa

Cena zł 50.—

ISBN 83-01-00407-x
ISSN 0010-7282