MODELLING INTERDEPENDENCIES IN HIERARCHICAL SETTLEMENT SYSTEMS

E. Sheppard

March 1983
WP-83-31

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

Many large urban agglomerations in the developed countries are either experiencing population decline or are growing at rates lower than those of middle-sized and small settlements. This trend is in direct contrast to the one for large cities in the less developed world, which are growing rapidly. Urban contraction and decline is generating fiscal pressures and fueling interregional conflicts in the developed nations; explosive city growth in the less developed world is creating problems of urban absorption. These developments call for the reformulation of urban policies based on an improved understanding of the dynamics that have produced the current patterns.

During the period 1979-1982, the former Human Settlements and Services Area examined patterns of human settlement transformation as part of the research efforts of two tasks: the Urban Change Task and the Population, Resources, and Growth Task. This paper was written as part of that research activity. Its publication was delayed, and it is therefore being issued now a few months after the dissolution of the HSS Area.

Andrei Rogers
former Chairman
of the Human Settlements
and Services Area

ABSTRACT

This paper discusses some fundamental difficulties faced by researchers attempting to model hierarchical settlement systems. Particular attention is paid to the problem of relating the effects of city size and of the regional location on growth prospects for a city. It is argued that the central issue here is a need to relate a multiregional specification of change to the hierarchical, overlapping regions that are typical of an urban system and reflect its city size distribution. A typology is provided of methods that convert interactions between arbitrarily defined regions into interactions between more meaningful functional, urban centered, regions. This is then used in an exercise that demonstrates how a conventional multiregional economic model may be restructured to allow use of a hierarchical set of functional regions, in such a way that regional economic theory may be used to ask questions about the effect of city size and regional location on urban phenomena.

CONTENTS

MODELLING INTERDEPENDENCIES IN HIERARCHICAL SETTLEMENT SYSTEMS

INTRODUCTION

The modeling of change in an urban system has been tackled at two broad levels. On the one hand, the relation of urban growth to city size has received extensive attention, testified to by the extent of literature on the benefits of city size and on the dynamics of city size distributions (see the reviews by Richardson, 1973; Carroll, 1982; Sheppard, 1982). The common criticism of this literature, voiced also in these reviews, is that the interdependencies between cities are ignored in such discussions. The second level represents an attempt to model a complete urban system with all the associated inter-urban interdependencies (see the plea by Simmons and Bourne, 1981). This literature, which tends to draw heavily on the methodology and theories developed for modeling systems of regions, can in turn be criticized for not taking into account the hierarchical nature of urban systems. In multi-regional models, all regions are essentially alloted the same importance, but in an urban system the extent of influence, and thus the importance of large metropolitan areas means that they should be treated differently from small cities and towns. This has resulted in the evolution of a distinctive sub-category of urban systems models that concentrate almost exclusively on the hierarchical nature of inter-urban interdependencies, and bear little apparent relation to multiregional approaches (Pred, 1971; Hudson, 1972; Bassett and Haggett, 1971; Pigozzi, 1980; Weissbrod, 1976).

The real situation clearly is some mixture of city size elements on the one hand, and the relative location and interdependency of cities on the other hand. Just as the growth, and thus the size of cities depends on links with the other cities, so it is also the case that the nature of these links depends on the size and sphere of influence of the various cities. An example of the importance of this issue is that the growth rate of cities of a given size in the United States depends on the region they are located in. Thus while the major metropoli of the northeast are declining, those of the Southwest continue to expand (Berry and Dahmann, 1977). The regions in which these cities are located are in turn an aggregation of cities and their rural hinterlands, and it is the prosperity of these clusters of cities that influences individual metropoli.

While the conclusion that city size and patterns of interurban interdependence influence one another is hardly surprising, researchers working on the urban system (and on other strongly hierarchical systems) face a particular problem. This may be posed as a question: how can inter-urban interdependencies and the hierarchical nature of urban systems be simultaneously taken into account? The purpose of this paper is to provide some steps on the way to answering this question. Section 1 motivates the discussion by illustrating the problems faced by urban system modelers. Section 2 presents some rather abstract ways of attempting to resolve this problem, while Section 3 illustrates how this type of solution can be applied to introducing hierarchical elements into a model of interurban flows and prices.

1. THE PROBLEMS OF URBAN SYSTEM HIERARCHIES

Hierarchical systems come in many forms. Some are capable of straightforward treatment, such as hierarchies with a strict top-down structure. But urban and regional systems are significantly more complicated than this (Rietveld, 1981b). In these systems, flows can go up and down between hierarchical levels, and also between cities at the same level. This is a basic principle stemming from central place theory (Christaller, 1933). But perhaps more importantly than this, there are no clear boundaries between branches of the hierarchical tree that can (see Pred, 1971) be used to represent an urban system. This makes it particularly difficult to identify the separate elements of an urban system.

The difficulty can perhaps be illustrated by an analogy. Biological systems are also strongly hierarchical, but at certain levels there are distinct individuals that can be separated from one another. Thus cells and individual plants and animals can be isolated physically from one another. As a result the influence of cells upon one another, and of cells on some larger level of aggregation such as an animal, can be modeled at least in principle by identifying the individual cells and then summing up the influence of each cell in turn to derive some aggregate effect.

Unfortunately this is much more difficult with urban systems. If an entire urban system is split into individual functional entities, in parallel to the cells or individuals of biology, it is generally agreed that these functional units can be well represented by a city together with its rural hinterland (cf. Kawashima and Korcelli, 1982). However, urban hinterlands overlap in two rather complicated ways. First of all, cities from high up in the urban hierarchy have hinterlands encompassing the hinterlands of many smaller cities, as is to be expected due to the hierarchical structure. But, at every level, there is no complete identification of each lower order hinterland to a single higher order hinterland. Thus, for example, two large cities can simultaneously have direct contact with, and influence on, a smaller city. Second, hinterlands defined around cities of the same level in an urban hierarchy overlap; rural areas along simultaneously to two separate hinterlands at the same hierarchical level.

This overlap makes it difficult to identify the units of analysis of an urban system. The nature of this problem is identical to that faced by multi-regional modelers. In order to meaningfully forecast regional change, the individual regions must first be identified. If this is not well done, then the regions between which flows are modeled may not represent functional clusters of activities, and any attempt to treat a region as such a cluster and to forecast its prosperity may well be unsuccessful. For this reason, regional modelers have turned to functional regions as their units of analysis. However, often such regions are only identified at one hierarchical level. This has had two effects. First, such regions are not useful for hierarchical models as mentioned in the introduction. Second the regions are non-overlapping, forcing sub-regions that belong to two larger regions to only be included in one of them. Any attempt to allow for a hierarchical structure is achieved by aggregating lower order regions into non-overlapping higher order regions (cf. Harris, 1980). But, again, this is a very severe way of representing the rather ambiguous manner in which urban-centered regions do divide up the national space of an economy.

A regionalization which excludes the possibility of overlaps between regions poses special problems for a model of inter-regional interdependencies. Two non-overlapping regions A and B are proposed, and then intra-regional interactions $A \rightarrow A$, $B \rightarrow B$ are identified and separated from inter-regional interactions $A \leftrightarrow B$. But if in fact the two regions overlap, then some flows $A \rightarrow A$ should in fact be classified as $A \rightarrow B$, and vice versa. These areas of ambiguity occur in those subregions of A and B that in fact represent a zone of overlap where both A and B simultaneously exert a direct influence. If flows are misclassified in this way, due to the enforced misclassification of the overlapping region, then a confusion is introduced similar in effect to that of improperly constructing regions in the first place.

The sophistication of this argument, however, must be confronted with the fact that data is collected for non-overlapping regions, for compelling administrative reasons. Thus at an operational level researchers are forced to use such regions. But the possibility exists of *ex post* adjustments to these regions, and to the flows between them, in such a way that the misclassified flows are more correctly classified. Then inter- and intra-regional interdependencies may be more adequately separated from one another, perhaps leading to better forecasts. Such corrections could also simultaneously take into account the overlaps of higher order regions. If such adjustments could successfully be made, then one important implication would be that the models, and extensive experience, developed for multi-regional analysis could be applied to modeling an urban system in a way that accounts for its real hierarchical nature.

The following section proposes some ways of making such *ex post* adjustments in a hierarchical urban system. The results presented here are complex and do not at this stage have the elegance necessary if they are to be practically useful. However, it is hoped that if the approach taken is sound then future work may lead to practical proposals, which would at least allow an estimate of the size of the misspecification error introduced by not accounting for the overlapping and hierarchical nature of urban centered regions.

2. INTERREGIONAL AND INTERURBAN INTERACTIONS: A TYPOLOGY

2.1. Interactions Amongst Well-Defined Regions

Consider the (artificial) case where a nation is divided into a non-overlapping set of urban-centered regions $\tilde{A},\tilde{B}=1,\ldots,R$, with metropoli A, B,...R. Let $p_{\tilde{A}\tilde{B}}$ be the probability that an interaction (of commodities or people), starting from some part of region $\tilde{A}$, flows directly to some part of region $\tilde{B}$. Assuming (reasonably) that the nation is an open system, let o represent the outside world, and assume for all $\tilde{A}$: $p_{\tilde{A}o} \geq 0$ with the inequality holding in at least one case. Then the R x R matrix of interactions, P is transitive. As a consequence, the matrix:

$$U = (I - P)^{-1} \qquad (1)$$

is finite and contains elements u_{ij} representing the probability that a unit of commodity shipped, or a person migrating, will ever reach j from i. U is a matrix of total influences or "geographical potentials", which in turn is related to the potential function of a Markov process (Seneta 1981; Sheppard 1979).

If functional regions represent the most meaningful units for analyzing spatial demoeconomics then this matrix P contains flows that can be identified with meaningful origins and destinations. The flows themselves are then more likely to be meaningful.

The regions used in the above analysis are themselves aggregates that are internally heterogeneous. However this heterogeneity is not random but may itself be susceptible to subdivision into functional subregions. Such a division of aggregate entities into disaggregate but still meaningful entities is simply a procedure of replacing a smaller group of loosely knit but heterogeneous functional regions by a larger number of less heterogeneous regions. We do not seek to maximize homogeneity in our groups, but rather to maximize the functional unity of each member.

If our regions, $\tilde{A}$, can each in turn be divided into a set of completely exclusive and mutually exhaustive functional subregions; a,b = 1,...,M, such that each subregion is within only one region, then we have a <u>well-defined</u> set of regions (Figure 1). Relating aggregate and disaggregate flows is then relatively simple, since:

$$p_{\tilde{A}\tilde{B}} = \sum_{a \in \tilde{A}} \hat{p}_a \cdot \sum_{b \in \tilde{B}} p^*_{ab} + \hat{p}_A \cdot p^*_{AB} \qquad (2)$$

where $\hat{p}_a$ is the probability that a randomly selected trip starting in region $\tilde{A}$ will originate in a, and p^*_{ab} is the probability that a trip from center a will terminate in center b. We distinguish here the flow from center A to center B simply to emphasize that it must be included. Indeed, if P^* is the matrix of direct interactions between subregions (including the subregions centered on the regional centers, A, B, etc. as in Figure 1):

$$P = G.W.P^*.G' \qquad (3)$$

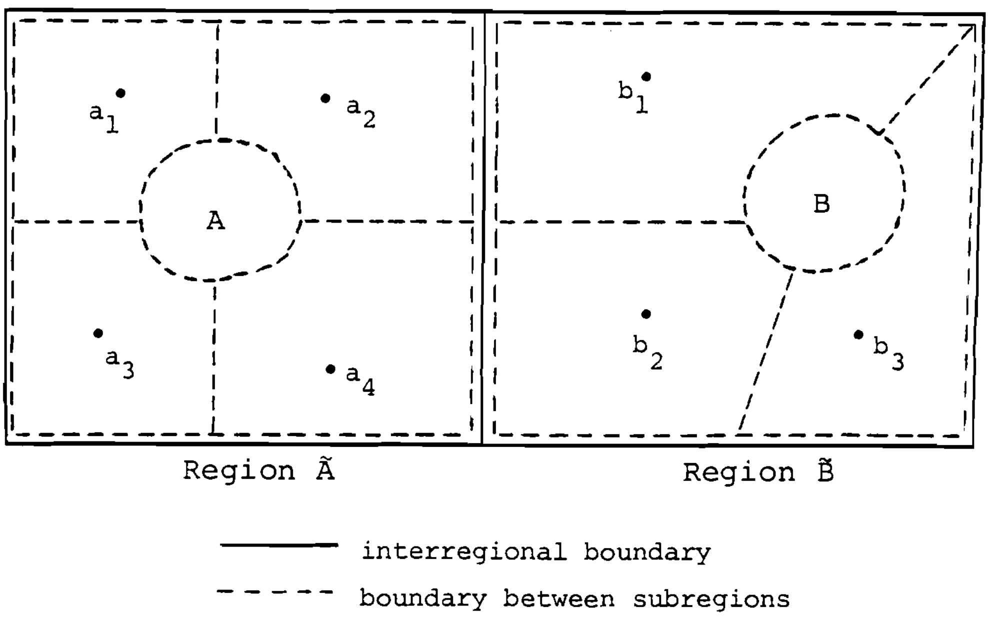


Figure 1. Well-defined regions.

where P^* is the M by M matrix of $\{p^*_{ab}\}$, G is an K by M aggregation matrix; with its rows indexed by regions and its columns by subregions. Entry g_{ij} is one if subregion j is a member of region i; zero otherwise. Note that $G'G = I$; $G.G' = H$; a diagonal matrix with h_{ii} equal to the number of subregions in region i. Finally, W is a diagonal weighting matrix with entry $w_{ii} = \hat{p}_i$. If P^*, W and G are known, P can immediately be derived. If P, W and G are known, P^* can be estimated in a "least biased" way (Sheppard 1975; Snickars and Weibull 1977):

$$\text{Max} - \sum_a \sum_b p^*_{ab} \log p^*_{ab} \tag{4}$$

subject to equation (2) and non-negativity conditions on p^*_{ab}.

Knowledge of W implies possessing some theory or data that provides knowledge about the propensity to make trips. This is a complex issue intimately linked with questions of accessibility that cannot be pursued here (cf. Sheppard 1980).

The relation between potential matrices U, and U^*, where U^* is the M by M matrix of potentials between all subregions:

$$U^* = (I - P^*)^{-1}, \tag{5}$$

can also be specified. From (1); (3):

$$U^{-1} = I - GWP^*G'$$

whence

$$G'U^{-1}G = I - WP^*$$

$$W^{-1}G'U^{-1}G = W^{-1} - P^*$$

$$I + W^{-1}G'U^{-1}G = I - P^* + W^{-1}$$

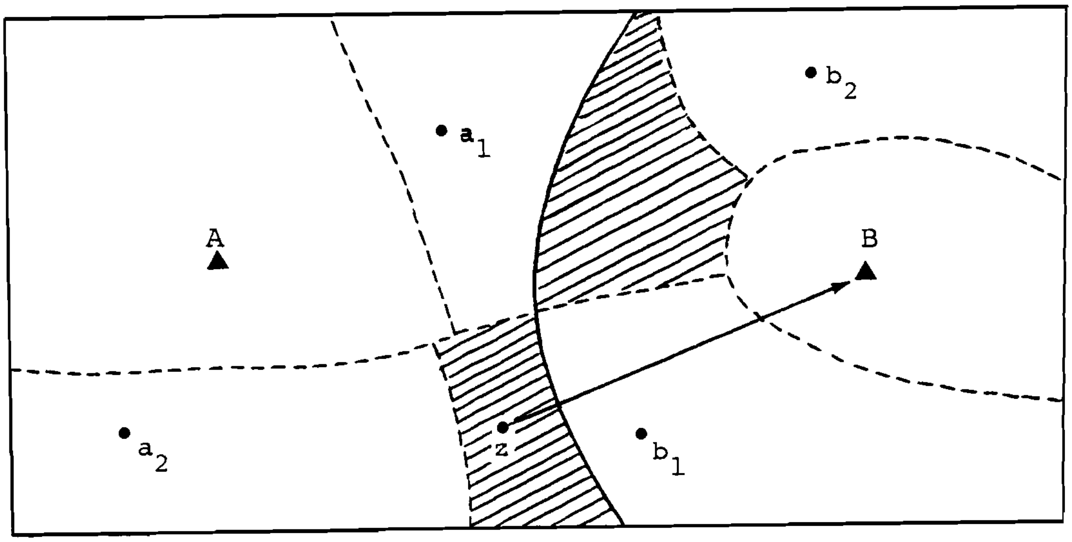
or

$$U^{*-1} = I - W^{-1}(I - G'U^{-1}G) \qquad (6)$$

2.2. Interactions Amongst Overlapping Regions

It is typically the case that we cannot construct well-defined aggregations of functional urban regions. There are four sources of indeterminancy. First, the functional territories surrounding lower order cities have a spatial extent that does not coincide with that of the areas of influence of the higher order centers (Figure 2). Second, it is impossible to draw precise boundaries between functional regions because they overlap. Third, different types of interaction will fall to low levels at different distances from any city; and fourth, areas may be erroneously classified into the wrong region. All of these sources of error imply that regional boundaries are fuzzy (Gale and Atkinson 1979). Thus an observed flow from $\tilde{A}$ to $\tilde{B}$ may in fact be more appropriately classed as an internal flow within the functional region of $\tilde{B}$.

As an example of the first source of error, consider Figure 2. Functional regions are typically defined in terms of the level of interaction between areal units and a city identified as the core of the region. However, location z in Figure 2 may have stronger direct contacts with A, than with B, thus leading to it being classified as part of region A, whereas indirect contacts $z \rightarrow b_1 \rightarrow B$ may be stronger still. If this is the case it is at least partially erroneous to represent z as a member of the functional region $\tilde{A}$. Concrete examples would be that an individual from z shops more at A than at B, but he/she obtains even more goods by placing orders in town b_1 with local merchants who buy from city B. Similarly a person in z may become unemployed due to layoffs in town b_1 responding to economic conditions at B, rather than due to conditions in A. In short indirect interactions may be more powerful than direct interactions, and, particularly on the



interregional boundary

boundary between subregions

Figure 2. Hierarchically overlapping urban centered regions.

fringe of functional regions, may operate in the opposite direction. If only direct interactions are considered the result is an identification error classifying some intraregional flows ($z \rightarrow b_1$) as if they were interregional, and vice-versa. The challenge of reconstructing meaningful flows is pursued in the next subsection.

2.3. Reconstructing Functionally Meaningful Interactions.

2.3.1. Observed Interactions

It has been traditional in regionalization problems to classify regional membership on the basis of allocating lower order urban centers to that regional center with which they have the highest direct interaction. This approach was pioneered by Nystuen and Dacey (1961) and has been used *inter alia* by Simmons (1974). Let us alternatively assume, for simplicity, that subregions whose centers are in a particular region should ideally be treated as part of that functional region. Thus suppose that the areas shaded in Figure 2 have been mis-classified as a result of considering only direct interactions with centers A and B. Also assume that all individuals in a subregion can be treated as responding to the same socioeconomic environment. Then the flows between regions must be modified to take into account the proportion of a subregion that is assigned to each larger region. Define this by a non-binary fuzzy membership function $p_a(\tilde{A})$ representing the possibility that a randomly selected individual from a̲ is in fact starting a trip from a location that is within $\tilde{A}$, where:

$$0 \leq p_a(\tilde{A}) \leq 1 \tag{7}$$

$p_a(\tilde{A})$ could, for example equal the proportion of the subregional population of a̲ residing in $\tilde{A}$. The probability that a move from a to b is in fact a move from $\tilde{A}$ to $\tilde{B}$ $[p^*_{ab}(\tilde{A}\tilde{B})]$, which we may call a fuzzy interaction, is:

$$p^*_{ab}(\tilde{A}\tilde{B}) = p^*_{ab}\, p_a(\tilde{A})\, p_b(\tilde{B}) \tag{8}$$

Then the total probability that a randomly selected trip in the system occurs from $\tilde{A}$ to $\tilde{B}$ is:

$$p_{\tilde{A}\tilde{B}} = \sum_a \hat{p}_a \sum_b p^*_{ab}(\tilde{A}\tilde{B}) \tag{9}$$

$$= \sum_a p_a(\tilde{A})\hat{p}_a \sum_b p_b(\tilde{B})p^*_{ab} \tag{10}$$

Or, generalizing (3):

$$P = \underline{P(G)}\, WP^* \underline{P(G)}' \tag{11}$$

where P(G) is an R by M matrix with i,j-th entry equal to $p_j(I)$, the probability that an individual from subregion j lives in region I. P(G) is thus a fuzzy generalization of G. P would then represent the matrix of *observed* interregional flows.

The relation between U^* and U is now more complex because P(G)'P(G) is not an identity matrix;

$$U^{*-1} = I - P^* = I - W^{-1}G^*[I - P(G)'U^{-1}P(G)G^*] \quad (12)$$

with $G^* = [P(G)'P(G)]^{-1}$.

2.3.2. Adjusted Interactions

Let us assume for simplicity that each subregional center is dominated by only one regional center. We might term this binary hierarchical dominance as "Christallerian" (Figure 3), where a hierarchical relation is taken to exist whenever there is a direct interaction between two locations.

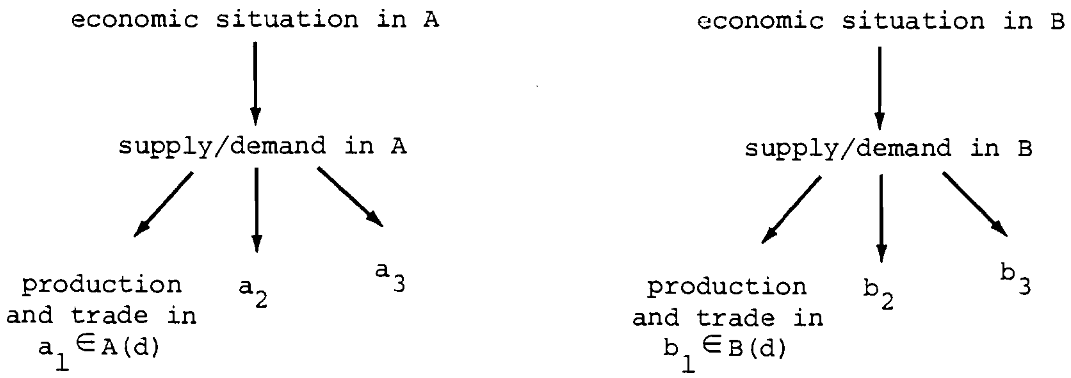


Figure 3. Christallerian hierarchical dominance.

$a_i \in A(d)$ is a statement that the center of subregion a_i is dominated by the (higher order) regional center of region $\tilde{A}$. Since this regional center cannot interact directly with subregions dominated by another regional center by assumption, interactions from regional center A to b_i, or from a_i to B, are excluded. Therefore the total interregional interactions represent direct flows from center A to center B, or from subcenters a_i to subcenters b_i. Thus no inter-regional flows occur from one level of the hierarchy to another. Formally:

$$\tilde{p}_{\tilde{A}\tilde{B}} = \sum_{a \in \tilde{A}(d)} \hat{p}_a \sum_{b \in \tilde{B}(d)} p^*_{ab} + \hat{p}_A \, p^*_{AB} \tag{13}$$

where p^*_{AB} is the probability that a trip from center A will travel directly to center B, and $\tilde{p}_{\tilde{A}\tilde{B}}$ is the probability that a trip is made directly from some point in the region dominated by A to some point in the region dominated by B.

Our Christallerian hierarchical structure implies:

$$p^*_{aB} = p^*_{bA} = p^*_{Ab} = p^*_{Ba} = 0$$

Notice a subtle but vital difference between equation (13) and equation (2). In (13) set membership is defined by functional dominance; in (2) it is given by the (well-defined) regionalization.

We wish to convert observed aggregate interactions $p_{\tilde{A}\tilde{B}}$ into functionally based interactions $\tilde{p}_{\tilde{A}\tilde{B}}$. The assumption underlying this aggregation is that a better specified set of relationships will be derived if areas indirectly dependent on region $\tilde{B}$, but classified as being located in region $\tilde{A}$ were to be re-identified with region $\tilde{B}$. The interactions observed across the given $\tilde{A}/\tilde{B}$ boundary must then be adjusted to allow for this reclassification; i.e., flows from regions dominated by B must be classified as being flows from $\tilde{B}$.

Defining $\tilde{P}$ as the matrix of adjusted interactions, $\tilde{p}_{\tilde{A}\tilde{B}}$:

$$\tilde{P} = G_d W P^* G_d'$$

$$G_d' \tilde{P} G_d = WP^*$$

whence, from (11)

$$P = \underline{P(G)} G_d' \tilde{P} G_d \underline{P(G)'} \qquad (14)$$

with the R by M matrix of dominance relations, G_d, having i,j-th element equal to one if subregion j is dominated by region i and zero otherwise. Once again $G_d' G_d = I$.

Equation (14) shows how an observed aggregate interregional interaction matrix P can be converted into a more meaningful matrix of flows between functionally defined regions, $\tilde{P}$. The necessary information for this procedure are estimates of (i) the probability $P_a(\tilde{B})$ that an individual in some functionally defined subregion a is actually residing within the aggregate functional region $\tilde{B}$, for all a = 1,...,M; $\tilde{B}$ = 1,...,R; and (ii) the regional center which hierarchically dominates each subregion. Then matrices G_d and P(G) can be constructed, and our functionally meaningful interaction matrix is:

$$\tilde{P} = [G_d P(G)']^{-1} \; P \; [G_d P(G)']^{-1} \qquad (15)$$

assuming the inverses exist. It is most important to note that no knowledge of the disaggregate flows is necessary; for purposes of altering P to $\tilde{P}$, estimations described in equation (4) can be bypassed.

2.3.3. Adjusted Interregional Flows with Fuzzy Hierarchical Relations

The Christallerian assumption that each subregion is dominated by only one regional center is clearly a simplification; we may generalize this by recognizing (with Pred, 1971, and others) that each subregion may be dominated to different degrees by several regional centers, making the set relation of dominance fuzzy. This case has been extensively treated by Ponsard (1977). He argues that many types of interactions occur between urban centered regions of various hierarchical positions. Define f_{ij}^k to be the direct interaction of type k between urban centers i and j. We normalize these as follows

$$g_{ij}^k = \begin{cases} f_{ij}^k \,/\, \max\limits_i \sum\limits_j f_{ij} & \text{if flow k is determined by supply considerations} \\ f_{ij}^k \,/\, \max\limits_j \sum\limits_i f_{ij} & \text{if flow k is determined by demand considerations} \end{cases}$$

Thus once the k-th type of interaction is defined as primarily supply or demand determined, the index g_{ij}^k measures the dominance of i over j (for given k) relative to the total dominance exerted by the most influential urban center. For each inter-urban link (i,j), we have a range of values of g_{ij}^k (one for each k), and we can construct a "fuzzy" matrix F; with i,j-th entry $\mu_{ij} = [\min\limits_k g_{ij}^k, \max\limits_k g_{ij}^k]$. Each entry in F is thus a double entry giving maximum and minimum values for the level of dominance of i over j. These values can be ranked and manipulated consistently (Ponsard 1977). Even if these ranges are reduced to one number (i.e., $g_{ij}^k = g_{ij}$ for all k) it still has a "fuzzy" interpretation as the degree of existence of dominance of i over j (the level of possibility that i has of dominating j).

Once interregional interactions between centers at different hierarchical levels are included (i.e., p_{Ab}, $p_{aB} \geq 0$), it no longer is possible to say that any one subregion is uniquely associated with any one region. Indeed any interaction observed from a_i to b_j represents an influence that only partly originates in region $\tilde{A}$. Total interaction from some region $\tilde{B}$ directly to subregion a_i is given by:

$$I_{\tilde{B}a_i} = \hat{p}_{\tilde{B}}\left[\sum_b \hat{p}_b\ p^*_{ba_i} + \hat{p}_B\ p^*_{Ba_i}\right] \qquad (16)$$

where $\hat{p}_{\tilde{B}}$ is the probability that a trip will be generated from region $\tilde{B}$ during a fixed time period. Define the level of direct dominance of region $\tilde{A}$ over some subregion a_i as $D_{a_i}(\tilde{A})$:

$$\begin{aligned} D_{a_i}(\tilde{A}) &= 1 - \left(\sum_{\tilde{B}\neq\tilde{A}} I_{\tilde{B}a}\right) \Big/ \left(\sum_{\tilde{B}} I_{\tilde{B}a}\right) \\ &= 1 - \sum_{\tilde{B}} D_{a_i}(\tilde{B}) \end{aligned} \qquad (17)$$

where

$$D_{a_i}(\tilde{B}) = I_{\tilde{B}a_i} \Big/ \left(\sum_{\tilde{B}} I_{\tilde{B}a_i}\right) \qquad (18)$$

Then the interaction from a_i to b_j that originates directly from region $\tilde{A}$ is given by $D_{a_i}(\tilde{A})p^*_{a_ib_j}$, and total interaction from region $\tilde{A}$ to region $\tilde{B}$ that directly originates within region $\tilde{A}$ is:

$$\tilde{p}_{\tilde{A}\tilde{B}} = \sum_{a\in\tilde{A}(d)} D_a(\tilde{A})\ \hat{p}_a \sum_{b\in\tilde{B}(d)} p^*_{ab} \qquad (19)$$

Define an R by M fuzzy dominance matrix, D, with i,j-th entry equal to $D_j(i)$, if i,j are members of the same functional region, zero otherwise. Then from equation (19):

$$\tilde{P} = DWP^{*}G_d' \tag{20}$$

Note that the i,j-th element of D is zero if the i,j-th element of G_d is zero, thus all feasible products of D and G_d are diagonal matrices.

An assumption underlying equation (20), with the inclusion of G_d, is that interactions into some subregion b of region $\tilde{B}$ thus represent interactions affecting region $\tilde{B}$. However, by extending the above arguments, if b in turn directly interacts with subregions associated with other regions, then a part of the flow into b is exported again out of $\tilde{B}$. Thus (20) may be modified to count only those flows to b that remain within $\tilde{B}$ as follows.

Define a matrix E′, of dimension M by R, with i,j-th elements:

$$E_a(\tilde{A}) = 1 - \left(\sum_{\tilde{B}\neq\tilde{A}} I_{a\tilde{B}}\right) \Big/ \left(\sum_{\tilde{B}} I_{a\tilde{B}}\right) \tag{21}$$

$$I_{a\tilde{B}} = \sum_{b} p_{ab}^{*} + p_{aB}^{*} \tag{22}$$

where $E_i(j)$ is zero if i is a subregion not belonging to region j. Then

$$\tilde{p}_{AB} = \sum_{a\in\tilde{A}(d)} D_a(\tilde{A})\ \hat{p}_a \sum_{b\in\tilde{B}(d)} p_{ab}^{*}E_b(\tilde{B}) \tag{23}$$

or

$$\tilde{P} = DWP^{*}E' \tag{24}$$

Again E has the same structure as D and G_d.

2.3.4. Adjusted Interregional Flows with Indirect Relations

In general, particularly in models with static interaction matrices, the influence of one location on another is given by the sum of all direct and indirect interactions, or the geographic potential difference along that link (Sheppard 1979). This is already implicit in the arguments of the previous section. To conceptualize this we shall discuss in turn a first approximation and the limiting case.

As a first approximation, interactions from any subregion g to another subregion h may contain interactions that partly originate in region $\tilde{A}$, and may affect flows from g that go from h directly to $\tilde{B}$. Then there is an element of interdependence from $\tilde{A}$ to $\tilde{B}$ that exists even in a flow between g and h when g and h do not belong to $\tilde{A}$ or $\tilde{B}$. The total flows from $\tilde{A}$ to $\tilde{B}$ should incorporate this:

$$\tilde{p}_{AB} = \sum_{g=1}^{M} \sum_{h=1}^{M} D_g(\tilde{A}) p^*_{gh} E_h(\tilde{B}) \tag{25}$$

where the fuzzy relations D and E represent, respectively; the possibility that a flow from g results from a flow into g from $\tilde{A}$, and the possibility that a flow into h will in turn directly affect $\tilde{B}$. Both of these possibility relations may be non-zero for any subregion and region, defining the generalized matrices $\overline{D}$ and $\overline{E}$. Then:

$$\tilde{P} = \overline{D} W P^* \overline{E}' \tag{26}$$

It should be noted that despite the generality of the two possibility functions D and E we are retaining a precise definition of which subregions should be assigned to which regional centers. If this were not done all of the structure of the problem would be lost.

Turning attention to the limiting case, we recall that U^* contains the total direct and indirect interactions between all pairs of subregions, and that U^* is finite if P^* is transitive. Now the total influence exerted on subregion b (the probability of ever reaching b from some randomly chosen starting point in the system) may be defined as the location potential at b; the b-th column sum of U^*. It seems, however, not relevant to assume that each subregion is equally likely to interact with other regions. Therefore we should define location potential, U_b, as weighted by overall interaction propensities:

$$U_b = \sum_{g \in \tilde{G}} \sum_{\tilde{G}} \hat{p}_{\tilde{G}} \hat{p}_g u^*_{gb} \tag{27}$$

Then the possibility that any given action at b is influenced by causes emanating from some region $\tilde{A}$ is:

$$D^*_b(\tilde{A}) = U_b^{-1} \sum_{g \in A} \hat{p}_{\tilde{A}} \hat{p}_g u^*_{gb} \tag{28}$$

The possibility that an action at b influences events in region $\tilde{A}$ is:

$$E^*_b(\tilde{A}) = \sum_{g \in A} u^*_{bg} \Big/ \sum_{\substack{g \\ g \neq 0}} u^*_{bg} \tag{29}$$

and, defining the $R \times M$ matrix D^* from (28) and the R by M matrix E^* from (29):

$$\tilde{P} = D^* W P^* E^{*\prime} \tag{30}$$

Now,

$$D^* = \hat{W} G_d W Q U^* \tag{31}$$

where $\hat{W}$ is an R × R diagonal matrix, $\hat{w}_{\tilde{H}\tilde{H}} = \hat{p}_{\tilde{H}}$; and Q is a diagonal matrix, $q_{ii} = U_i^{-1}$.

$$E^{*\prime} = U^* Q^* G_d' \tag{32}$$

where Q^* is a diagonal matrix, $q_{ii}^* = \left(\sum_{\substack{g \\ g \neq 0}} u_{ig}^* \right)^{-1}$. Therefore

$$\tilde{P} = \hat{W} G_d W Q U^* P^* U^* Q^* G_d' \tag{33}$$

Recalling $U^* = (I - P^*)^{-1} = \sum_{k=0}^{\infty} P^{*k}$, then $P^* U^* = (I - P^*)^{-1} - I$, and

$$\tilde{P} = \hat{W} G_d W Q U^* (U^* - I) Q^* G_d' \tag{34}$$

Finally, $\tilde{P}$ may be related to observed aggregate interactions, P, since from equation (11) $P^* = H^{-1} \underline{P(G)}' P \underline{P(G)} H^{-1} W^{-1}$, whence $U^* = \left[I - H^{-1} \underline{P(G)}' P \underline{P(G)} H^{-1} W^{-1} \right]^{-1}$ where $H = \underline{P(G)}' \underline{P(G)}$ and therefore

$$\tilde{P} = \hat{W} G_d W Q \left[I - H^{-1} \underline{P(G)}' P \underline{P(G)} H^{-1} W^{-1} \right]^{-1} \cdot \left[\left(I - H^{-1} \underline{P(G)}' P \underline{P(G)} H^{-1} W^{-1} \right)^{-1} - I \right] Q^* G_d' \tag{35}$$

Equation (35), in all its complexity, describes how an observed matrix of interactions between regions defined as in Figure 2 may be converted to a "meaningful" matrix of interregional flows. In this case meaningful flows between regions $\tilde{A}$ and $\tilde{B}$ are the sum of all direct interregional flows weighted by the probabilities that those flows ever originated in region $\tilde{A}$ and will ever terminate in region $\tilde{B}$. Once again this conversion may be made without any knowledge of the disaggregated flows matrix. We do however require information, at the disaggregate level, of $\hat{p}_a$, $\sum_{\substack{g \in A \\ g \neq 0}} u_{ag}$, and $\sum_{\tilde{B}} \sum_{b} \hat{p}_{\tilde{B}} \hat{p}_b u_{ba}$. These three terms may be interpreted as the propensity for a subregion to generate interaction; the general accessibility of the rest of the system from that subregion, and the accessibility of that subregion from the rest of the system. It may well be possible to provide estimates of these without a knowledge of individual pairwise flows.

2.4. Summary

Assume that we have an empirical system of regions, each hierarchically decomposed into subregions that themselves do not coincide in extent with regional boundaries (Figure 2). Suppose further that the interactions between these regions have been observed, or modeled, as P. However because they are not well defined with respect to the subregions, these interactions do not represent interdependencies between functionally meaningful units. In order to convert them to more meaningful interactions, P must be converted into a theoretically more defensible matrix of interactions $\tilde{P}$. The nature of this conversion depends on how interregional links are conceptualized. However, it may be shown in each case that although (or rather, because) regional interactions P and $\tilde{P}$ depend on a larger, more disaggregated, interactions matrix P^*, it is not necessary to know subregional interactions in order to perform this conversion. We only need to know how P and P are related to P^*. The results are summarized in Table 1, where the choice of procedure is seen to depend on how interdependencies are conceptualized.

3. FROM REGIONAL ECONOMICS TO URBAN SYSTEMS MODELS

Modeling the spatial development of regions has typically involved defining a set of regions and relating them together in some way. The dominance that any one region might exert

Table 1. Converting observed into meaningful interregional interactions in a hierarchical urban system.

CONCEPTUALIZATION OF INTERDEPENDENCIES	CONVERSION EQUATION	DERIVATION	INFORMATION NECESSARY FOR CONVERSION
<u>CASE A</u> Only direct interregional interactions considered between centers at the same hierarchical level in different regions.	$\tilde{P} = [G_d P(G)']^{-1} P [G_d P(G)']^{-1}$	From equations (11), (14)	Probability that an area functionally a part of one region is mistakenly classified in another region, [P(G)]. The allocations of subregions to functional regions (G_d).
<u>CASE B</u> Interregional interactions between different levels of the hierarchy allowed. Direct interregional flows weighted by the likelihood that each subregion is directly influenced by events from within that subregion.	$\tilde{P} = \left[P(G)(D'D)^{-1}D'\right]^{-1} P [G_d P(G)']^{-1}$	From equations (11), (18), (20) (note D'D is diagonal)	As above, plus knowledge for each subregion of the proportion of direct interaction terminating there that originates in the same functional region (D).
<u>CASE C</u> As for case B, except the proportion of flows terminating in any subregion is reduced by the likelihood of that subregion immediately contacting places outside the functional region.	$\tilde{P} = \left[P(G)(D'D)^{-1}D'\right]^{-1} P \cdot \left[E(E'E)^{-1}P(G)'\right]^{-1}$	From equations (11), (21), (23) (note E'E is diagonal)	As above, plus knowledge of the likelihood that each subregion will directly interact with other places within the same functional region (E).
<u>CASE D</u> As for case C, except the flow between any pair of subregions g,h is a flow between any pair of regions A,B ($g,h \not\in A,B$), weighted by the likelihood that flows exist from A to g and from h to B.	$\tilde{P} = \left[P(G)(\bar{D}'\bar{D})^{-1}\bar{D}'\right]^{-1} P \cdot \left[\bar{E}(\bar{E}'\bar{E})^{-1}P(G)'\right]^{-1}$	From equations (11), (25), (26)	As above, plus knowledge of the likelihood of any subregion receiving direct interaction from, or sending direct interaction to, each region ($\bar{E},\bar{D}$ are expanded versions of E,D).
<u>CASE E</u> As for case D, except the probabilities of contact of A to g and h to B are given by the likelihood of any contact by direct or indirect means (geographical potentials) between these places.	$\tilde{P} = \hat{W}G_d WQ\left[I - H^{-1}P(G)'PG)H^{-1}W^{-1}\right]^{-1} \cdot \left[\left(I - H^{-1}P(G)'PP(G)H^{-1}W^{-1}\right)^{-1} - I\right]Q^{*}G_d'$	From equations (11), (30), (31), (32), (34), (35)	As <u>case B</u> plus knowledge of the likelihood that a trip in the system will originate from any subregion ($\hat{W}G\hat{W}$), and knowledge of the overall accessibility of any subregion from the rest of the system (Q) and to the rest of the system (Q^*).

over others is to be determined from an empirical analysis of the relative strength of interregional links. By contrast, models of change in urban systems have tended to impose a "top down" structure; growth impulses are seen as diffusing through the urban hierarchy moving quickly between large cities and into the functional regions of those cities (Hudson 1972). Indeed many analyses have demonstrated the existence of short period space-time lags in urban responses to impulses, reflecting this process (for a recent example see Pigozzi, 1980). However both of these conceptions are only partly correct.

The hierarchical structure, implying cities connected together dendritically, cannot allow for many other interactions that are clearly important both up and across the hierarchy (Pred 1971). Once these loops are allowed for, the responses of cities to growth impulses can be brought about by all sorts of direct and indirect transmission routes through the system and can thus occur more than once for any given city, and in a temporal order that eventually bears little relation to the original hierarchical structure. This may explain why the most successful empirical demonstrations of leads and lags are limited to responses that occur within one to three months of the initial impulse. Bennett (personal communication), for example, has suggested one month as the maximum time lag at which meaningful results can be obtained for the British urban system. It is, perhaps, only during the first pass of an impulse through the system that there can be any hope of detecting a meaningful pattern. Later on, the various spatial feedbacks will disturb any regular sequencing of responses that might be hoped for. Not only, then, are there loops and cycles in the system of inter-urban interdependencies, but there apparently must be geographical biases to these flows. If they were dependent only on the *in situ* characteristics of the cities involved then cities of the same type would be identically affected, and there would be no regional differentiation of city performance for cities of a particular type.

On the other hand, to assume no hierarchical structure to the intra-national space economy at all can also lead to difficulties. If a nation is divided into a few large regions, even when meaningfully defined as centered on the major cities, little can be said about those other cities located within these functional regions simply because the scale of analysis is inappropriate. If we divide the nation into very many very small units, the sheer size of the problem is such that in the absence of any structure imposed on the interaction patterns the number of interregional interactions to be modeled is enormous (9,625,206 in the case of counties in the United States, a scale used by Harris, 1980).

A model that allows for regional and urban aspects of demoeconomic change would ideally incorporate the advantages of both the above approaches. Interregional interdependencies may be represented as links between the cities of major functional urban regions, including those between lower order cities within those regions. Hierarchical relations will be represented by strong flows between cities of different hierarchical order but within the same branch of the hierarchy. Finally, the model itself would have a hierarchical structure allowing representation of interregional links between the fewer higher-order functional regions and the more numerous lower-order regions.

Urban system theory and plain common sense inform us that different types of interdependencies are important at different scales. If one considers wage or price formation, for example, the scale of analysis at which these are determined depends on the geographic scope of the institutions involved. Nation-wide unions and/or nation-wide corporations will set certain wages at a national scale. Examples are auto-workers' wages in the United States, miners' wages in Britain, and federal government wages everywhere. On the other hand in industries where the unions do not have nation-wide penetration (such as textiles in the U.S.) or in corporations whose operations are resricted to certain regions (regional retailing companies), a regional scale is appropriate, with actions in one region affecting those in others providing a greater geographical variation nation-wide. Finally, single enterprise companies and highly localized corporations, particularly if associated with labor organizations whose policies are locally determined, will set wages at a local scale; wages that show the greatest geographical variation and the lowest level of spatial autocorrelation. Similar arguments may be made about price formation; certain prices show a strong correlation between cities as they are set nationally by suggested retail prices (for instance, standard brand-name commodities). Others are set regionally (such as in agricultural commodity markets) or locally (personal services); with the links between locations being at best indirect leading to correspondingly less well correlated prices. This would suggest that different types of activities should be modeled at different regional scales (or equivalently at different positions within the urban hierarchy) within a nested regional structure.

3.1. A Theoretical Illustration

Consider, as an example, an economy that is nationally focused on one major (capital) city, while below this three nested lower order levels of hierarchical functional regions may be identified. We shall refer to these levels by the index $h = 1,\ldots,4$; with $h = 1$ representing the highest order functional region, encompassing the nation and centered on the capital city. Let us further suppose, as is frequently the case, that the process of regionalization by which subregions for each level are identified is a strict hierarchical classification based on direct interactions as illustrated in Figure 2. I shall attempt to show how regional and hierarchical considerations can be linked together in this context.

As an example of a regional specification consider an interregional model of price and profit determination in a capitalist economy, specified at a given point in time, where the levels of production and trade within and between regions, and the real wage, are given. Define a_{ij}^{mn} as the amount of good m, produced in region i, that is shipped to region j to produce a unit of good n there. If A is the matrix of these coefficients, π is the rate of profit (assumed equal everywhere), and $\underline{p}'$ is a row vector of all prices p_i^m of goods m in regions i (including transportation), then in competitive equilibrium with no joint production (Sheppard 1980, 1981):

$$\underline{p}' = (1 + \pi)\underline{p}'A \qquad (36)$$

If the economy is producing a surplus of commodities over demands the non-negative matrix A has a principal eigenvalue less than one which has associated with it the only eigenvector of A that is positive, by the Perron-Frobenius theorem. This eigenvalue, equal to $(1 + \pi)^{-1}$ implies a positive rate of profit and a unique price vector $\underline{p}'$ given by the associated left hand eigenvector. Hourly money wages in sector n of region j, given by the real wage weighted by prices, are:

$$w_j^n = \sum_i \sum_{m\in\Pi} a_{ij}^{mn} p_i^m T_n^{-1} + \sum_i \sum_{m\in\Pi} a_{ij}^{mn} \tau_{ij}^m p_i^t \qquad (37)$$

where Π is the set of goods consumed by workers (including transportation inputs that ship such goods), or the set of wage goods. a_{ij}^{mn} is the amount of wage good m consumed per day by a worker in industry n, τ_{ij}^m is the transportation needed to ship m from i to j, p_i^t is the transport price in i, and T_n is the length of the working day in hours. It can be shown in this system that profits are inversely related to input quantities of labor, to the real wage, and to the length of the working day.

Typically A is partitioned into a relatively small number of regions which may or may nor be arbitrarily defined. Interregional input-output models are then tied into a national econometric model in some way, with the sum of regional variables being made consistent with national aggregates. Recalling that an urban hierarchy suggests that different economic variables are determined at various geographical scales, this problem may be approached differently. Partition the set of N sectors into four groups identified by h = 1,...,4; each group representing the sectors whose prices may be regarded as being determined at the h-th hierarchical (or spatial, cf. Curry 1972) scale. Note that some sectors may have prices determined at more than one scale (an example would be goods produced by national and regional corporations), so the total number of scale specific sectors could exceed N. For simplicity we will ignore this possibility.

We then require a model with two principal features. First, the prices of different goods are determined at different scales. Second, the interactions between regions defined at hierarchical levels above the lowest level must be consistent aggregations of lower level interactions. One way to incorporate these is as follows. Assume N_h sectors in R_h functional regions have their prices determined at hierarchical level h. Then at the (lowest) level h = 4:

$$\underline{p}_4' = (1 + \pi)\left\{\underline{p}_4'A_4 + \left[\left(\underline{p}_1'\tilde{A}_1G_1 + \underline{p}_2'\tilde{A}_2\right)G_2 + p_3'\tilde{A}_3\right]G_3\right\} \qquad (38)$$

where $\underline{p}_h'$ is the (1 by N_hR_h) vector of prices in the N_h sectors and R_h regions for which prices are determined at hierarchical level h. A_h (N_hR_h by N_hR_h) is the interregional input-output matrix at level h. G_h (N_hR_h by $N_{h+1}R_{h+1}$) is a binary matrix specifying which subregions at level h+1 are dominated by each regional center at level h. The i,j-th element of G_h is one if j represents a sector in a subregion s that is dominated by the region r represented by row i. G_h thus disaggregates prices set at level h into price inputs for all subregions at level h+1. $\tilde{A}_h$ represents interregional interactions at level h that are consistent aggregations of lower level interactions. Similarly:

$$\underline{p}_3' = (1 + \pi)\ \underline{p}_3'\tilde{A}_3 + [\underline{p}_1'\tilde{A}_1G_1 + p_2'\tilde{A}_2]G_2 \qquad (39)$$

$$\underline{p}_2' = (1 + \pi)[\underline{p}_2'\tilde{A}_2 + p_1'\tilde{A}_1G_1] \tag{40}$$

$$\underline{p}_1' = (1 + \pi)\underline{p}_1'\tilde{A}_1 \tag{41}$$

Defining $\underline{p}'$ as the partitioned vector $[\underline{p}_1' \vdots \underline{p}_2' \vdots \underline{p}_3' \vdots \underline{p}_4']$:

$$\underline{p}' = (1 + \pi)\underline{p}'\tilde{A} \tag{42}$$

where

$$\tilde{A} = \left[\begin{array}{c|c|c|c} \tilde{A}_1 & \tilde{A}_1G_1 & \tilde{A}_1G_1G_2 & \tilde{A}_1G_1G_2G_3 \\ \hline 0 & \tilde{A}_2 & \tilde{A}_2G_2 & \tilde{A}_2G_2G_3 \\ \hline 0 & 0 & \tilde{A}_3 & \tilde{A}_3G_3 \\ \hline 0 & 0 & 0 & \tilde{A}_4 \end{array}\right]$$

3.2. Regionalization Issues

To complete the specification of (42) $\tilde{A}_h$ must be defined. As noted in section 2 of this paper there are a number of approaches to this which vary only in the way in which direct interactions are weighted. For this example we choose case C from Table 1, which from equation (24) implies:

$$\tilde{A}_h = D_{h,h+1}\tilde{A}_{h+1}E'_{h,h+1} \tag{43}$$

The definition of input-output coefficients implies that W from (24) is not necessary. $D_{h,h+1}$ and $E_{h,h+1}$ have essentially the analogous definition to those of equation (24): they have the i,j-th element non-zero if row i refers to some sector $m \in N_h$ in some region $r \in R_h$, and j refers to a sector $n \in N_h$ for some subregion $s \in R_{h+1}$ which is dominated by r. An exception is aggregation from the subnational to the (single region) national level. Here regional boundaries have disappeared and we have a simple aggregation problem:

$$\tilde{A}_1 = \Gamma_h \tilde{A}_h \Gamma_h' \tag{44}$$

where the i,j-th element of Γ_h is one if row i refers to some sector $n \in N_1$ in the nation, and column j refers to some sector $m \in N_1$ in some region $r \in R_h$; zero otherwise.

Three kinds of empirical cases may now be identified. First, we may possess direct observations on flows of goods $m \in N_h$ between regions $r \in R_h$ for all levels h. If so then these may be used directly to construct $\tilde{A}_1$, $\tilde{A}_2$, $\tilde{A}_3$ and A_4, and aggregations of the form (43) are not needed. Second, we may observe flows of all goods directly between all functional regions of the lowest hierarchical level: an interregional input-output matrix of dimension R_4N by R_4N, which we may define as A. If Δ_4 is a R_4N_4 by R_4N binary matrix with i,j-th element equal to one if row i represents a sector $m \in N_h$ in subregion r, and j represents the same region and subsector, then

$$A_4 = \Delta_4 \; A \; \Delta_4' \tag{45}$$

$$\tilde{A}_3 = D_{3,4} \; A \; E_{3,4}' \tag{46}$$

$$\tilde{A}_2 = D_{2,4} \; A \; E_{2,4}' \tag{47}$$

and

$$\tilde{A}_1 = \Gamma_4 \; A \; \Gamma_4' \tag{48}$$

From (42) and (45)-(48):

$$\tilde{A} = \begin{bmatrix} \Gamma_4 & & & \\ & D_{23} & & \\ & & D_{34} & \\ & & & \Delta_4 \end{bmatrix} \begin{bmatrix} A & & & \\ & A & & \\ & & A & \\ & & & A \end{bmatrix} \begin{bmatrix} \Gamma_4' & & & \\ & E_{24}' & & \\ & & E_{34}' & \\ & & & \Delta_4' \end{bmatrix} \begin{bmatrix} I & G & G_1G_2 & G_1G_2G_3 \\ & I & G_2 & G_2G_3 \\ & & I & G_3 \\ & & & I \end{bmatrix}$$

The third case occurs if we have an interregional input-output matrix for flows between all sectors, but defined for regions whose scale is closest to that of a hierarchical level above the lowest (say at h = 3), and furthermore these regions do not conform with well-defined functional regions at this scale. For example, we have an observed matrix A_3 which has to be converted to $\tilde{A}_3$. Using case c of Table 1:

$$\tilde{A}_3 = J^{-1}A_3K^{-1} \tag{49}$$

where

$$J = P(G)(D_{3,4}'D_{3,4})^{-1}\ D_{3,4}'$$

$$K = E_{3,4}(E_{3,4}'E_{3,4})^{-1}\ P(G)'$$

whence

$$\tilde{A}_2 = D_{2,3}J^{-1}A_3K^{-1}E_{2,3}' \tag{50}$$

$$\tilde{A}_1 = \Gamma_3J^{-1}A_3K^{-1}\Gamma_3' \tag{51}$$

and A_4 may be estimated, as $\hat{A}_4$, from

$$\underset{(a_{ij}^{nm})}{\text{MAX}} - \sum_{i,j=1}^{R_4} \sum_{n,m=1}^{N_4} a_{ij}^{nm} \log a_{ij}^{nm}$$

subject to:

$$\sum_{\substack{i \in A \\ i=1}}^{n_A} \sum_{\substack{j \in B \\ j=1}}^{n_B} a_{ij}^{nm} = a_{AB}^{nm} \qquad \forall_{A,B,n,m}\ ; \quad A,B \in R_3\ ;\ n,m \in N_3 \tag{52}$$

$$a_{ij}^{nm} \geq 0 \qquad \forall_{i,j,n,m}$$

Here equation (52) matches the aggregate flows [from the n_A subregions, i, of region A, to the n_B subregions, j, of region B] to the observed values contained in the matrix A_3; a total of N_3R_3 constraints. In the absence of any other information about these flows this reduces to the estimate:

$$\hat{a}_{ij}^{nm} = \frac{1}{n_A n_B} a_{AB}^{nm} \tag{53}$$

Equation (53) provides a minimally biased prior estimate of the probability of each of the disaggregate flows occurring, that is subject only to being consistent with the more aggregate flows. If we had extra information, of course, these estimates would differ (Sheppard 1975; Snickars and Weibull 1977). Thus

$$\hat{A}_4 = G'NA_3NG \tag{54}$$

where N is a R_3N_3 by R_3N_3 diagonal matrix with i,i-th entry equal to $(n_j)^{-1}$ where j is the region represented in row i.

It thus follows that when A_3 is known $\tilde{A}$ in equation (42) is:

$$\tilde{A} = \begin{bmatrix} \Gamma_3 J^{-1} & & & \\ & D_{2,3} J^{-1} & & \\ & & J^{-1} & \\ & & & G'N \end{bmatrix} \begin{bmatrix} A_3 & & & \\ & A_3 & & \\ & & A_3 & \\ & & & A_3 \end{bmatrix} \begin{bmatrix} K^{-1}\Gamma_3' & & & \\ & K^{-1}E_{2,3}' & & \\ & & K^{-1} & \\ & & & NG \end{bmatrix} \begin{bmatrix} I & G_1 & G_1 G_2 & G_1 G_2 G_3 \\ & I & G_2 & G_2 G_3 \\ & & I & G_3 \\ & & & I \end{bmatrix}$$

3.3. Causal Structure

Two simplifications seem to exist in the type of hierarchical model represented by equation (42). First, national prices ($\underline{p}_1'$) seem to have no geographical variation, and secondly, that lower order goods are not consumed as inputs for higher order goods. As a result this model has apparently a rigid top down structure (cf. Rietveld 1981b). However neither of these interpretations need strictly speaking be true.

As regards the first, it should be noted that prices are f.o.b. prices set at the factory gate. In other words the price paid in various regions is not uniform; only the price prior to shipment is set. Transportation, the costs of which are added to this, are a lowest order good. In order to know transport costs, and thus local delivered prices, it is necessary to know the increase in cost per unit of distance, and the location of production and consumption. Note that for goods whose price is set at higher hierarchical levels the location given is at best aggregate, since the fine details do not affect price formation. The price paid for a unit of good $k \in N_1$ delivered in subregion $j \in R_4$ to industry m there is:

$$E(\theta_j^k) = p^k + \sum_{i=1}^{R} a_{ij}^{kn} p_i^t c_{ij}^k \qquad (55)$$

where c_{ij}^k is the amount of the transportation good used up in shipping a unit of k from i to j, and p_i^t is the price of a unit of transportation services bought at i. Here θ_j^k is the cost of good k delivered to j. θ_j^k is a stochastic variable with its expectation depending on the weighted average of transportation costs (equation 55). The variance of θ_j^k will be zero only in the event that shipments satisfy some globally optimal criterion such as the spatial price equilibrium of Takayama and Judge (1964). Equation (55) includes information which is not necessary to determine prices in equation (42), because it requires knowledge of a_{ij}^{kn}, the production and shipment of a high order good $k \in N_1$ between low order regions $i,j \in R_4$. The equations for price determination require only an aggregated version of this. For provision of the transport good, t in region i as an input to the production of good m in region j we need only to know:

$$a_{ij}^{tm} = \sum_{n=1}^{N} a_{ij}^{nm} c_{ij}^{n} \tag{56}$$

Turning to the second issue, the spatially uniform production price of some good $k \in N_g$ at the regional scale R_g need not imply that lower order goods $m \in N_h$ $(h < g)$ are not used as inputs. Rather it simply reflects the fact that the set of cities producing good k are cities of order g and above. For example goods $k \in N_1$ are only produced at one location in the nation, R_1; goods $\ell \in N_2$ are produced at most at only one location within each region $j \in R_2$, and so on. In other words the hierarchical level, g, at which any good is produced is defined by the lowest value for g such that no region $j \in R_g$ produces this good at two or more different locations with two or more different production prices.

This in turn implies that intraregional variations in the price of low order goods do not lead to significant intraregional variations in the production price for a good $k \in N_g$ in a region $j \in R_g$. We may modify the price determination equations to allow for this, explicitly introducing inputs of lower order goods.

Take as an example a good at the highest level. The equation determining production prices becomes:

$$p^k = \left[\sum_{\ell \in N_1} a^{\ell k} p^{\ell} + \sum_{g=2}^{4} \bar{d}_{gk}\right](1 + \pi) \tag{57}$$

where $\bar{d}_{gk}$ is the spatial average cost of all inputs of goods from the set N_g per unit of production of good k:

$$\bar{d}_{gk} = \frac{1}{R_g} \sum_{r=1}^{R_g} \sum_{m \in N_g} \sum_{i \in R_g} a_{ir}^{mk} p_i^m$$

$$= \frac{1}{R_g} \sum_m \sum_i p_i^m \sum_r a_{ir}^{mk} \tag{58}$$

The fact that good k is produced at just one location (perhaps the capital city) must be either because the spatial variance of $\bar{d}_{gk}$ is small relative to p^k or because of institutional restrictions. The latter possibility will not be discussed here. The spatial variance of $\bar{d}_{gk}$ would be relatively small if the a_{ir}^{mk}'s were either small or have a small spatial variance, and the p_i^m's either are small or have a small spatial variance. In either event it would be reasonable to use:

$$\bar{d}_{gk} \simeq \sum_m \bar{p}_g^m \bar{a}_g^{mk} \tag{59}$$

with $\bar{d}_{gk} = 0$ if $\bar{d}_{gk} \ll \min_{\ell \in N_1} a^{\ell k} p^{\ell}$ where $\bar{p}^m = \frac{1}{R_G} \sum_{i \in R_g} p_i^m$ and $\bar{a}^{mk} = \frac{1}{R_g} \sum_i \sum_r a_{ir}^{mk}$ (with $i, r \in R_g$) are the spatial averages of prices and input coefficients. Then, for $k \in N_1$:

$$p^k = \left[\sum_{\ell \in N_1} a^{\ell k} p^{\ell} + \sum_{g=2}^{4} \sum_{m \in N_g} \bar{a}^{mk} \bar{p}^m\right](1 + \pi) \tag{60}$$

Define the weighting vector, of dimension 1 by $N_g R_g$:

$$\underline{\alpha}'_{gk} = \left[\bar{a}^{1k}/R_g, \ldots, \bar{a}^{N_g k}/R_g,\ \bar{a}^{1k}/R_g, \ldots, \bar{a}^{1k}/R_g, \ldots, \bar{a}^{N_g k}/R_g\right]$$

Then equation (60) becomes:

$$p^k = \left[\sum_{\ell \in N_1} a^{\ell k} p^{\ell} + \sum_{g=2}^{4} \underline{p}'_g\, \underline{\alpha}_{gk}\right](1 + \pi) \qquad (61)$$

Equation (61) re-introduces the prices of all goods into the determination of all other goods. The same reasoning may be applied to second and third order goods. As a result, equation (42) becomes:

$$\underline{p}' = (1 + \pi)\underline{p}'\tilde{\tilde{A}} \qquad (62)$$

where

$$\tilde{\tilde{A}} = \left[\begin{array}{c|c|c|c} \tilde{A}_1 & \tilde{A}_1 G_1 & \tilde{A}_1 G_1 G_2 & \tilde{A}_1 G_1 G_2 G_3 \\ \hline \alpha_{21} & \tilde{A}_2 & \tilde{A}_2 G_2 & \tilde{A}_2 G_2 G_3 \\ \hline \alpha_{31} & \alpha_{32} & \tilde{A}_3 & \tilde{A}_3 G_3 \\ \hline \alpha_{41} & \alpha_{42} & \alpha_{43} & A_4 \end{array}\right]$$

and

$$\alpha_{rs} = [\underline{\alpha}_{ri_1} \ \cdots \ \alpha_{ri_{N_s}}] \text{ where } i_j \in \text{set of industries at hierarchical level } s.$$

In this modification of the hierarchical model, lower order goods are reintroduced as inputs to higher order goods in the form of spatial averages. However the spatial variance of prices is still zero at geographical scales below that scale at which each industry is realized. In this way a hierarchical spatial structure of prices is retained without the restriction of a "top down" causal structure.

4. CONCLUSION

This paper has presented some faltering steps in the direction of reorienting multiregional models so that they can be used to discuss the hierarchical nature of human settlement systems. The theme throughout is a search to integrate perspectives emphasizing regional variations and city size as factors mediating urban growth and change. The results show that it is in principle possible to construct relatively simple rules that translate multiregional into hierarchical systems. However the approach taken is basically technical; reducing essentially to aggregation and re-aggregation methods. One example was provided, but it remains for future research to determine if the methodology is applicable to more sophisticated multiregional models. It is also highly questionable as to whether such a technical adjustment is sufficient to capture the essential differences, and interrelations, between regional and urban systems. However it does seem at least to be a necessary initial step in constructing a hierarchical approach incorporating geographical interdependencies.

REFERENCES

Bassett, K., and P. Haggett (1971) Towards Short-Term Forecasting for Cyclic Behavior in a Regional System of Cities, in M. Chisholm, A. Frey, and A. Haggett (eds). *Regional Forecasting*. London: Butterworths, pp.389-413.

Berry, B.J.L., and D. Dahmann (1977) Population Redistribution in the United States in the 1970's. *Population and Development Review* 3:443-471.

Carroll, G.R. (1982) National City Size Distributions: What Do We Know After 67 Years of Research? *Progress in Human Geography* 6:1-43.

Christaller, W. (1933) *Central Places in Southern Germany*. Translated by C.W. Baskin. Englewood Cliffs, N.J.: Prentice Hall, 1966.

Curry, L. (1972) A Bivariate Spatial Regression Operator. *The Canadian Geographer* 16:1-14.

Gale, S., and D. Atkinson (1979) On the Set Theoretic Foundations of the Regionalization Problem. Pages 65-108 in *Philosophy in Geography*, edited by S. Gale and G. Olsson. Dordrecht, Holland: D. Reidel.

Harris, C.C. (1980) New Developments and Extensions of the Multiregional, Multi-industry Forecasting Model. *Journal of Regional Science* 20:159-171.

Hudson, J.S. (1972) *Geographical Diffusion Theory*. Northwestern University Publications in Geography No. 19. Evanston, Illinois: Northwestern University.

Karlström, U. (1980) *Urbanization and Industrialization: Modelling Swedish Demoeconomic Development from 1870-1914*. RR-80-44. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Kawashima, T., and P. Korcelli (eds) (1982) *Human Settlement Systems: Spatial Patterns and Trends*. CP-82-S1. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Kelley, A.C., and J.G. Williamson (1980) *Modeling Urbanization and Economic Growth*. RR-80-22. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Keyfitz, N. (1980) Do Cities Grow by Natural Increase or by Migration? *Geographical Analysis* 12:142-156.

Korcelli, P. (1981) *Migration and Urban Change*. WP-81-140. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Ledent, J. (1980) *Comparative Dynamics of Three Demographic Models of Urbanization*. RR-80-1. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Nystuen, J.D., and M. Dacey (1961) A Graph Theory Interpretation of Nodal Regions. *Papers and Proceedings of the Regional Science Association* 7:29-42.

Pigozzi, B. (1980) Inter-urban Linkages Through Polynomially Constrained Distributed Lags. *Geographical Analysis* 10:340-352.

Ponsard, C. (1977) Hiérarchie des Places Centrales et Graphes ϕ-flous. *Environment and Planning* 9A:1233-1252.

Pred, A. (1971) Large City Interdependence and the Preelectronic Diffusion of Innovations in the U.S. *Geographical Analysis* 3:165-181.

Richardson, H.W. (1973) *The Economics of Urban Size*. Westmead: Saxon House.

Rietveld, P. (1981a) A Review of Multiregional Economic Models. Paper presented at the IIASA Conference on Practice and Prospects of Multiregional Economic Modeling. Laxenburg, Austria, November 1981.

Rietveld, P. (1981b) *Causality Structures in Multiregional Economic Models*. WP-81-50. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Rogers, A., and F. Willekens (1978) *Migration and Settlement: Measurement and Analysis*. RR-78-13. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Sanderson, W. (1980) *Economic-Demographic Simulation Models: A Review of their Usefulness for Policy Analysis.* RR-80-14. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Seneta, E. (1981) *Non-negative Matrices and Markov Chains.* New York: Springer Verlag.

Sheppard, E. (1975) Entropy, Theory Construction and Spatial Analysis. *Environment and Planning* 8A:741-752.

Sheppard, E. (1979) Geographic Potentials. *Annals of the Association of American Geographers* 69:438-447.

Sheppard, E. (1980) *Spatial Interaction in Dynamic Urban Systems.* WP-80-103. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Sheppard, E. (1981) Regional Capitalist Production Structures. Paper presented at the North American Regional Science Association meetings, Montreal, November 1981.

Sheppard, E. (1982) City Size Distributions and Spatial Economic Change. *International Regional Science Review* 7:127-152.

Simmons, J.W. (1974) *Canada as an Urban System: A Conceptual Framework.* Research Paper No.62. Center for Urban and Community Studies, University of Toronto.

Simmons, J.W., and L.S. Bourne (1981) Urban and Regional Systems — Qua Systems. *Progress in Human Geography* 5:420-31.

Snickars, F., and J. Weibull (1977) A Minimum Information Principle: Theory and Practice. *Regional Science and Urban Economics* 7:137-168.

Takayama, T., and G.G. Judge (1964) Equilibrium Among Spatially Separated Markets. *Econometrica* 32:510-524.

Weissbrod, R.S.P. (1976) *The Diffusion of Relative Wage Inflation in Southeast Pennsylvania.* Evanston IL: Northwestern University, Studies in Geography No.23.