SPATIAL STRUCTURE AND SPATIAL INTERACTION

By

Engdawork Desta

A DISSERTATION

Submitted to
Michigan State University
in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Department of Geography

1988

ABSTRACT

SPATIAL STRUCTURE AND SPATIAL INTERACTION

by

Engdawork Desta

Traditional, gravity-type, spatial interaction models are commonly employed for predicting spatial flows. It has been reasoned that the models are not fully *specified* because they miss a vital variable representing the influence of *spatial structure* on spatial interaction.

Spatial structure, which is defined as the relative location of interacting centers, is assumed to vary from area to area. The absence of a variable representing this variation in the interaction model formulation causes the *parameters* of the *spatial separation* variable (usually distance) to vary locationally too. This happens because the distance measure, in addition to representing the effect of *spatial separation* which is expected to be *constant* over space, is also found to measure accessibility, which is location specific. Therefore, including a variable to represent the effect of *spatial structure* during calibration is believed to *specify* the model correctly. The *prediction*

capacity of the model is at the same time expected to improve as a result of removing the *misspecification* by way of including a *spatial structure* variable.

The attempt in this research, thus, has been to substantiate the role played by *spatial structure* in upgrading the usefulness of the *traditional* family of spatial interaction models in terms of their *specification* and *prediction capability*. The exercise involved calibration of the different versions of the *traditional* family of spatial interaction models using four different spatial flow data (telephone calls and rail passengers flows in Ethiopia, and Vehicle and bus passengers flows in the State of Michigan). During the calibration of each model various ways of entering the effect of *spatial structure* have been considered with the help of a computer program designed for this purpose. Comparisons are then made between the *traditional* spatial interaction models with *spatial structure* variables and those without them using coefficient of variation for assessing improvements in the model *specification*, and standardized-root-mean-square-error (SRMSE) as a goodness of fit measure.

The results of the exercise show that all the model calibrations do not behave in exactly the same way when the *spatial structure* variable of one type or another is considered. The effect varies not only by data type but also by type of model calibration. In general, this study has shown that the various ways of entering the effect of *spatial structure* impacts the models differently. In the case of the *doubly constrained* model calibration, including *spatial structure* variables reduces the models' usefulness. At best, the models remained unaffected. The constraints of this model, which are defined similarly as the *spatial structure* variables, seem to play the role of the *spatial structure* variables in addition to serving as balancing factors.

The approach used in this research is believed to contribute to further understand the behavior of spatial interaction models and the practical significance of *spatial structure* effect. Suggested future research includes a closer look at *spatial structure* as a new variable versus the balancing factors which are available as integral parts of the *production, attraction and doubly constrained* model calibrations.

ACKNOWLEDGMENTS

I wish to thank my advisor, Dr. Bruce Wm Pigozzi, for giving me the opportunity to work with and learn from him over the last four years, and for his helpful criticism of this manuscript. His encouragement, whenever I turned to him for advice, had always given me the vigor to work with enthusiasm until the completion of this dissertation.

I am also indebted to my committee members: Dr. Assefa Mehretu, Dr. Rene Hinojosa, and Dr. Richard Lyles for their constructive suggestions during the formulation of the problem and progress of this research.

Thanks are due to Mr. Jim Moore for writing the computer program necessary for this research. Mr. Moore's assistance in getting my requests answered at every level of the development of the program is highly appreciated.

I also wish to extend my appreciation to the Department of Transportation of the State of Michigan for making part of the data used in this work available to me.

Finally, thanks are in order to the Department of Geography for offering me the assistantship to pursue my study at Michigan State University, and Sage Foundation for its financial support to complete the write-up of this Dissertation.

TABLE OF CONTENTS

LIST OF TABLES

Page

TABLE

LIST OF FIGURES

INTRODUCTION

Geographers are interested in describing and explaining the distribution and variation of phenomena over space. The *gravity model* is among the many tools they use to investigate spatial patterns and spatial processes. This model is the most widely used mathematical descriptions of spatial interaction. It is commonly applied in such areas as transportation, marketing, migration and various locational studies.

The gravity model measures interaction by relating the magnitude (often population size) of *pairs* of interacting places such as cities, and some measure of the *spatial separation* between them. The basic assumption expressed by this model may be stated as follows: *interaction between two places is directly proportional to the relative attraction of each place and inversely proportional to some function of the distance between them.* Large cities, for example, are expected to generate or attract more interaction than small cities. And, the farther two places are apart, the less they interact (U.S. Department of Commerce, Bureau of Public Roads, 1965).

In spite of its popularity, the gravity model still requires improvements so that spatial interactions can be estimated more accurately. A number of studies have indicated that the *traditional* gravity model, which represents scale and distance impacts, is not fully *specified.* That is, the model does not include all the variables that explain interaction over space. An important element that the *traditional* gravity model does not consider has been noted as a *spatial structure* effect (Gould, 1975; Fotheringham, 1981, 83, 84; Baxter, 1986).

Spatial structure is also sometimes referred to as *spatial pattern.* However the two are not treated in exactly the same way. Usually, spatial pattern represents a subjective description of the configuration of phenomena over space. On the other hand, *spatial structure* is more specific because some form of *measurement* is applied to make it an *objective* representation of spatial pattern so that comparison of the locational importance of centers could easily be made. Hence, numerical values are assigned to individual locations within a system of interacting origins and destinations. This might be performed in a number of ways; the time it takes to reach one center from all the other centers of the system, or cost of travel, or simply the distance separating them.

Indices of *spatial structure* derived from only distance values may not be sufficient however to represent the relative locational importance of a center. In the case of cities, for instance, the volume and types of services rendered by them are also important spatial characteristics. Therefore, in order to make objective comparisons between locations in a given system of origins and destinations of flows, some function of the size and/or characteristics of each city may be used together with one or the other of the distance measures to model individual relative locational importance. This approach which is similar to Stewart's *population potential* (1941), is described in more detail in Chapter 2.

Interaction over space is defined as the movement or flow of people or goods between locations. The problem challenging many geographers and scholars in other disciplines, is determining the factors which explain the volume of flow between origins and destinations. Most of the discussions in this respect have revolved around the role of the *friction factor*, normally, the coefficient of the distance measure (Price, 1948; Anderson, 1950). In spite of the volume of work produced, there are still unanswered, but important, practical and theoretical questions regarding interaction over space. The effect of *spatial structure* on interaction has not been given sufficient attention. Some of the attempts made in this regard are discussed in Chapters 1 and 2.

This study is designed to investigate the strength of the influence of *spatial structure* on spatial interaction. The work involves observation of the behavior of the *traditional* (also called *general*) family of gravity models (Wilson, 1967) using various types of spatial flow data and various methods for the inclusion of the effect of *spatial structure*.

The problem statement, the specific questions of the research and hypotheses are presented in Chapter 3. The various types of spatial flow data used in this study and, the methods of analysis are described in Chapter 4. Chapter 5 has the results and analysis. Finally, a summary of the study and concluding remarks are presented in Chapter 6.

CHAPTER 1.

SPATIAL INTERACTION MODELS

1.1. The Gravity Model

Spatial interaction models are adaptations of the *gravity model* advanced by Newton in 1686. In simple terms, Newton's Law of Gravitation states as follows (Pollard 1969, p. 70):

> Every body attracts every other body with a force proportional to their two masses directed toward each other, and also proportional to the inverse square of their distance apart.

The mathematical representation of this statement is generally given as:

$$F_{12} = g(M_1 M_2)/d_{12}^2 \qquad (1)$$

where:

F_{12} = attraction force between M_1 and M_2
M_1, M_2 = two attracting masses
g = universal constant
d_{12} = distance between M_1 and M_2

Spatial interaction models used by geographers are modifications of Equation 1. As introduced earlier, potential interaction between two places (generally towns or areas) is explained in terms of some measurement of the size of the places and some measurement of the distance between them. Later improvements of interaction models are also based on similar principles.

1.2. Developments in Spatial Interaction Modeling

Carey (1858, 59) made the first adaptation of the gravity concept of human interaction. According to him, (Carothers 1956, p. 95):

> The force of interaction between two concentrations of population, acting along a line joining their centers, is directly proportional to the centers and inversely proportional to the square of the distance between them.

His spatial interaction formulation is thus given as:

$$T_{ij} = g(P_i P_j)/d_{ij}^2 \qquad (2)$$

where:

T_{ij} = the force of attraction between concentration i and j
P_i, P_j = two concentration of population
d_{ij} = distance between concentration i and j
g = a constant of proportionality (similar to the gravitational constant of physics)

Later, Ravenstein (1885) applied the law of gravitation to social interactions. Many researchers have since then been involved in studying spatial flows by further modifying Newton's gravity model for social science purposes (for example, Young, 1924; Reilly, 1929; Price, 1948; Anderson, 1955, Wilson, 1967, 71; Griffith, 1976).

The most problematic component of the gravity model when applied to social science purposes has been the *friction factor* (Stewart, 1941; Price, 1948; Anderson, 1955; Zipf, 1941, 42). Opposition to the use of the gravity-type spatial interaction model, as a result, concentrated mostly on the inability of the distance variable to measure the true effect of *friction* between interacting locations. Probably, one of the first men to react against the gravity-type spatial interaction model was Stouffer (1940). According to him, distance and interaction are not necessarily related. Instead, he writes (p. 846):

> ... the number of persons going a given distance is directly proportional to the number of *opportunities* at that distance and inversely proportional to the number of *intervening opportunities.* (emphasis added)

By *opportunities* is meant, the number of potential destinations that are available to a given origin. Stouffer's intervening opportunities model initiated further research (for example, Isbell, 1944; Stewart, 1947; Strodtbeck, 1949; Carrothers, 1956; Shneider, 1959; Tomazinis, 1962; Clark, 1962; Miller, 1972; Baxter and Ewing, 1979) and today it is, probably, the main contender of the gravity-like interaction models (Price, 1948; Ikle, 1954; and Carroll, 1955).

Unlike the ideas of Zipf (1941, 42, 46, 47), and Stewart (1941, 47, 48, 50), among others, the distance relationship with interaction has been reported to vary instead of being a simple inverse of the square value. Thus, various exponents for distance were introduced by Price (1948) and Anderson (1955) such that Equation 2 may be modified and restated as:

$$T_{ij} = g(P_i P_j)/d_{ij}^{\beta} \qquad (3)$$

The beta (β) value represents the effect of *spatial separation* between interacting population concentrations which is assumed to vary for various types of spatial flows within and between a system of origins and destinations.

The above modification has been found too simple and insufficient to account for observed differences in generation or attraction of flows of centers in different circumstances. Suggestions were forwarded to modify the population (masses) factors (Roether, 1949; Stewart, 1950;

Dodd, 1950 and Anderson, 1955, 56). This has been done in two ways: first, to account for variations in population characteristics such as income, education and gender, some form of weights have been attached to the population variables. A modified version of Equation 3 with the weights considered is given as:

$$T_{ij} = g(\theta P_i \; \phi P_j)/d_{ij}^{\beta} \qquad (4)$$

where:

T_{ij} = interaction between i and j
P_i, P_j = population of i & j
θ = weighting factors for P_i
ϕ = weighting factors for P_j
β = parameter to be estimated
g = scale parameter
d_{ij} = distance between i and j

The second suggestion, in addition to the weights, is raising the population factors to some power. The reason for this is to control for size and/or agglomeration effects of population or activities. With the mass exponents included the new spatial interaction model is stated as follows:

$$T_{ij} = g(\theta P_i^{\Omega} \; \phi P_j^{\alpha})/d_{ij}^{\beta} \qquad (5)$$

where:

T_{ij} = interaction between i and j
P_i, P_j = population of i & j
θ = weighting factors for P_i
ϕ = weighting factors for P_j

Ω, α, β = parameters to be estimated
g = scale parameter
d_{ij} = distance between i and j

Later, Equation 5 has been further refined and a family of four interaction models have been developed (Wilson, 1971). If, for example, in an origin-destination flow matrix of *n* by *m* dimension, the attributes, that is, ***propulsiveness power*** of the origins (θP_i) is measured by V_i and the ***attractiveness*** of the destinations (ϕP_j) is measured by W_j, the ***total flow constrained*** version, the first of the family of spatial interaction models, could be stated as follows:

$$T_{ij} = g\, V_i^{\Omega}\, W_j^{\alpha}\, d_{ij}^{\beta} \qquad (6)$$

where:

T_{ij} = estimated flow from i to j
V_i = origin propulsiveness
W_j = destination attractiveness
d_{ij} = distance between i and j
g = scale parameter
Ω, α, β = parameters to be estimated

Equation 6 is very similar to Equation 5. In Equation 6, information is available for only total flows. The scaling constant g insures that the sum of the estimated interactions from the model is equal to the sum of the actual interactions. This is done by comparing total actual flows with total predicted flows. The scaling constant is computed using the following formula:

$$g = T / \sum_{i}^{n}\sum_{j}^{m} V_i^{\Omega} W_j^{\alpha} d_{ij}^{\beta} \qquad (7)$$

where:

$$\sum_{i}^{n}\sum_{j}^{m} V_i^{\Omega} W_j^{\alpha} d_{ij}^{\beta} = \text{estimated total flow}$$

$$T = \sum_{i}^{n}\sum_{j}^{m} T_{ij} = \text{actual total flow}$$

The *production constrained* model, the second version of the set, is stated as follows:

$$T_{ij} = A_i O_i W_j^{\alpha} d_{ij}^{\beta} \qquad (8)$$

where:

T_{ij} = estimated flow from i to j

$A_i = 1/(\sum_{j}^{n} W_j^{\alpha} d_{ij}^{\beta})$ = production constraint (balancing factor)

O_i = known origin totals

W_j = destination attractiveness

d_{ij} = distance between i and j

In this model, information about the number of trips leaving each center is known but where they end up is not. The constraint, A_i, insures that the estimated outflow is equal to the known outflow, that is, $\sum_{j}^{n} T_{ij} = O_i$ for all j. The *production constrained* model forecasts destination inflows that are unknown.

The third version, the *attraction constrained* model, on the other hand, forecasts unknown origin outflows on the basis of known incoming flows. The constraint, B_j, insures that the estimated inflow is equal to the known inflow, that is, $\sum_{i}^{m} T_{ij} = D_j$ for all i. This model is given as:

$$T_{ij} = B_j D_j V_i^{\Omega} d_{ij}^{\beta} \qquad (9)$$

where:

T_{ij} = estimated flow from i to j
$B_j = 1/(\sum_{i}^{m} V_i^{\Omega} d_{ij}^{\beta})$ = attraction constraint (balancing factor)
V_i = origin propulsiveness
D_j = known destination totals
d_{ij} = distance between i and j

Finally, the *doubly constrained* version, the most refined of the family of gravity models, is given as:

$$T_{ij} = A_i B_j O_i D_j d_{ij}^{\beta} \qquad (10)$$

where:

T_{ij} = estimated flow from i to j
O_i = known outflow totals
D_j = known inflow totals
$A_i = 1/(\sum_{j}^{n} B_j D_j d_{ij}^{\beta})$
$B_j = 1/(\sum_{i}^{m} A_i O_i d_{ij}^{\beta})$

This model is formulated from known inflows and known outflows. Because they are interdependent, the production and attraction constraints, A_i and B_j, are optimized iteratively. Often a value of 1.00 is assigned to B_j and with that A_i is estimated. Then using the estimate of A_i, B_j is calculated. The process continues until successive values of A_i and B_j show no change. Their purpose is the same as given for the *production* and *attraction constrained* versions (Equations 8 and 9). That is, A_i and B_j mathematically insure that $\sum_{j}^{n} T_{ij} = O_i$ and $\sum_{i}^{m} T_{ij} = D_j$ respectively. These constraints control that the estimated total outflows and total inflows balance with the actual total outflows and total inflows with respect to each *propulsion* and *attraction* center within the given system of origins and destinations. Since the constraints of the *singly* and *doubly constrained* models are origin and destination accessibilities, they can also be used as measures of relative locational competition with respect to each other (Wilson, 1967).

In each of these versions, it is same principle operating which derives the simple gravity model. That is, the estimated flow between given *pairs* of locations is an increasing function of the *propulsiveness* and *attractiveness* of the locations and a decreasing function of the *spatial separation* between them. Instead of using distance as a measure of *spatial separation* between locations, cost of travel (Isard and Freutel, 1954) and travel time (Carroll, 1955; Anderson, 1956) have also been suggested.

One of the major reservations of using the *traditional* set or family of gravity models in the social sciences has been associated with the lack of adequate theoretical backing to support their adaptation. In trying to overcome this problem, Wilson (1967) introduced a plausible theoretical explanation of the gravity-type spatial interaction models based on statistical mechanics. In 1970 he showed the derivation of the model by entropy maximization methodology.

Long and Uris (1971) suggested a change in the definition of spatial interaction after empirically observing the improvement of their work over previous attempts. Their definition reads as follows (p. 155):

> The amount of interaction originating at one place and terminating at another is proportional to the population of the places and inversely proportional to the distance between the places and to the population lying at lesser distance from the origin than the destination does. The size of the effects will differ for interactions taking place within and between city size groups, that is, ... for interactions within and between hierarchies.

What has been done is actually a *synthesis of the gravity* and *intervening opportunities* models as shown below:

$$T_{ij} = g(P_i^{\Omega} P_j^{\alpha})/(d_{ij}^{\beta} I_{ij}^{\tau}) \qquad (11)$$

where:

T_{ij} = passenger from i to j
P_i, P_j = population of i and j
d_{ij} = distance between P_i and P_j
I_{ij} = the sum of population of centers intervening between i and j
g = scale parameter
Ω, α, τ = parameters to be estimated

A test of Equation 11 on airline data of 1960 is reported to have given improved results (Long and Uris, p. 156). This approach has been contested by Miller (1972). According to him the *gravity* and *intervening opportunities* methods are *competing* measures of interaction implying that they may not be used together. The reason being that the *gravity* model hypothesizes decreasing interaction with increasing distance while the *opportunities* model hypothesizes decreasing interaction with increasing intervening population.

When the *gravity* model is compared with the *intervening opportunities* model, Long and Uris found out that "for trips from large to small cities, *intervening opportunities* has a slight edge and for trips from small to large cities *distance* does" (p.160). Miller's work, however, shows that *opportunities* are better explanatory variables than *distance* although the way he interprets his results is not sufficiently clear.

However, more attention has been given to the gravity-type spatial interaction models. Refinements of these models have been intensified over the past fifty years. Some of the works in this regard are presented in the following Chapter.

CHAPTER 2.

THE QUESTION OF SPATIAL STRUCTURE

2.1. Theoretical Perspective

The question of the *friction* (also called *distance-decay, deterrence, distribution* or *spatial separation*) factor, as indicated earlier, has been one of the most important concerns in spatial interaction modeling. Curry (1972, p. 132) comments that

> Any calibration is specific to particular spatial pattern of origins and destinations and may be substantially meaningless. Different degrees of clustering will exhibit different *frictional* terms even if friction is known to be constant. (emphasis added)

The clearest reflection of the problem is revealed in attempts to calibrate a simple gravity model (Equation 6, p. 10) for *individual nodes* of a network. For example, one could estimate the parameters of Equation 6 for each i (if the number of flows from each were sufficient). Thus, it is possible to produce estimates of beta, the distance parameter, for each node i. If the model is properly *specified*, these beta values should all be equal. However, the simple gravity (and more complex ones too, Equations 8, 9 and 10, pp. 11-12) produces beta values which vary significantly, within the same network. The general hypothesis is that variation in the beta values occurs because the *traditional* gravity models do not include the influence of *spatial structure* on interaction necessitating a closer look at the problem summarized by Curry. If the general hypothesis is right, that is, if *the spatial configuration of interacting centers, that is the relative locations of points in a network (which is presumably different for different points within a given system of origins and destinations) is believed to be the cause for the friction factor to vary over space*, introducing a *spatial structure* variable into the *traditional* gravity models will give *constant frictional* values over space.

Cliff and Martin (1974), basing their argument on the works of Olsson (1970), Curry (1972), Johnston (1973), disagree by noting that the *frictional* effect of distance is not seriously affected by a *map pattern*, particularly for interurban flows. According to them, the *confounding* effects of spatial interaction variables may only be relevant for intraurban interactions.

Long (1969) argues that the greatest weakness of the gravity model is its lack of concern for the influence of alternative destinations of travel between pairs of cities. For him, Stouffer's *intervening opportunities* model is not satisfactory because it totally ignores the effect of distance. Long suggests that the position of a city within a *spatial structure* and *distance* together with *intervening opportunities* need be considered (Equation 11, p. 15). "The components of the gravity model are," he writes, " elements in a traveller's decision and he makes these decisions on the basis of alternatives" (p.107). Similar notions of distribution of *opportunities* other than the interacting bodies influencing total interaction has been extensively discussed by Rushton (1969), and Glejser and Dramies (1969). As in the previous case, the shortcoming of the gravity model is emphasized because of its failure to take into account the *spatial structure* effect although Glejser and Dramies did not come up with a satisfactory mechanism of measuring this effect and incorporating it with the gravity model.

Using the same argument, Ewing (1974) writes in agreement with those researchers who believe in the influence of alternative locations on interaction. To this end he writes (p. 85):

> The distance to alternative destinations change as the origin point changes and each different set of such distances will differently affect the migration probabilities from their respective origins.

Ewing assumes that this condition equally applies to the *intervening opportunities* model. Cesario (1975) also elaborates on the same notion, in a step by step manner; how a failure to consider relative locations in the gravity model affects interaction.

The *friction* factor also worries Gould (1975). In a study based on Curry's finding, (p.83) he states:

> ... distance parameters appear to index the relative accessibility of a location .... The supposed effect of distance is not really a distance effect in any traditional or behavioral sense, but rather a matter of where a group is located within a set of information generation regions.

Gould's comment substantiates Curry's observation. The distance exponents of a single system of origins and destinations do not remain *constant* over space. Instead of measuring the spatial separation effect, they represent accessibilities of the individual origins and destinations, making the behavioral interpretation of the *friction* factor impossible.

Probably, it was Claesson (1964) who first formally presented the relationship between *spatial structure* and *spatial interaction* (Griffith, 1975). Claesson's findings, include: (i) the presence of a systematic change in attraction fields attributable to population variations taking place within the center; that is, as the population, and hence, the hierarchical position of an urban center increases, distance becomes less of a *deterrence*; (ii) the existence of a relationship between the attractiveness of a given place within the urban hierarchy and that of its subordinate centers; and (iii) the fact that differences between these fields are not only influenced by population and distance but also by the *size* and *spacing* of neighboring centers (Griffith, p. 733). After experimenting with the effect of the location of centers, Griffith states (p. 738):

> The gravity models fail to discriminate between components of interaction, that is, the masses and distance variables, by not being able to capture their individual effects explicitly.

The attempt to measure *spatial structure* and consider it in calibrating the traditional gravity models is, in part, to be able to separate the effect of these individual variables.

A thorough discussion of the spatial relationships between individuals, and individuals and groups has been made by Knippernberg and DeVos (1983). They note the importance of the *context*, referred to differently as *structural*, *compositional*, *group-level*, or *neighborhood* effects influencing individual decisions. According to them, the *traditional* gravity model places too much emphasis on *individual* effect, that is, the effect of the *propulsiveness* and *attractiveness* power of *individual* centers, without giving attention to the *context* under which each interacting location finds itself (pp. 120-128). Therefore, it may be appropriate to consider influence *multi-directionally*, that is, how locations as a *group* relate with an *individual* center in addition to how *individual* centers relate with a *group* within a system of spatially interacting locations.

The assumption "that the choice of one place over another has nothing to do with all other places in the system is," Sheppard writes, "open to serious question" (1978, p. 397). In another work, he makes it even more explicit by stating the *interdependence* between *spatial interaction* and *spatial structure* (1979, p. 438). In support of the influence of alternative locations in trip making he, writes (1984, p. 370):

> ... the spatial structure or relative location of population profoundly affects the number of trips made of any given length.... what spatial structure of population refers to is the spatial distribution of those attributes affecting trip generation propensity or destination attractiveness.

The attributes of population, such as income and education, are spatially distributed and the pattern of their distribution varies from one location to another. The distance separating the population locations can also have various attributes, such as, bad or good communication and costly fares. All of them can have some influence on interaction. Consideration of accessibilities of alternative locations is assumed to capture the effect due to variations in the relative locations of places and make the gravity model more useful.

In referencing Curry (1972), Sheppard comments that the unpredictable variation of the distance exponent, or beta, may be due to the fact that the spatial pattern of origins and destinations is different in each study area (1984, p. 370). He also notes that there is a possibility for the beta estimator to be affected by *spatial structure* of population if some of the statistical assumptions made during model calibration are ignored (p. 372).

The problem associated with the estimation of the distance parameter in the gravity-type spatial interaction models has been the prime concern of Fotheringham since 1979. He writes that differences in *spatial opportunities* over space result in *different distance exponents* (1983, p. 34). He (1981) remarks strongly that accurate estimates can be made if other elements of the spatial system, that is, the size and configuration of origins and destinations, are also modeled. According to him, (pp. 15 & 20):

> The usual interpretation of the *distance decay* parameter which is taken to describe the relationship between the observed interaction pattern and distance is actually false because

> the *spatial structure* effect is not taken care of. (emphasis added)

and

> .... the principle which operates but which is ignored in gravity modeling is that the more accessible a destination is to all other destinations in a spatial system, the less likely it is that that destination is a terminating point for interaction from any given origin.

With such qualifying statements, Fotheringham develops a set of equations he calls *competing destinations* models. He does this by adding, to the traditional family of gravity models, a variable that represents the *competing* position of an interacting destination in relation to other destinations available for an interacting origin. The variable he then introduces represents the *spatial structure* effect which he believes will correctly specify the *traditional* gravity models.

The additional variable makes the model statement more complete. As a result, *over* and *under* predictions of spatial flows, which result from the *misspecification* of the *traditional* gravity models, are expected to be corrected. (Fotheringham, 1983, Baxter, 1988). This is in addition to the constraints of the family of gravity models (pp. 11-12) which can also be, as indicated earlier, *measures of competition* (Wilson 1967).

In summary, although the gravity model borrowed from physics is the basis of spatial interaction models, a number of important modifications have been made to make it suitable for social science purposes. But, the changes made to the gravity model have not resulted in sufficiently improved spatial interaction models so far. The problems associated with *correctly specifying* spatial interaction models, as have been addressed by the different researchers, have far reaching theoretical and practical consequences.

The models have not served the purpose in the social sciences satisfactorily because the theories of physical and social sciences are probably only partially compatible with each other as far as spatial interaction modeling is concerned (Shneider, 1969). As a result, many alterations have been made to the *basic model.* As Knudsen (1986) remarks, spatial interaction models are associated with very high error margins also. If a model is not based on a well founded theory, satisfactory performance may not be expected from its application. Hua (1979, p. 117) also writes, "... the failure to yield a truly predictive model is rooted in their lack of a causal foundation". Emphasis therefore needs to be placed on identifying the underlying problem of the model *statement* or *specification.* This study attempts to address this problem by measuring the effect of *spatial structure* variables on the distance parameter and origin-destination flow estimates of the *traditional* gravity models.

The practical aspect of the problem of spatial interaction modeling is actually dependent upon the soundness of the theory on which it is based. If the theory, which is the guiding principle, is not clear in the first place, interpretation of results will become difficult. Hence, for example, the inability to precisely determine the role of the *spatial separation* element is both a question of theory and a practical problem of methodology. Hence, attempts made to improve the performances of interaction models should look to theoretical explanations and the means for their practical realization.

2.2. Measuring Spatial Structure

2.2.1. Introduction

Having established the influence of *spatial structure* on spatial flow theoretically, the next step is to verify it in empirical terms. If flows reflect not only the influence of the factors of *propulsiveness* and *attractiveness* of locations but also the impact exerted by the specific *spatial position* each interacting origin and/or destination occupies in relation to all others, then, the *traditional* gravity model statements are not complete since they do not consider the influence of the *spatial structure* of the interacting locations (Lendent, 1981). As explained before, the absence of a variable of *spatial structure* in the gravity model statement *misspecifies* it, and when calibrating this *misspecified* model, the distance parameters will vary spatially because of the influence of *spatial structure*. This has to be expected because the number, size, and relative location of any given system of interacting places are rarely identical. What this implies is that parameter estimates of spatial flow will be *location specific*, that is, different for individual nodes in a network (Thomas and Hagget, 1974). But how is *spatial structure* measured so that its effect could be accounted for?

Let us, for example, take a hypothetical situation where there is only one destination and a number of origins. Let us assume that the origins have the same power to generate trips and they are equidistant from the single destination. The destination will not face any competition because there is no other destination. If all else is assumed constant, the flow from any one of the origins to the destination measured by a gravity-type model will be identical. The influence of *spatial structure* which is made up of a number of origins of *equal propensity to generate trips* and a single destination will be *constant*.

Suppose one more destination is added to the picture. If this new destination is also equidistant to all the origins, and its *attractiveness* power is the same as the already existing destination, the flow from the origins will be shared between them since the *power of attractiveness* of the two destinations is the same, all other things being equal.

If we alter the *attractiveness* power of one of the destinations, immediately, the pattern of flow changes because of the change in the *relative attractiveness* of the two destinations. The two destinations may not, however, be competing with each other as long as there is still a need for additional destinations. *Spatial structure* becomes meaningful here if it represents the position of the destinations relative to each other.

If the *propulsiveness* power of the origins is altered, the number of trips generated by each of them is expected to vary. The flow structure will likewise change giving a varying relative importance to individual origins.

If *spatial structure* is defined as the configuration of locations in a given system of origins and destinations, then both the above conditions qualify. But, some kind of distinction may be necessary between destinations competing with each other for flows from origins, and origins competing for a destination.

Such hypothetical situations may be useful to direct thought processes although reality may not lend itself to the very simplistic assumptions. It is only by design that we can think of *uniform attractiveness or propulsiveness* of destinations and origins. For all practical purposes they may not exist. Because of spatial variations in the distribution of phenomena, each spatial element exhibits a unique locational advantage or disadvantage. In the case of urban centers, for example, the pattern of flows is expected to follow, partly, the relative accessibility of each center involved in the interaction process. As suggested by the various researchers, the *traditional* gravity model does not capture this aspect of spatial interaction. Measuring *spatial structure* and incorporating its effect in this model is an attempt to *specify* it more correctly so that accurate predictions and generalizations could be made.

2.2.2. Illustrating Spatial Structure Measurement:

2.2.2.1. Population Potential-type Spatial Structure Measure

A population potential-type approach is used to create *spatial structure* variables (Stewart, 1948; Sheppard, 1979).

Origin and destination based *spatial structure* measures with and without the interacting origins and destinations are computed using simple interaction and distance matrices created for this purpose. Fotheringham has applied a similar notion to observe the effect of *spatial structure* on interaction (1981). Observing the difference in the level of influence between a vector and matrix accessibility values is one of the questions addressed in this study.

The mathematical form of population potential is:

$$V_j = k \sum_{i=1}^{n}(P_i/d_{ij}) \qquad (12)$$

where:

V_j = influence
P_i = population of i
d_{ij} = distance between i & j
n = total centers
k = constant

Suppose we have flow data of a system of five interconnected origins and destinations as presented below:

Table 1

Interaction (Flow) Matrix

Orig/Dest	1	2	3	4	5	Total
1	---	5	2	6	9	22
2	7	---	2	3	2	14
3	3	6	---	8	9	26
4	2	8	1	---	4	15
5	6	3	7	1	---	17
Total	18	22	12	18	24	94

Table 2

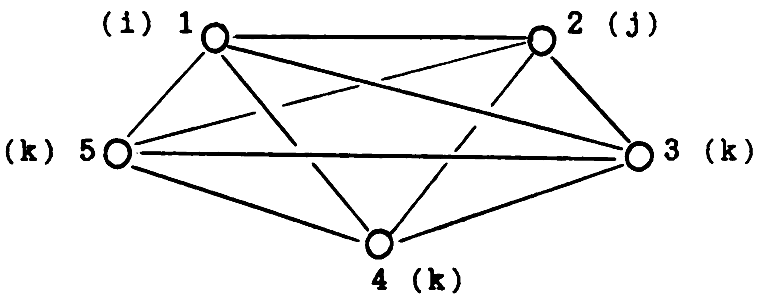
Distance Matrix

Orig/Dest	1	2	3	4	5
1	---	2	6	8	2
2	2	---	3	7	4
3	6	3	---	4	5
4	8	7	4	---	7
5	2	4	5	7	---

The graph of this data set may be represented as follows:

Figure 1

A System of Five Interacting Locations



Suppose in the picture above, location 1 (i) is flow origin, location 2 (j) flow destination, and the rest are alternative destinations for origin 1. (For our purpose, let us concentrate on the flow from origin 1 (i) to destination 2 (j)). A *spatial structure* measure that will be made a component of the *traditional* gravity model to predict the flow from origin 1 to destination 2 may be computed following the population potential-type approach (Equation 12).

The *spatial structure* measurement may be *origin based*, that is, computing the accessibility of location 1 to all the destinations. The formula and the diagram associated with it is as follows:

$$Z_{ij} = 1/\sum_{\substack{k=1 \\ (k\neq i,j)}}^{n}(W_k/d_{ik}) \qquad (13)$$

where:

Z_{ij} = accessibility of an interacting origin with destinations

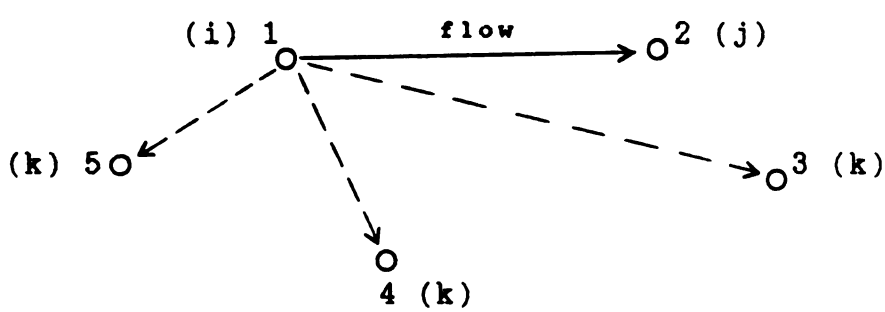
W_k = a weight of destination k, in this case total inflow

d_{ik} = distance between the interacting origin i and destination k.

n = total locations in the system, in this case 5.

Figure 2

Origin Based Potential-type Spatial Structure without the Interacting Destination



From the given data and distance matrices, this origin based accessibility measure results in the following values:

Table 3a

Origin Based Potential-type Accessibility

Orig/Dest	1	2	3	4	5
1	---	**.062^**	.040	.040	.066
2	.080	---	.057	.053	.064
3	.060	.081	---	.066	.067
4	.105	.115	.113	---	.119
5	.096	.072	.059	.059	---

[For example, **.062^** = 1/((12/6)+(18/8)+(24/2))]

If location 2, the interacting destination with origin 1, is included in the computation, since in interactions, nodes serve as origins and destinations of flows simultaneously, then Z_{ij} becomes Z_i which is a vector of the following values:

$$Z_i = 1/\sum_{\substack{k=1 \\ (k \neq i)}}^{n} (W_k/d_{ik}) \qquad (14)$$

Figure 3

Origin Based Potential-type Spatial Structure with the Interacting Destination

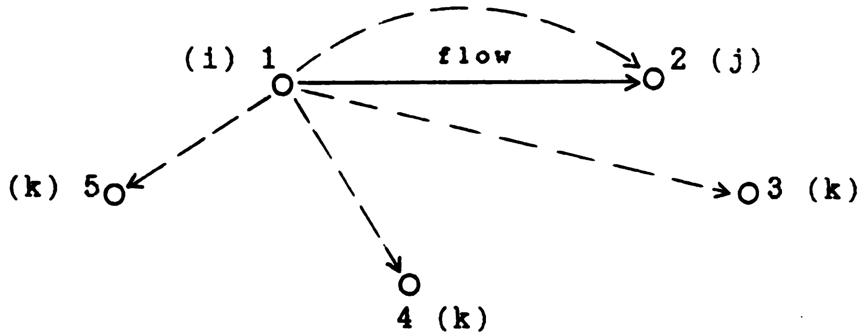


Table 3b

Origin Based Potential-type Accessibility

O&D	Values
1	.037
2	.046
3	.051
4	.085
5	.051

Alternatively, a *destination based* accessibility could be computed (Figure 4) as a measure of the relationship of an interacting destination, location 2, with other destinations, locations 3, 4 and 5, available for a given interacting origin (Fotheringham, 1981). This time the formula becomes:

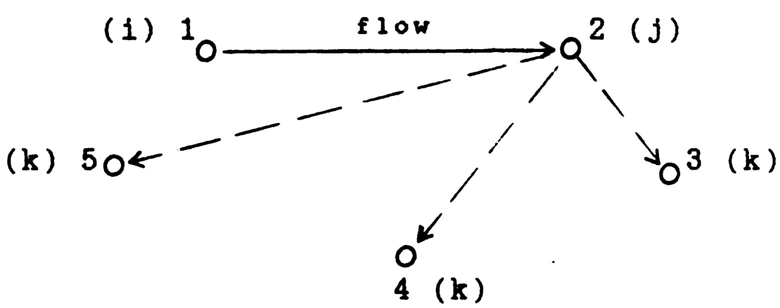
$$P_{ij} = 1/\sum_{\substack{k=1 \\ (k \neq i,j)}}^{n} (W_k/d_{jk}) \qquad (15)$$

where:

P_{ij} = accessibility of an interacting destination with other destinations
W_k = a weight of destination k, in this case total inflow
d_{ik} = distance between the interacting origin i and destination k.
n = total locations in the system (5)

Figure 4

Destination Based Potential-type Spatial Structure without the Interacting Origin



From the flow and distance matrices above, the destination based accessibility, that is, P_{ij}, gives the following matrix values:

Table 4a

Destination Based Potential-type Accessibility

O&D	1	2	3	4	5
1	---	**.080^**	.060	.105	.096
2	.062	---	.081	.115	.072
3	.040	.057	---	.113	.059
4	.040	.053	.066	---	.059
5	.066	.064	.067	.119	---

[For example, **.080^** = 1/((12/3)+(18/7)+(24/4))]

If location 1, the interacting origin with location 2, is considered as one of the destinations and is included in the computation, then P_{ij} becomes P_j and a vector of the following values will result:

$$P_j = 1/\sum_{\substack{k=1 \\ (k \neq j)}}^{n} (W_k/d_{jk}) \qquad (16)$$

Figure 5

Destination Based Potential-type Spatial Structure with the Interacting Origin

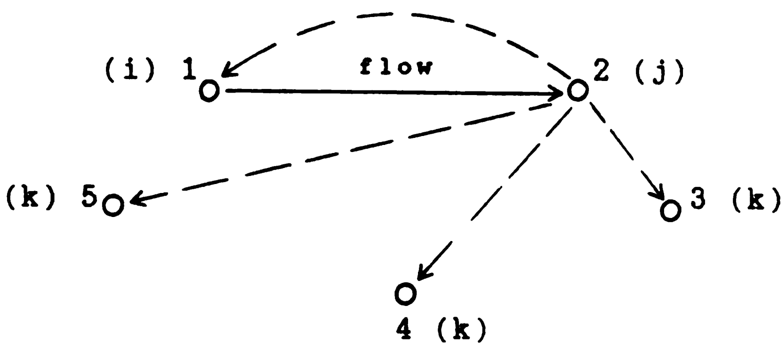


Table 4b

Destination Based Potential-type Accessibility

O&D	1	2	3	4	5
Values	.037	.046	.051	.085	.051

As could be observed in the tables above, the values of the origin based potential-type *spatial structure* measures (Tables 3a and 3b) are similar with the Fotheringham-type (destination based) *spatial structure* values (Tables 4a and 4b) except that the respective tables are transpose of each other.

2.2.2.2. Distance-type Spatial Structure Measure

A question addressed in this study also relates with comparing connectivity matrices (Griffith and Jones, 1980) as a purely distance measure of *spatial structure.* Instead of a population potential-type approach, therefore, origin and destination based *spatial structure* indices with and without the interacting centers, are produced, in a similar manner conducted above, using the distance values separating each location of the system of origins and destinations for the same interaction and distance matrices. Considering the same centers, that is, location 1 and 2 as the interacting origin and destination respectively, an origin based accessibility of only the distance variable may be stated as follows:

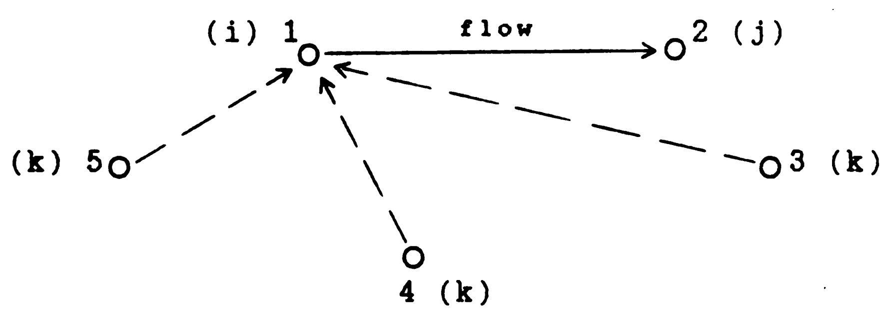
$$Z_{ij*} = 1/(V_i/\sum_{\substack{k=1 \\ (k\neq i,j)}}^{n} d_{ik}) \qquad (17)$$

where:

Z_{ij*} = distance accessibility of an interacting origin with destinations
V_k = a weight of origin i,in this case total outflow
d_{ik} = distance between the interacting origin i and destination k.
n = total locations in the system (5)

Figure 6

Origin Based Distance-type Spatial Structure Variable without the Distance between the Interacting Origin and Destination



The matrix values, that is, indices of *spatial structure* derived using Equation 17 are given in Table 5a.

Table 5a

Origin Based Distance-type Accessibility

O&D	1	2	3	4	5
1	---	**.070^**	.049	.048	.070
2	.063	---	.051	.042	.047
3	.106	.135	---	.119	.111
4	.104	.107	.135	---	.107
5	.070	.049	.047	.044	---

[For example, **070^** = 1/((18/6)+(18/8)+(18/2))]

If the distance between the interacting origin and destination, that is, centers 1 and 2, is included in the computation, then, Z_{ij*} becomes Z_{i*} which is a vector of the following values:

$$Z_{i*} = 1/(V_i/\sum_{k=1}^{n} d_{ik}) \qquad (18)$$

Figure 7

Origin Based Distance-type Spatial Structure Variable with the Distance between the Interacting Origin and Destination

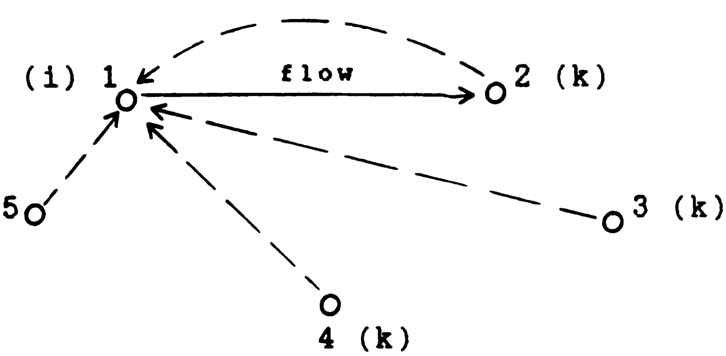


Table 5b

Origin Based Distance-type Accessibility

O&D	Values
1	.043
2	.037
3	.088
4	.084
5	.038

Similarly, the equation of the destination based version of the distance accessibility indices may be stated as follows:

$$P_{ij*} = 1/(W_j/\sum_{\substack{k=1 \\ (k\neq i,j)}}^{n} d_{kj}) \qquad (19)$$

where:

P_{ij*} = distance accessibility of an interacting destination with other

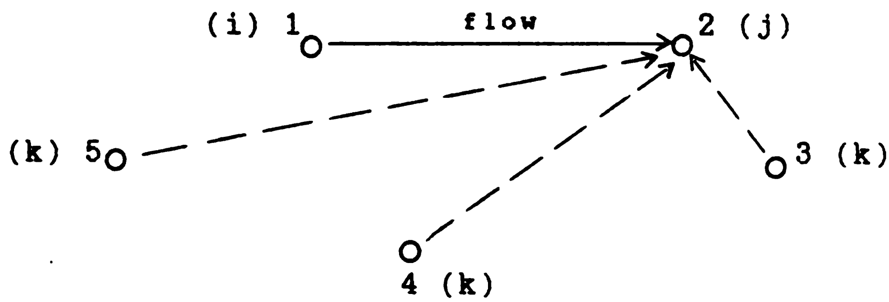
W_j = a weight of the interacting destination,in this case total inflow

d_{kj} = distance between the other destinations and the interacting destination.

n = total locations in the system (5)

Figure 8

Destination Based Distance-type Spatial Structure Variable without the Distance between the Interacting Destination and origin



The matrix values of Equation 15 is presented below:

Table 6a

Destination Based Distance-type Accessibility

O&D	1	2	3	4	5
1	---	**.063^**	.106	.104	.070
2	.070	---	.135	.107	.049
3	.049	.051	---	.135	.047
4	.048	.042	.119	---	.044
5	.070	.047	.111	.107	---

[For example, **.063^** = 1/((22/3)+(22/7)+(22/4))]

If the distance between the interacting centers, that is, locations 1 and 2 is included in the derivation of the *spatial structure* indices, then, P_{ij*} becomes P_{j*} and the matrix collapses to a vector values shown in Table 6b below

$$P_{j*} = 1/(W_j/\sum_{k=1}^{n} d_{ij}) \qquad (20)$$

Figure 9

Destination Based Distance-type Spatial Structure Variable with the Distance between the Interacting Destination and origin

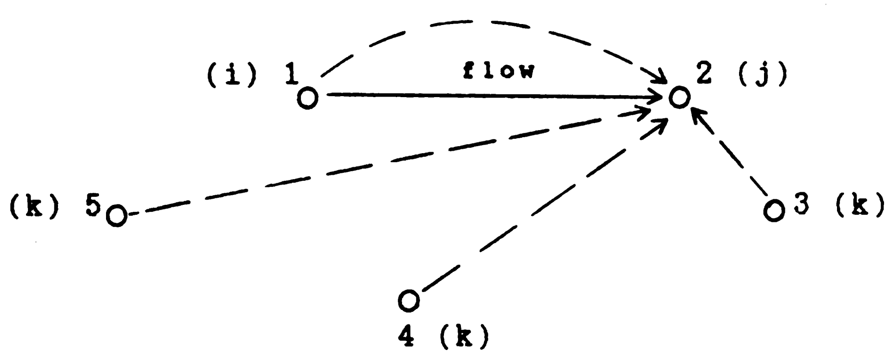


Table 6b

Destination Based Distance-type Accessibility

O&D	1	2	3	4	5
Values	.043	.037	.088	.084	.038

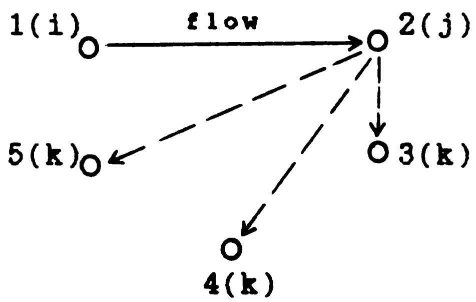
As in the case of the population potential-type *spatial structure*, the origin and destination based distance-type *spatial structure* values (Tables 5a and 5b versus Tables 6a and 6b respectively) are transpose of each other. For close inspection, the figures above are presented together below.

Figure 10

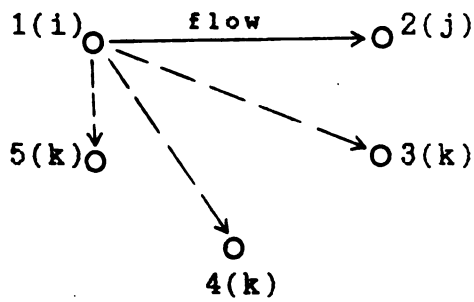
Summary of Spatial Interaction Measures

Destination to destinations accessibility	Origin to destinations

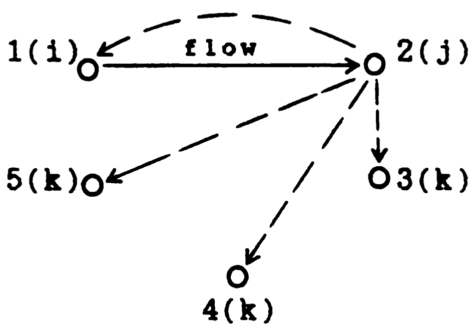
Without the interacting origin (P_{ij})



Without the interacting destination (Z_{ij})



With the interacting origin (P_j)



With the interacting destination (Z_i)

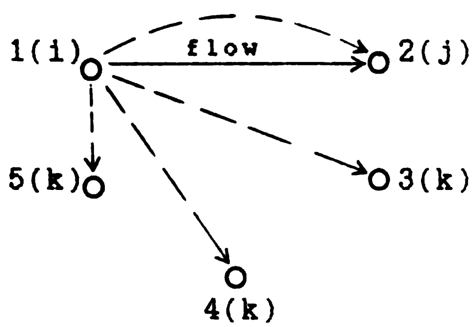
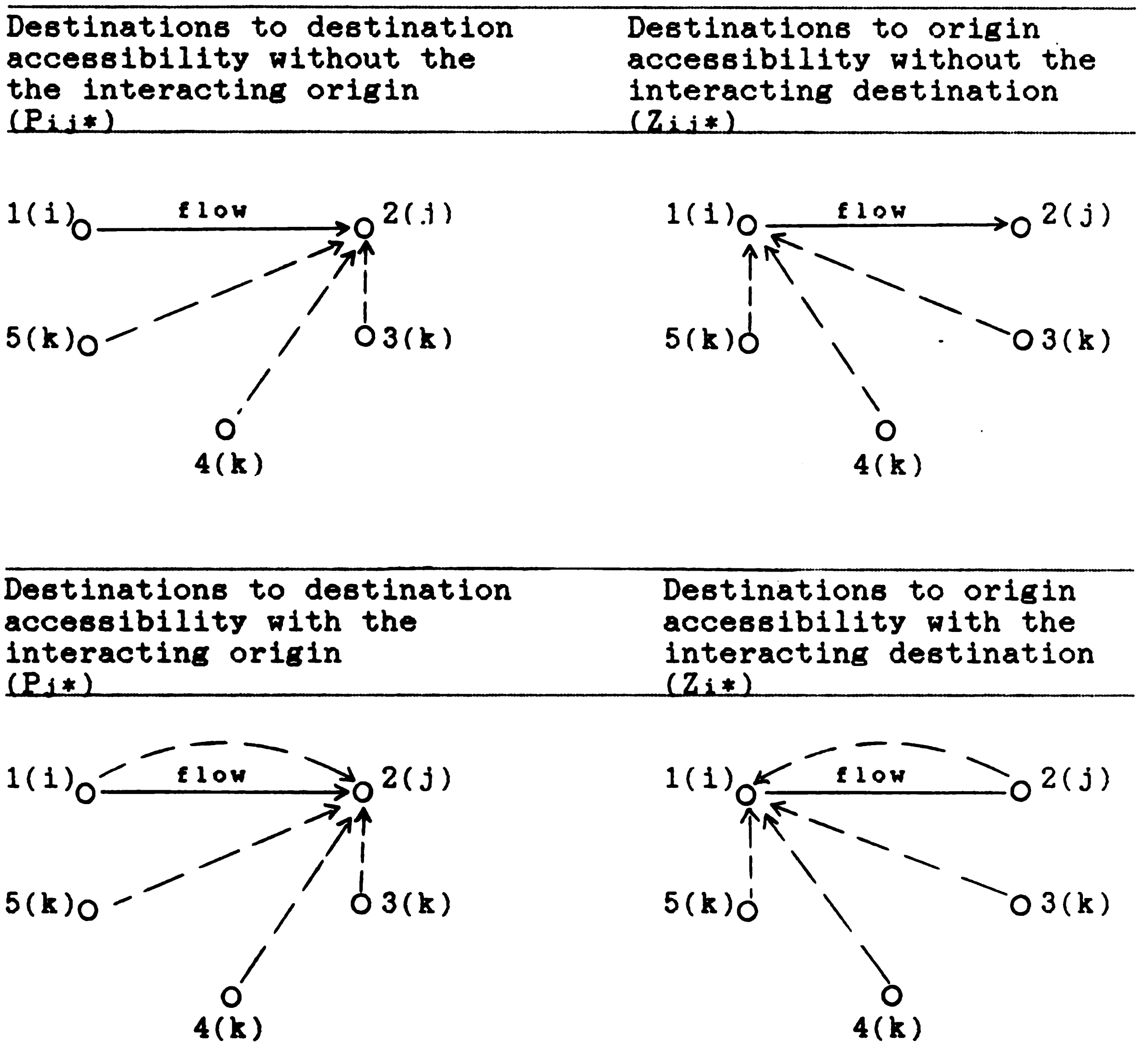


Figure 10 (con´d).



As indicated earlier, the matrices above show that, these pairs: (P_{ij} and Z_{ij}), (P_j and Z_i), (P_{ij*} and Z_{ij*}), and (P_{j*} and Z_{i*}) are identical except that each one of them is a transpose of the other. The transpositional difference between the pairs is a function of the direction of measuring the accessibility of an interacting destination (applied by Fotheringham) and the accessibility of an interacting origin (introduced here).

The vector, P_j, the destination based *spatial structure* measure, is the transpose of A_i, the constraint of the simple production constrained model (Equation 8, p. 11) because Z_i is defined in the same way as A_i in this exercise. The difference between the population potential measure and the destination to destinations or origin to destinations measure is, as indicated above, the number of centers included in the computation of the indices. The transpositional difference between Z_i and P_j however, cannot be removed partly because of the direction of measuring *spatial structure* and partly because vectors Z_i and P_j exclude one center and the matrices, Z_{ij} and P_{ij}, leave out two of them.

If all the points in the network are considered in computing the *spatial structure* values (Equation 12, p.30), all the matrices will become vectors of identical values in every respect implying that the direction of measuring accessibility does not matter. Consequently, instead of relative *origin or destination* accessibilities, there will be relative *location* accessibilities. This would imply that Fotheringham's destination based accessibilities may not be more meaningful than the origin based accessibilities since it is the few excluded points in a given network, while computing *spatial structure* measures, that explain the difference between the two approaches.

If origin to destinations accessibility is interpreted as how the origin relates with the destinations, a reverse measure of destinations to origin accessibility may be similarly interpreted as how the destinations relate themselves with the origin. A composite (*multi-directional*) accessibility measure may, therefore, be a more complete representation of the influence of *spatial structure*.

This study concentrates on observing the effect of these various ways of accounting for *spatial structure* on the *specification* and *prediction capabilities* of the family of the *traditional* gravity models discussed in Chapter 1.

CHAPTER 3.

AN EXPANSION OF THE PROBLEM

3.1. Introduction

The influence of *spatial structure* on spatial interaction has been, as referenced earlier, well known by geographers and others for a long time. However, attempts to include its influence in the calibration of spatial interaction models have not been made until recently (Griffith and Jones, 1980; Fotheringham, 1983; Baxter, 1986). These researchers conclude, as some others have done before them, that the gravity model formulation of spatial interaction needs to consider the influence of *spatial structure* for accurate parameter estimation and for a better model prediction capability.

This study is concerned with analyzing the effect of *spatial structure* on spatial interaction in various types of spatial flows. The major problem of the research, hence, involves calibration of the *traditional* family of spatial interaction models presented in Chapter 1 with and without the various formulations of *spatial structure* variables. The influence of *spatial structure* on spatial interaction is, then evaluated both in terms of model *specification* and model *performance* measured by the *coefficient of variation* of betas over space and *standard error of estimates* respectively. More about methods is presented in Chapter 4.

3.2. Research Questions

A major problem of the *traditional* gravity model is that it does not perform as well when applied to locations different from where it is calibrated. Researchers suggested that this happens because the model is *misspecified* (Curry, 1976; Baxter, 1982; Fotheringham,1983). *Misspecification* occurs, basically, for one or more of three reasons: if a model does not have all the explanatory variables; if a model has more variables than are actually necessary; and/or if a wrong functional form has been applied in the process of calibrating the model (Pindyck and Rubinfeld, 1980; Kennedy, 1984).

In the case of the gravity model, it is usually the first and to a limited extent the third reasons that explain the problem of model *misspecification.* Absence of an explanatory variable when it is needed in a model *misspecifies* the model; and unless and until the nature of the missing variable is known the model´s reliability is suspect (Fotheringham, 1980).

The major research questions addressed in this study are stated hereunder.

3.2.1. Curry (1972) and Gould (1975) write that parameters of the distance variable in the gravity model are indices of accessibility. Fotheringham (1983) states that the degree of *misspecification* of the gravity model depends on the *structure* of the spatial system under consideration. How will introducing *spatial structure* variables affect the *specification* of the *traditional* gravity models across *various spatial systems* of origins and destinations?

3.2.2. If all origins and destinations of a given spatial system are equal in size and are also equidistant from each other, then the population potential-type *spatial structure* measure becomes constant for all the centers. But this is very unlikely. If, only varying sizes of origins and destinations, that is, *intervening opportunities*, are considered, the role of distance, and hence the relative spatial locations of origins and destinations will be totally disregarded. It is conceivable too that the effect of *intervening opportunities* and spatial separation be complementary or competing aspects of spatial interaction (Miller, 1972). How will the distance-type *spatial structure* measure affect the *specification* and *goodness of fit* of the gravity model?

3.2.3. Will the *specification* and *goodness of fit* of the gravity model be different if the *spatial structure* measure considered in the calibration process includes the interacting origin or destination?

3.2.4. The *total flow constrained* and the *production constrained* versions (Equations 6 and 8 respectively, pp. 10 & 11) of the family of spatial interaction models are reported to be the most *misspecified* ones (Fotheringham, 1983). Observation of these models shows that they are different from the *attraction* and *doubly constrained* models because they do not have a destination constraint value. The *total flow constrained* model does not have an origin constraint value either. The *attraction constrained* model, however, has been reported to perform well, although it also does not have a *production constraining* value. It has been suggested that these *origin* and *destination constraints* are also measures of relative *accessibility* or *competitions* (Wilson, 1967) since they are similarly defined as P_j and Z_i (the potential-type destination and origin based accessibilities respectively). What will be the effect of using the *constraints* or *accessibility* values, that is, A_i and B_j, as *spatial structure* measures, on the *specification* and *predictive capability* of the gravity model?

3.2.5. If a model could be fully *specified*, that is, have all the relevant variables, then the parameter estimates are expected to correctly reflect the behavior of spatial interaction, provided calibration of the model is performed with the right functional form, that is, power, exponentiation or a combination of the two (Thomas and Haggett, 1982). But good *specification* of a model may not be necessarily associated with *high goodness of fit* measure (Pindyck, 1980; Fotheringham, 1984). How will the goodness of fit of the models be affected by the addition of *spatial structure* variables?

3.3. Hypotheses

3.3.1. Introduction

As has been pointed out the *traditional* gravity model has been found deficient because it does not take into account the *spatial structure* effect. If this spatially varying component of spatial interaction is captured, the assumption is that it may be possible to follow similar approaches to construct models for use at various places and times. Before recommendations are forwarded, however, the significance of *spatial structure* variable in influencing spatial interaction has to be established. Hence, to evaluate the statistical significance of the questions raised above, the following hypotheses have been formulated with respect to the family of gravity models described in Chapter 1, and the various spatial interaction data presented in Chapter 4.

3.3.2. Major Hypotheses

> The *specification* and *predictive capacity* of the family of spatial interaction models are significantly improved when variables of *spatial structure* are considered in calibrating them.

3.3.3. Sub-hypotheses

3.3.3.1. The significance of the *specification* of the *traditional* family of interaction models is improved for all spatial flows when destination based population potential-type *spatial structure* measures are used in estimating the distance parameter.

The models with the destination based population potential-type *spatial structure* measure, P_{ij}, (Equation 15, p. 36) to be calibrated to test hypothesis 1 are the following:

a) *Total flow constrained* version:

$$T_{ij} = g\, V_i^{\Omega}\, W_j^{\alpha}\, d_{ij}^{\beta}\, P_{ij} \qquad (21)$$

where:

T_{ij} = flow from i to j

$g = T / \sum_{i}^{n}\sum_{j}^{m} V_i^{\Omega}\, W_j^{\alpha}\, d_{ij}^{\beta}\, P_{ij}$ = scale parameter

V_i = origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$P_{ij} = 1/\sum_{\substack{k=1 \\ (k \neq i,j)}}^{n} (W_j/d_{ij})$ = *spatial structure* measure

b) *Production constrained* version:

$$T_{ij} = A_i\, O_i\, W_j^{\alpha}\, d_{ij}^{\beta}\, P_{ij} \qquad (22)$$

where:

T_{ij} = flow from i to j

$A_i = 1/(\sum_{j}^{m} W_j^{\alpha}\, d_{ij}^{\beta}\, P_{ij})$ = scale parameter

O_i = known origin propulsiveness
W_j = destination attractiveness
d_{ij} = distance between i and j

$$P_{ij} = 1/\sum_{\substack{k=1 \\ (k\neq i,j)}}^{n}(W_j/d_{ij}) = \textit{spatial structure}\text{ measure}$$

Since the interest in this exercise is associated with the *distance-decay parameter* (ß), the mass parameters, Ω and α are set to 1.0 and will be ignored from now on. A value of -1 is assigned for all the parameters of the *spatial structure* measures because flow between an origin and destination and accessibility of the interacting destination to other destinations or accessibility of the interacting origin and other destinations is assumed to have inverse relationship. The inverse relationship corrects overprediction and underprediction of the *traditional* gravity models for accessible and inaccessible locations respectively (Fotheringham, 1983).

3.3.3.2. The significance of the *specification* of the *traditional* family of gravity models is improved when a destination based distance-type *spatial structure* measure is considered in estimating the parameters of the models for each data set.

The models to be calibrated to test hypothesis 2 are the following:

a) *Total flow constrained* version:

$$T_{ij} = g\ V_i\ W_j\ d_{ij}^{\beta}\ P_{ij*} \qquad (23)$$

where:

T_{ij} = flow from i to j

$g = T/\sum_{i}^{n}\sum_{j}^{m}(V_i\ W_j\ d_{ij}^{\beta}\ P_{ij*})$ = scale parameter

V_i = origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$P_{ij*} = 1/(W_j/\sum_{k=1\,(k\neq i,j)}^{n} d_{ij})$ = *spatial structure* measure

b) *Production constrained* version:

$$T_{ij} = A_i\ O_i\ W_j\ d_{ij}^{\beta}\ P_{ij*} \qquad (24)$$

where:

T_{ij} = flow from i to j

$A_i = 1/(\sum_{j}^{m} W_j\ d_{ij}^{\beta}\ P_{ij*})$ = scale parameter

O_i = known origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$P_{ij*} = 1/(W_j/\sum_{k=1\,(k\neq i,j)}^{n} d_{ij})$ = *spatial structure* measure

3.3.3.3. The significance of the *specification* of the spatial interaction models also improves when origin based population potential-type *spatial structure* measures are considered in estimating the distance parameter of the models for each data set.

The models to be calibrated to test hypothesis 3 are:

a) *Total flow constrained* version:

$$T_{ij} = g\ V_i\ W_j\ d_{ij}^{\beta}\ Z_{ij} \qquad (25)$$

where:

T_{ij} = flow from i to j

$g = T/\sum_{i}^{n}\sum_{j}^{m} V_i\ W_j\ d_{ij}^{\beta}\ Z_{ij}$ = scale parameter

V_i = origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$Z_{ij} = 1/\sum_{k=1\ (k \neq i,j)}^{n}(W_j/d_{ik})$ = *spatial structure* measure

b) *Production constrained* version:

$$T_{ij} = A_i\ O_i\ W_j\ d_{ij}^{\beta}\ Z_{ij} \qquad (26)$$

where:

T_{ij} = flow from i to j

$A_i = 1/(\sum_{j}^{m} W_j\ d_{ij}^{\beta}\ Z_{ij})$ = scale parameter

O_i = known origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$$Z_{ij} = 1/\sum_{\substack{k=1 \\ (k \neq i,j)}}^{n}(W_j/d_{ij}) = \textit{spatial structure} \text{ measure}$$

3.3.3.4. The significance of the *specification* of the *traditional* family of gravity models is improved when an origin based distance-type *spatial structure* measure is considered in estimating the parameters of the models for each data set.

The models to be calibrated to test hypothesis 4 are:

a) *Total flow constrained* version:

$$T_{ij} = g\ V_i\ W_j\ d_{ij}^{\beta}\ Z_{ij}^{*} \qquad (27)$$

where:

T_{ij} = flow from i to j

$$g = T/\sum_{i}^{n}\sum_{j}^{m} V_i\ W_j\ d_{ij}^{\beta}\ Z_{ij}^{*} = \text{scale parameter}$$

V_i = origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$$Z_{ij}^{*} = 1/(W_j/\sum_{\substack{k=1 \\ (k \neq i,j)}}^{n} d_{ik}) = \textit{spatial structure} \text{ measure}$$

b) *Production constrained* version:

$$T_{ij} = A_i\ O_i\ W_j\ d_{ij}^{\beta}\ Z_{ij}^{*} \qquad (28)$$

where:

T_{ij} = flow from i to j

$A_i = 1/(\sum_{j}^{m} W_j \, d_{ij}^{\beta} \, Z_{ij*})$ = scale parameter

O_i = known origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$Z_{ij*} = 1/(W_j / \sum_{k=1, (k \neq i,j)}^{n} d_{ij})$ = *spatial structure* measure

3.3.3.5. The significance of the *specification* of the family of gravity models is improved when destination based composite *(multi-directional) spatial structure* is used in estimating the parameters of the models.

The models to be calibrated to test hypothesis 5 are:

a) *Total flow constrained* version:

$$T_{ij} = g \, V_i \, W_j \, d_{ij}^{\beta} \, P_{ij} \, P_{ij*} \qquad (29)$$

where:

T_{ij} = flow from i to j

$g = T / \sum_{i}^{n} \sum_{j}^{m} V_i \, W_j \, d_{ij}^{\beta} \, P_{ij} \, P_{ij*}$ = scale parameter

V_i = origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$$P_{ij} = 1/\sum_{\substack{k=1 \\ (k \neq i,j)}}^{n} (W_j/d_{ij}) = \textit{spatial structure} \text{ measure}$$

$$P_{ij*} = 1/(W_j/\sum_{\substack{k=1 \\ (k \neq i,j)}}^{n} d_{ij}) = \textit{spatial structure} \text{ measure}$$

b) *Production constrained* version:

$$T_{ij} = A_i \, O_i \, W_j \, d_{ij}^{\beta} \, P_{ij} \, P_{ij*} \qquad (30)$$

where:

T_{ij} = flow from i to j

$A_i = 1/(\sum_{j}^{m} W_j \, d_{ij}^{\beta} \, P_{ij} \, P_{ij*})$ = scale parameter

O_i = known origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$$P_{ij} = 1/\sum_{\substack{k=1 \\ (k \neq i,j)}}^{n} (W_j/d_{ij}) = \textit{spatial structure} \text{ measure}$$

$$P_{ij*} = 1/(W_j/\sum_{\substack{k=1 \\ (k \neq i,j)}}^{n} d_{ij}) = \textit{spatial structure} \text{ measure}$$

3.3.3.6. The significance of the *specification* of the family of gravity models is improved when origin based composite *multi-directional* instead of a *uni-directional spatial structure* measure is used in estimating the parameters of the models.

The models to be calibrated to test hypothesis 6 are:

a) *Total flow constrained* version:

$$T_{ij} = g\ V_i\ W_j\ d_{ij}^{\beta}\ Z_{ij}\ Z_{ij}* \qquad (31)$$

where:

T_{ij} = flow from i to j

$g = T/\sum_{i}^{n}\sum_{j}^{m} V_i\ W_j\ d_{ij}^{\beta}\ Z_{ij}\ Z_{ij}*$ = scale parameter

V_i = origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$Z_{ij} = 1/\sum_{k=1\ (k \neq i,j)}^{n}(W_j/d_{ij})$ = *spatial structure* measure

$Z_{ij}* = 1/(W_j/\sum_{k=1\ (k \neq i,j)}^{n} d_{ij})$ = *spatial structure* measure

b) *Production constrained* version:

$$T_{ij} = A_i\ O_i\ W_j\ d_{ij}^{\beta}\ Z_{ij}\ Z_{ij}* \qquad (32)$$

where:

T_{ij} = flow from i to j

$A_i = 1/(\sum_{j}^{m} W_j\ d_{ij}^{\beta}\ Z_{ij}\ Z_{ij}*)$ = scale parameter

O_i = known origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$Z_{ij} = 1/\sum_{k=1\ (k \neq i,j)}^{n}(W_j/d_{ij})$ = *spatial structure* measure

$Z_{ij}* = 1/(W_j/\sum_{k=1\ (k \neq i,j)}^{n} d_{ij})$ = *spatial structure* measure

3.3.3.7. The *specification* of the family of the spatial interaction models is best when the specific origins and destinations (in a system of origins and destinations), and the distance between the interacting origin and destination are considered in the measurement of *spatial structure* and model parameters estimated.

The *total flow constrained* models to be calibrated to test hypothesis 7 are the following. The *production constrained* versions which could be stated similarly have also been calibrated.

a) With origin based population potential-type *spatial structure measure*

$$T_{ij} = g\, V_i\, W_j\, d_{ij}^{\beta}\, Z_i \qquad (33)$$

where:

T_{ij} = flow from i to j

$g = T / \sum_{i}^{n} \sum_{j}^{m} V_i\, W_j\, d_{ij}^{\beta}\, Z_i$ = scale parameter

V_i = origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$Z_i = 1 / \left(\sum_{k=1\,(k \neq i)}^{n} W_k / d_{ik} \right)$ = *spatial structure* measure

b) With origin based *spatial structure* measure of only the distance-type *spatial structure* measure

$$T_{ij} = g\ V_i\ W_j\ d_{ij}^{\beta}\ Z_i^* \qquad (34)$$

where:

T_{ij} = flow from i to j

$g = T/\sum_{i}^{n}\sum_{j}^{m} V_i\ W_j\ d_{ij}^{\beta}\ Z_i^*$ = scale parameter

V_i = origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$Z_i^* = 1/(V_i/\sum_{k=1}^{n} d_{ik})$ = *spatial structure* measure

c) With destination based population potential-type *spatial structure* measure

$$T_{ij} = g\ V_i\ W_j\ d_{ij}^{\beta}\ P_j \qquad (35)$$

where:

T_{ij} = flow from i to j

$g = T/\sum_{i}^{n}\sum_{j}^{m} V_i\ W_j\ d_{ij}^{\beta}\ P_j$ = scale parameter

V_i = origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$P_j = 1/\sum_{\substack{k=1 \\ (k \neq j)}}^{n} (W_k/d_{jk})$ = *spatial structure* measure

d) With destination based *spatial structure* measure of only the distance-type *spatial structure* measure :

$$T_{ij} = g\ V_i\ W_j\ d_{ij}^{\beta}\ P_j^* \qquad (36)$$

where:

T_{ij} = flow from i to j

$g = T/\sum_{i}^{n}\sum_{j}^{m} V_i\ W_j\ d_{ij}^{\beta}\ P_j^*$ = scale parameter

V_i = origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$P_j^* = 1/(W_j/\sum_{k=1}^{n} d_{jk})$ = *spatial structure* measure

3.3.3.8. The *specification* of the *traditional* gravity models are significantly improved if they are calibrated with the constraints of the *production* and *attraction* constrained models.

The models to be calibrated to test hypothesis 8 are the following:

a) *Total flow constrained* version:

$$T_{ij} = g\ V_i\ W_j\ d_{ij}^{\beta}\ A_i\ B_j \qquad (37)$$

where:

T_{ij} = flow from i to j

$g = T / \sum_{i}^{n} \sum_{j}^{m} V_i W_j d_{ij}^{\beta} A_i B_j$ = scale parameter

V_i = origin propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

$A_i = 1/(\sum_{j}^{m} W_j d_{ij}^{\beta})$ = *spatial structure* measure

$B_j = 1/(\sum_{i}^{n} V_i d_{ij}^{\beta})$ = *spatial structure* measure

a) *Production constrained* version:

$$T_{ij} = A_i O_i W_j d_{ij}^{\beta} B_j \qquad (38)$$

where:

T_{ij} = flow from i to j

$A_i = 1/(\sum_{j}^{m} W_j d_{ij}^{\beta} B_j)$ = scale parameter

O_i = known origin propulsiveness

$B_j = 1/(\sum_{i}^{n} V_i d_{ij}^{\beta})$ = *spatial structure* measure

W_j = destination attractiveness

d_{ij} = distance between i and j

3.3.3.9. A gravity model with origin or destination based *spatial structure* measure results in the same significant *goodness of fit* of predicted flows.

The beta values of the models stated above are estimated following the procedure described in the following chapter. Evaluation and analysis of the hypotheses are then presented in Chapter 5.

CHAPTER 4.

DATA AND METHODS OF ANALYSIS

4.1. The Data

4.1.1. Introduction

Model building is based on a number of assumptions. In order for the model to have wide applicability, the assumptions have to be evaluated using various data sets. If, on repeated experiments, the results of the model are consistently validating the assumptions, then, generalizations about certain events can be made safely.

It is desirable that models be understandable with minimum effort. Models are also expected to account for *causality.* To produce a useful model, probably, the best strategy is to subject it to various types of data and observe its performance.

In this study, four data sets are used; two of them are from the State of Michigan, and two are from Ethiopia. The diversity of these data sets in type and place of origin, is believed to give the opportunity for comparing the effect of *spatial structure* on spatial interaction.

4.1.2. The Michigan Vehicle Flow Data

Vehicle flow data for 25 major cities (Appendix 1a) in the State of Michigan were obtained from the Michigan Department of Transportation. Some adjustments had been made to the data in order to bring them to the common year of 1975. This year was chosen because many other surveys had been made during this time. The 1975 annual vehicle flow in the state was used as a weight to transform the data which forms a complete 25 by 25 origin and destination flow matrix. The distance matrix accompanying this data has been taken from an intercity distance map of the State of Michigan published in 1978. It reports distances along the shortest road path between pairs of cities.

4.1.3. The Michigan Public Passenger Flow Data

These data are the average daily flow of bus passengers between 15 cities (Appendix 1b) within the State of Michigan for a typical month of 1985. The source of the data is, again, the State of Michigan Department of Transportation. The data forms a 15 by 15 origin and destination flow matrix. The distance matrix is taken from the same source used for the vehicle flow data.

4.1.4. The Ethiopian Rail Passengers Flow Data

These are intercity flows for 22 cities (Appendix 1c) in Ethiopia for 1982. The flow of passengers here is along a rail line and, historically, the cities were established as a result of the railway line. The distance matrix is, therefore, a combination of one or more of the segments making up the separation between two cities along the line.

4.1.5. The Ethiopian Telephone Flow Data

The telephone-call data are for 13 major cities, regional capitals, (Appendix 1d) for 1984. These flows are different from the rest in that they may be influenced very little by neighborhood effects. Yet, they are interaction data and the gravity formulation has been applied to predict their flow volumes. The distance matrix is the shortest distance along main road routes.

Both the rail passengers and the telephone-call data are obtained from publications of the Ministry of Transport and Communication of Ethiopia.

4.2. Methods of Analysis

The usual statistical methods applied in interactional studies are various forms of multivariate approaches. The most common ones are regression and category analysis (Ewing, 1974; Stetzer, 1976; Miller and Mayer, 1984; Baxter, 1979). While the *total flow constrained* version of the family of spatial interaction models is often calibrated with the *ordinary least squares* approach without any serious difficulty, the parameters of the *production, attraction* and *doubly constrained* models are best estimated using *maximum likelihood* estimators (Wilson, 1971; Oppenheim, 1979).

In this study, the distance exponents of, especially, the origin specific models are estimated using a special computer program written for this purpose because insufficient number of observations limited the use of the regression approach. The program is based on Baxter's simple program designed for system wide calibration of the basic family of gravity models (1976). Using the computer program, the best distance exponents have been derived from a range of values through an intensive search process as described below.

Each data set is subjected to a search process to derive the best distance exponent for each point in each system. The *best beta* is that distance *exponent* which gives the *best goodness of fit* as measured by the standard-root-mean-square-error, SRMSE (Knudsen and Fotheringham 1986, p. 132). The formula of the SRMSE is:

$$\{\sum_i\sum_j(t_{ij} - \bar{t}_{ij})^2/m \times n\}^{1/2}/(\sum_i\sum_j t_{ij}/m \times n) \qquad (39)$$

where:

t_{ij} = matrix of observed flows
$\bar{t}_{ij}$ = matrix of predicted flows
$m \times n$ = dimension of matrix

In order to derive the *best beta*, a range of values extending between -5.00 and 2.00 is used. (This range has been arrived at in a preliminary experiment conducted on a sample of the data sets. About 90% of the runs fell within this range). The computer program searches through the range beginning at -5.00 by a step of 0.01, and when it reaches 2.00, it prints out the *best beta.* Between -5.00 and 2.00 there are 700 runs at an increment of 0.01. The *best beta* is one of these 700 distance exponents which results in the *best goodness of fit* (i.e., lowest SRMSE). When the *best exponent* happens to be outside the given range, reruns are done until it is obtained. In order to obtain an origin specific distance exponent, the above procedure is repeated for all the cities (nodes) of the various data sets. Thus, over five and a half million computer runs have been processed to come up with the *best* *beta* and *goodness of fit* values for all the various flows described above.

It is possible to improve the fit of the model by changing the increment, for example, to 0.001. The problem with reducing the increment is, the increased length of time the computer requires to identify the *best beta.* Instead of 700 runs as in the case of 0.01 increment within the same range, a total of 7000 model runs are required to get the best distance exponent for one city. The time is increased even more when calibrating a *production, attraction*, or *doubly constrained* model. The longest time is associated with the *doubly constrained* model since it needs to converge the balancing factors by an iterative process. In general, the computer gives quick results if models have few variables, are not constrained, and the matrix order is small.

The SRMSE of estimates is preferred to such *goodness of fit* measures as the Chi-square, Absolute Entropy Difference, Phi Statistics, Coefficient of Determination and Absolute Psi since it has been shown to be the best of them (Fotheringham and Knudsen, 1986), especially, for models of spatial interaction. All the others are produced by the program, however, optimality is determined by the SRMSE. The *goodness of fit* of each model with a variable or variables of *spatial structure* is then compared with the *goodness of fit* of the *traditional* spatial interaction model, hereafter, referred to as the *basic model* (the model without any of the designated *spatial structure* variables).

The effect of *spatial structure* on the spatial distribution of the beta values is assessed with measures of *coefficient of variation* (C.V.). The distance exponents (beta values) derived from calibrating the spatial interaction models with *spatial structure* variables are compared with the beta of the *basic model* by way of *coefficient of variations*. If all the *distance exponents* are similar over space, the *coefficient of variation* becomes zero. Higher *coefficient of variations* are, therefore, indications of higher variations in the distribution of betas. *A spatial interaction model with high coefficient of variation is considered misspecified because the distance exponent is not measuring spatial separation alone. It is also representing the effect of spatial structure variable* (Curry, 1972; Gould, 1975; Fotheringham, 1983). The exercise here is to see if the introduction of a *spatial structure* measure significantly affects the *specification* of the *traditional* spatial interaction models by way of reducing the C.V. of the betas.

The *significance* of the difference between two *coefficient of variations* is determined using the following formula introduced by Gregory (1963, p. 130) and applied in interaction model assessments by Fotheringham (1983, p. 29):

$$S.E. = |V_1 - V_2| / [(V_1^2/2n_1) + (V_2^2/2n_2)]^{1/2} \quad (40)$$

where:

V_1 = coefficient of variation of the first case
V_2 = coefficient of variation of the second case
n_1 = observation of the first case
n_2 = observation of the second case

Regression approach has been also used to see model significance for system-wide calibrations. Although the level of *goodness of fit* may be observed from the method applied for origin specific calibrations, it is not possible to see the improvement in model *specification* since a single beta value is estimated for all the centers of a given system of origins and destinations. The regression approach renders the possibility of appreciating the importance of an included variable by way of increasing the adjusted *coefficient of determination* (R^2).

CHAPTER 5.

RESULTS AND ANALYSIS

5.1. Introduction

The empirical results are presented in this chapter. In all cases, the estimation process is done according to the procedure explained in Chapter 4. Each data set is subjected to the same standard of calibration to derive the beta values which give the *best goodness of fit* measured by SRMSE for each origin of each data set.

Since the problem of *misspecification* and *prediction capability* are reported to be more pronounced in the *total flow constrained* (Equation 6, p. 10) and *production constrained* (Equation 8, p. 11) versions (Fotheringham, 1983), this study also gives more attention to these models. However, reference will also be made to the *doubly constrained* model (Equation 10, p. 12) for part of the discussion.

5.2. The Origin Specific Total Flow Constrained Model

This model is given as:

$$T_{ij} = g\ W_j\ d_{ij}^{\beta_i} \qquad (41)$$

where:

T_{ij} = interaction between T_{ij}

$g = T / \sum_{i}^{n}\sum_{j}^{m} V_i\ W_j\ d_{ij}^{\beta_i}$

W_j = weight of destination

d_{ij} = distance between i and j

$T = \sum_{i}^{n}\sum_{j}^{m} T_{ij}$

The *propulsiveness* variable (V_i) is usually ignored since it is a constant subsumed in the g. The beta value estimated by the computer search process is specific to a single origin. Parameters Ω and α are as stated in Chapter 3, set to 1 and hence forced not to vary.

5.2.1. The Telephone Data

Table 7 shows a summary of the *coefficient of variations* of the estimated beta values along with the *goodness of fit* (SRMSE) measures. The models that resulted in significantly improved goodness of fit are only two:

$$(a)\ T_{ij} = g\ W_j\ d_{ij}{}^{\beta_i}\ P_{ij}* \tag{42}$$

$$(b)\ T_{ij} = g\ W_j\ d_{ij}{}^{\beta_i}\ P_j* \tag{43}$$

$P_{ij}*$ and P_j* as defined in Equations 19 (p. 41) and 20 (p. 43) are destination based distance-type configurational measures. The *average-standardized-root-mean-square-error*, (ASRMSE) of the models have been reduced from 27.98 to 21.28 in the case of model 42 and to 21.17 in the case of model 43. The two models are very similar because the difference between $P_{ij}*$ and P_j* is the excluded or included distance which lies between the interacting origin and destination. thus, the correlation between $P_{ij}*$ and P_j* is, normally, expected to be high.

In Table 7, none of the models with the *spatial structure* measure of one sort or another resulted in lower *coefficient of variation* to make them signifcantly better specified than the *basic* model. However, the models with variables $Z_{ij}*$, P_j* and the composite *bi-directional* variable (P_j,P_j*) are better *specified* than the *basic model* because the *coefficient of variation* associated with the models with these variables, 1.15, 0.90, and 1.08 respectively are lower than the *coefficient of variation* of the *basic model* which is 1.23. The *spatial structure* measure which best *specified* the *basic model* is, however, variable P_j* although not statistically significant. In fact, the model with P_j* is the best in terms of *goodness of fit* too.

Table 7

Origin Specific Total Flow Constrained Model Calibration of the Telephone Data.

Model		C.V.	G.F.	Ma.B.	Mi.B.	N.B.	P.B.
Basic		1.23	27.98	4.73	-1.39	2	11
with	Z_{ij}	1.80	31.80	5.85	-3.99	3	10
	$Z_{ij}*$	1.15	28.05	4.79	-1.22	2	11
	P_{ij}	1.42	32.16	5.15	-1.33	2	11
	P_j	1.24	31.77	7.19	-1.00	1	12
	$P_{ij}*$	1.34	21.28^	0.22	-0.93	10	3
	P_j*	0.90	21.17^	0.11	-0.94	10	3
	C_j	1.54	31.39	4.95	-1.95	2	11
	$P_{ij}, P_{ij}*$	1.89	22.41	0.68	-1.41	10	3
	P_j, P_j*	1.08	22.17	0.47	-1.44	10	3
	$Z_{ij}, Z_{ij}*$	2.31	32.65	4.19	-3.85	10	3

The lower the C.V.(coefficient of variation) the better
The lower the G.F.(goodness of fit) the better
^ = significantly better than basic at 0.05 level
Ma.B. = maximum beta
Mi.B. = minimum beta
N.B. = total number of centers with negative betas
P.B. = total number of centers with positive betas

In the origin specific *total flow constrained* model of the telephone flow data, the hypothesis of significant improvement in model *specification* has not been achieved. Variables $P_{ij}*$ and P_j* have, however, significantly improved the *goodness of fit* of the *basic model.*

5.2.2. The Bus Data

The results from the bus data are summarized in Table 8. Three models have better *goodness of fit* than the *basic model.* They are models with $P_{ij}*$ and P_j* (the distance-type *spatial structure* measures) and the following with origin based distance-type *spatial structure* variable.

$$T_{ij} = g\ W_j\ d_{ij}^{\beta_i}\ Z_{ij}* \qquad (44)$$

However, it is the models with variables $P_{ij}*$ and P_j* which resulted in significant improvement in *goodness of fit.* The *goodness of fit* measures due to variables $P_{ij}*$ and P_j* are almost identical for the reason explained earlier.

The *spatial structure* variable that significantly improved the *goodness of fit* of the basic model also significantly improved its *specification.* Significant improvement in model *specification* are also attained when the destination based accessibility measures, P_j and P_{ij}

Table 8

.Origin Specific Total Flow Constrained Model
Calibration of the Bus Data

Model		C.V.	G.F.	Ma.B.	Mi.B.	N.B.	P.B.
Basic		6.36	82.21	0.96	-0.87	11	4
With	Z_{ij}	19.15	82.92	1.30	-0.88	9	6
	Z_{ij*}	8.11	82.18	1.05	-0.79	7	8
	P_{ij}	3.95^	88.90	1.03	-1.23	11	4
	P_j	2.34^	88.74	0.93	-1.24	11	4
	P_{ij*}	1.46^	75.00^	0.12	-0.64	11	4
	P_{j*}	0.74^	74.62^	0.01	-0.67	14	1
	C_j	2.45^	90.13	1.15	-1.52	4	11
	P_{ij},P_{ij*}	1.88^	83.07	0.44	-1.17	11	4
	P_j,P_{j*}	1.88^	82.29	0.66	-1.25	13	2
	Z_{ij},Z_{ij*}	4.93	83.92	1.39	-0.80	8	7

(with and without the interacting origin respectively) are included in the calibration of the basic model. The best *specification* is provided by the model with variable P_{j*}. The *coefficient of variation* is reduced to 0.74 from 6.36.

Including B_j (measured in this exercise by C_j), the constraint of the *attraction constrained* model (Equation 9, p.12), which may be considered, as mentioned earlier, as a measure of competition, and the pairs (P_{ij},P_{ij*}) and

(P_j,P_j*) also give significantly improved *specifications*. The pair of variables represent composite *bi-directional* measure (the population potential-type and the distance-type *spatial structure* variables together).

C_j is different from the *production constraint* value, A_i (Equation 8, p.11) because it uses the *propulsiveness* instead of the *attractiveness* values as a measure of mass of a center. Thus, the origin specific model with the *attraction constraint* (a variable in this case) may be given as:

$$T_{ij} = g\ W_j\ d_{ij}^{\beta_i}\ C_j \qquad (45)$$

where:

$$C_j = 1/(\sum_i^n V_i\ d_{ij}^{\beta})$$

The *coefficient of variation* of Equation 45 is 2.45. It is better *specified* than the model with P_{ij} (the destination based potential-type *spatial structure* measure without the interacting origin). But the *goodness of fit* of Equation 45 is worse than the *basic model.* So also are the equations with variables P_{ij} and P_j.

Here, the hypothesis of significant improvement in *goodness of fit* as well as *specification* has been found valid for models with the distance-type *spatial structure* measures, that is, variables P_{ij*} and P_{j*}. The others, P_{ij}, P_j, C_j, and the composites (P_{ij},P_{ij*}) and (P_j,P_{j*}) are found important for only *specification* purpose.

5.2.3. The Railway Data

The spatial distribution of the centers of this data is different from the rest of the data sets. Here, all of the centers are strictly along the railway line (Appendix 1c). The treatment of this data is, however, the same as the others. As shown in Table 9, the models with the same *spatial structure* measures, P_{ij*} and P_{j*} (Equations 42 and 43 respectively, p. 78) have significantly better *goodness of fit* than the *basic model.* The pairs of composite variables (P_{ij},P_{ij*}) and (P_j,P_{j*}) which are actually *multi-directional* measures also give significant *goodness of fit.* As the correlation between P_{ij*} and P_{j*} is high, the difference between the *goodness of fit* of the models with these *spatial structure* measures is small. In this case, in fact, the *goodness of fit* measures of the models are identical (71.79). Similar situation is observed in the case of the *goodness of fit* of the composite *multi-directional* variables. If the precision level is increased to more than two decimal places, differences however insignificant will be observed. The *goodness of fit* of the model with the origin based distance-type *spatial structure* measure, Z_{ij*} is only marginally better than the *goodness of fit* of the *basic model.*

Unlike the previous cases, it is the *specification* of the *basic model* which turned out to be the best. However, the *coefficient of variation* of this model (1.13) is not significantly better than the *coefficient of variation* of the models with the destination based *spatial structure* variables, P_j (1.14) and P_{ij} (1.38) and also the models with variables Z_{ij*} (1.36) and C_j (1.18).

In this data set, there is no overlap between models of the significantly better *goodness of fit* and the *basic model* which, in this particular case, is the *best specified.* This would mean that good predictions are attainable if *spatial structure* variables like P_{ij*} or P_{j*} are considered but at the expense of *badly specified* models. This is because the *coefficient of variation* of the models with the same *spatial structure* variables shows that the betas of these models are highly spatially variable and hence are very *badly specified.*

Table 9

Origin Specific Total Flow Constrained Model
Calibration of the Railway Data

Model		C.V.	G.F.	Ma.B.	Mi.B.	N.B.	P.B.
Basic		1.13	113.43	0.20	-1.06	18	4
With	Z_{ij}	2.90	121.01	0.79	-1.08	13	9
	$Z_{ij}*$	1.36	113.19	0.40	-0.98	17	5
	P_{ij}	1.38	127.31	0.34	-1.07	15	7
	P_j	1.14	126.88	0.21	-1.16	16	6
	$P_{ij}*$	4.14	71.79^	0.31	-1.01	3	19
	P_j*	3.06	71.79^	0.26	-1.09	10	12
	C_j	1.18	115.78	0.16	-1.05	16	6
	$P_{ij}, P_{ij}*$	2.15	79.26^	2.73	-0.19	7	15
	P_j, P_j*	8.87	79.01^	2.63	-1.50	10	12
	$Z_{ij}, Z_{ij}*$	4.43	120.85	0.98	-1.00	12	10

The hypothesis of significantly better model *specification* than the *specification* of the *basic model* is, therefore, not confirmed in this specific case. Variables $P_{ij}*$ and P_j* are accepted only for improving the *goodness of fit* of the *basic model*.

5.2.4. The Vehicle Data

In this case also the models with the destination based *spatial structure* variables (P_{ij*} and P_{j*}) resulted in the best and significant improvement in *prediction power.* Other models with variables P_{ij}, P_j, the pairs (P_{ij}.P_{ij*}), (P_j,P_{j*}), and C_j are also significantly better in their capability of *replicating* the data they are calibrated with than the *basic model*.

The model with the origin based *spatial structure* measure of only the distance variable, Z_{ij*}, has a lower *coefficient of variation* than the *basic model* but not significantly so to be accepted under the *specified* hypotheses. The models that gave significantly improved fit are also significantly *better specified* than the *basic model.* Variable P_{j*} is the most effective in *specifying* the model and also giving the *best goodness of fit*.

The hypothesis of significantly better *specification* is thus true for models with destination based *spatial structure* variables including variable C_j. The *goodness of fit* of the same models has also shown significant improvement over the *goodness of fit* of the *basic model.*

Table 10

Origin Specific Total Flow Constrained Model
Calibration of the Vehicle Data

Model		C.V.	G.F.	Ma.B.	Mi.B.	N.B.	P.B.
Basic		4.68	168.57	1.87	-2.25	18	7
With	Z_{ij}	6.74	171.82	1.95	-2.26	15	10
	$Z_{ij}*$	3.69	168.66	1.89	-2.18	17	8
	P_{ij}	3.21^	160.48^	0.83	-1.75	17	8
	P_j	2.51^	160.54^	0.81	-1.76	19	6
	$P_{ij}*$	2.46^	139.28^	1.02	-5.00	21	4
	P_j*	1.73^	138.73^	0.94	-0.76	21	4
	C_j	2.40^	162.10^	0.84	-1.81	18	7
	P_{ij},$P_{ij}*$	2.61^	170.26	2.90	-1.49	21	4
	P_j,P_j*	2.26^	168.95	2.82	-1.52	21	4
	Z_{ij},$Z_{ij}*$	8.34	171.93	2.63	-2.26	16	9

5.3. The Origin Specific Production Constrained Model

The basic origin specific version of the *production constrained* model is given as:

$$T_{ij} = A_i \, O_i \, W_j \, d_{ij}^{\beta_i} \qquad (46)$$

where:

T_{ij} = flow from i to j

$A_i = 1/(\sum_{j}^{n} W_j\, d_{ij}^{\beta})$ = scale parameter

O_i = known production propulsiveness

W_j = destination attractiveness

d_{ij} = distance between i and j

5.3.1. The Telephone Data

The results from estimation of the distance exponent using the *production constrained* model and the telephone data are summarized in Table 11.

Unlike in the case of the *total flow constrained* estimates, the destination based distance-type *spatial structure* measures variables P_{ij*} and P_{j*} (Equation 42 and 43, p.78) have not been found to improve the *production constrained* model calibration. The origin based distance-type *spatial structure measure* has not been important either. For this data set and models, the best *goodness of fit* comes from using the origin based population potential-type *spatial structure* measure, Z_{ij}. The composite variable (Z_{ij},Z_{ij*}) is also a significant contributor; so also are variables P_{ij}, P_j and C_j. The model statements that have shown improvements discussed above are the following:

$$T_{ij} = A_i\, O_i\, W_j\, d_{ij}^{\beta_i}\, Z_{ij} \qquad (47)$$

$$T_{ij} = A_i\ O_i\ W_j\ d_{ij}^{\beta_i}\ Z_{ij}\ Z_{ij}* \qquad (48)$$

$$T_{ij} = A_i\ O_i\ W_j\ d_{ij}^{\beta_i}\ C_j \qquad (49)$$

$$T_{ij} = A_i\ O_i\ W_j\ d_{ij}^{\beta_i}\ P_{ij} \qquad (50)$$

$$T_{ij} = A_i\ O_i\ W_j\ d_{ij}^{\beta_i}\ P_j \qquad (51)$$

In this case, the *spatial structure* variables identified for significantly improving the *goodness of fit* also significantly improved the *specification* of the same models (with the exception of the origin based composite variable). The best *specified* model, that is the model with variable C_j is not, however, the best predictor. Its *coefficient of variation* drops to 0.56 from 0.78.

Table 11

Origin Specific Production Constrained Model
Calibration of the Telephone Data

Model		C.V.	G.F.	Ma.B.	Mi.B.	N.B.	P.B.
Basic		0.78	0.56	-0.22	-5.32	13	0
With	Z_{ij}	0.72^	0.29^	-0.11	-2.73	13	0
	$Z_{ij}*$	0.84	0.57	-0.13	-5.20	13	0
	P_{ij}	0.70^	0.36^	0.12	-3.58	12	1
	P_j	0.63^	0.37^	-0.02	-3.65	13	0
	$P_{ij}*$	1.18	1.71	-0.19	-8.85	13	0
	P_j*	1.05	2.08	-0.25	-7.48	13	0
	C_j	0.56^	0.39^	-0.01	-3.85	13	0
	P_{ij},$P_{ij}*$	1.19	1.77	0.10	-8.17	13	0
	P_j,P_j*	1.02	1.92	-0.17	-9.00	13	0
	Z_{ij},$Z_{ij}*$	0.80	0.30^	-0.03	-2.62	13	0

The hypothesis of significantly better model *specification* because of considering the attraction constraint, B_j, measured by C_j has been found to be true here. The *goodness of fit* of the same model is also significantly improved. Further, the origin based population potential-type *spatial structure* measure, Z_{ij} has also resulted in significantly improved *goodness of fit*. Although the *specification* of the basic model with variable

Z_{ij} has not been improved significantly, the change for the better is appreciable enough to receive serious attention.

5.3.2. The Bus Data

There is not a single *spatial structure* variable that significantly contributed to the *goodness of fit* of the basic model here (Table 12). Not only are the models with the additional *spatial structure* variables deteriorating in their predictive capacity, none, except the one with variable P_{ij}, has even a marginally better *goodness of fit*.

Variable P_{j*} reappears as a significantly better *specifier* of the *basic model* although the best *specified* model is the one with the pair variable (P_j,P_{j*}). Variables P_{ij}, P_j , P_{ij*} and C_j also contribute to the improvement of the *basic model* but not significantly.

Table 12

Origin Specific Production Constrained Model
Calibration of the Bus Data

Model		C.V.	G.F.	Ma.B.	Mi.B.	N.B.	P.B.
Basic		1.17	0.59	1.20	-1.41	13	2
With	Z_{ij}	1.59	0.68	1.27	-1.27	12	3
	$Z_{ij}*$	1.32	0.93	1.29	-1.90	13	2
	P_{ij}	0.99	0.56	0.69	-2.22	12	3
	P_j	0.84	0.87	0.54	-2.23	12	3
	$P_{ij}*$	0.81	1.30	0.02	-2.34	14	1
	P_j*	0.67^	1.32	-0.16	-2.44	15	0
	C_j	0.85	0.85	0.47	-2.37	12	3
	P_{ij},$P_{ij}*$	0.82	1.45	0.14	-2.69	14	1
	P_j,P_j*	0.56^	1.41	-0.16	-2.80	15	0
	Z_{ij},$Z_{ij}*$	2.17	0.83	1.39	-1.24	12	3

The hypothesis of model improvement is, hence, accepted for (P_j,P_j*), and P_j* *spatial structure* measures and that only for better *specifying* the model. Only variable P_{ij} has a marginally better *goodness of fit* than the *basic model*.

5.3.3. The Railway Data

In the case of the railway data, the best *goodness of fit* of the *production constrained* model is attained without including any of the *spatial structure* variables (Table 13). But the *coefficient of variation* is significantly improved with the addition of a number of them.

Table 13

Origin Specific Production Constrained Model
Calibration of the Railway Data

Model		C.V.	G.F.	Ma.B.	Mi.B.	N.B.	P.B.
Basic		7.22	0.94	13.52	-1.98	14	8
With	Z_{ij}	7.39	1.04	11.77	-2.27	13	9
	Z_{ij*}	6.07	0.98	13.53	-1.76	14	8
	P_{ij}	67.78	1.96	8.33	-2.57	14	8
	P_j	28.04	1.05	8.35	-2.58	15	7
	P_{ij*}	2.93^	*1.27	>100.00	-1.22	13	9
	P_{j*}	2.09^	1.35	8.50	-1.33	14	8
	C_j	12.01	0.96	10.62	-2.46	14	8
	P_{ij},P_{ij*}	1.84^	1.11	11.50	-1.19	14	8
	P_j,P_{j*}	2.47^	1.02	33.75	3.22	18	4
	Z_{ij},Z_{ij*}	5.89	1.05	11.78	-2.05	14	8

This time it is the composite variable (P_{ij},P_{ij*}) which *specify* the *basic model* best. Variables P_{j*}, P_{ij*} and the pair (P_j,P_{j*}) also significantly improve the *specification* of the *basic model*. Although not significantly, variables Z_{ij*}, and the pair (Z_{ij},Z_{ij*}) are also better *specified*.

Although its *specification* is significantly better than the *basic model*, the *coefficient of variation* associated with the destination based distance-type *spatial structure* is not the result of the best beta estimates. The best beta of one of the centers could not be estimated because the constraint, A_i had been reduced to a value of zero.

The hypothesis of model improvement is again, valid here only partially for the indicated *spatial structure* measures.

5.3.4. The Vehicle Data

Variables P_j and C_j improved the *goodness of fit* of the *basic model* significantly. In terms of *specification*, P_{ij*}, the pairs (P_{ij},P_{ij*}) and (P_j,P_{j*}) are significant measures. Variables P_{ij}, C_j and the composite (P_j,P_{j*}) also improve the *basic model's coefficient of variation* but they are not significant.

Table 14

Origin Specific Production Constrained Model
Calibration of the Vehicle Data

Model		C.V.	G.F.	Ma.B.	Mi.B.	N.B.	P.B.
Basic		0.48	0.76	-1.19	-5.86	25	0
With	Z_{ij}	0.51	0.85	-1.07	-6.01	25	0
	$Z_{ij}*$	0.50	0.88	-1.10	-5.77	25	0
	P_{ij}	0.40	0.75	-1.04	-4.84	25	0
	P_j	0.38	0.69^	-1.11	-4.91	25	0
	$P_{ij}*$	0.32^	1.06	-0.85	-4.43	25	0
	P_j*	0.41	1.06	0.87	-4.49	25	0
	C_j	0.36	0.74^	-1.02	-4.86	25	0
	$P_{ij}, P_{ij}*$	0.33^	1.16	-0.91	-4.91	25	0
	P_j, P_j*	0.31^	1.10	-0.97	-4.98	25	0
	$Z_{ij}, Z_{ij}*$	1.27	0.82	4.58	-5.06	23	2

As can be seen in the table above, the *spatial structure* measures which improve the *specification* of the *basic model* do not at the same time improve the *goodness of fit* of the *basic model*.

5.4. Origin Specific Doubly Constrained Model

In estimating the distant exponent of the interaction data of the various data sets, none of the *spatial structure* measures significantly improved the *specification* and/or *goodness of fit* of the *basic model*. In only the bus data, variables P_{ij} and Z_{ij} showed better model *specification* and *prediction* respectively. Variables P_j and P_{j*} also *specified* the basic model better.

In general, the *doubly constrained* spatial interaction model statement turned out to be the best *specified* without any *spatial structure* measure. Its *goodness of fit* is also the best across the various data sets. Fotheringham's work (1983) also confirms the same thing although in his case, only one method of including the *spatial structure* influence (P_{ij}) is used.

5.5. System Wide Calibration

Unlike the origin specific distance exponents, which are as many as the number of origins in the system, this one has a single, over-all, beta value representative of all of the nodes involved in the calibration process. It is, in a way, an averaging condition although the mean of the system wide and individually estimated betas have not been found to balance with each other.

The following table shows the details of the goodness of fit and the beta values of the different models which resulted from subjecting them to the same computer search process as the origin specific model calibrations. To illustrate the effect of the *spatial structure* measure on spatial interaction, the different data sets are calibrated using the *production constrained* model statements. As could be seen in the table, the *goodness of fit* of the telephone and vehicle data are improved by variables P_{ij}, P_j, and C_j. The bus data shows no improvement at all; and the rail data *goodness of fit* is improved by variables Z_{ij} and the composite (Z_{ij},Z_{ij*}). The *goodness of fit* of the four data sets could be compared since the SRMSE is a standardized measure. Thus, in general, the rail data has the worst and the telephone the best *goodness of fit*.

Except for some cases, the spatial structure variables that significantly improved the goodness of fit of the origin specific calibrations also to improved the goodness of fit of the system-wide calibrated models. The composite variable (Z_{ij},Z_{ij*}) of the telephone data, Z_{ij} and again the pair (Z_{ij},Z_{ij*}) of the railway data, and P_{ij} of the vehicle data show improvements in goodness of fit in the system-wide calibration. These variables have not significantly affected the goodness of fit of the origin specific calibrations of the respective data sets.

Since it is not possible to test the significance of model improvement in a system wide calibration following the same procedure as in the origin specific case, a multiple regression estimation has been performed. This approach allows one to observe whether or not adding a *spatial structure* variable adds appreciably to the *coefficient of determination* (R^2) or to the T statistics, and whether the models and the variables included in the calibration are significant or not.

Table 15

System Wide Production Constrained Model Calibrations

	Telephone		Bus		Rail		Vehicle	
Model	Beta	G.F.	Beta	G.F.	Beta	G.F.	Beta	G.F.
Basic	-0.39	0.70	-2.60	0.89	3.01	2.28	-1.68	1.58
With								
Z_{ij}	-0.21	0.41	0.00	0.95	1.50	2.27	-1.44	-1.64
$Z_{ij}*$	-2.90	0.70	-0.06	0.96	1.50	2.37	-1.55	1.60
P_{ij}	-0.59	0.51	-0.39	1.07	1.58	2.37	-1.56	1.43
P_j	-0.93	0.44	-0.62	0.98	1.50	2.35	-1.78	1.38
$P_{ij}*$	-0.52	1.79	-0.47	1.68	10.71	4.15	-2.02	1.72
P_j*	-0.62	1.79	-0.65	1.62	10.69	4.15	-2.13	1.71
C_j	-0.91	0.48	-0.63	0.96	2.72	2.42	-1.80	1.36
$P_{ij}P_{ij}*$	-0.70	1.74	-0.50	2.00	14.40	2.76	-2.10	1.72
P_j,P_j*	-1.08	1.72	-0.90	1.85	14.30	2.85	-2.40	1.66
$Z_{ij}Z_{ij}*$	-1.00	0.42	0.20	1.02	2.40	2.22	-1.30	1.67

As mentioned in Chapter 4, normally, the *production constrained* version is calibrated using *maximum likelihood* estimation techniques as the introduction of the *balancing* *factors* makes the relationship of the variables somewhat complex. However, if a model's *specification* is correct, (which this study assumes to be true when *spatial structure* variables are considered), the choice among the available calibration methods (*ordinary least squares* versus *maximum likelihood*) has been reported not to be that critical (Stetzer, 1976; Baxter and Ewing, 1979; Baxter 1982; Willenkens, 1982; Sen, 1985).

The use of the *ordinary least squares* multiple regression model in place of *maximum likelihood* may not be commendable in all circumstances, since the statistical assumptions required by the different estimation techniques may not always be met. It is, therefore, advisable to examine the nature of the distribution of the variables before attempting to use the *least squares* multiple regression approach for all gravity model versions. In this study, application of the *ordinary least squares* method saved time without losing very much precision of the regression estimates. The use of the *ordinary least squares* method is encouraged because a preliminary investigation of a *maximum likelihood* and a *least squares* estimates of a *production constrained* model resulted in similar distance parameters.

The variables have been square-root (SQR) transformed before calibrating the models. The choice of SQR over the commonly used logarithmic transformation is made because the former approach resulted in better adjusted R^2 values than the latter method and also the problem of zero interaction is not encountered.

As could be observed in Table 16, the effect of the *spatial structure* measures on the interaction of the different data sets is more or less consistent. However, there are some clear variations on the increment of the R^2 value between the data sets and also models.

For the telephone data, the different *spatial structure* variables hardly affected the *goodness of fit* value. Also, there is not appreciable difference between the *fit* of the *total flow* and *production constrained* models. This may mean that the usefulness, and in turn, problems of the different versions of the gravity-like models could possibly be data specific. In the other data sets, differences observed in the R^2 value because of calibrating the *total flow constrained* or the *production constrained* models are more pronounced than differences because of calibrating models with *spatial structure* variables. This can mean that the improvement in the R^2 values is a condition of the *production constraint* (A_i) rather than the effect of an additional *spatial structure* variable.

Table 16

Coefficient of Determination from System Wide Multiple Regression Calibration of the Total and Production Constrained Models

M O D E L		Telephone R^2	Bus R^2	Rail R^2	Vehicle R^2
Basic	Total Flow Consd.	0.89 (14.61)	0.63 (0.79)	0.39 (27.91)	0.40 (7.66)
	Production Consd.	0.93 (15.34)	0.81 (0.80)	0.62 (28.09)	0.59 (7.58)
Z_{ij}	Total Flow Consd.	0.90 (14.21)	0.55 (4.0)	0.40 (27.69)	0.45 (7.37)
	Production Consd.	0.92 (16.13)	0.66 (4.0)	0.63 (27.71)	0.61 (7.38)
Z_{ij}*	Total Flow Consd.	0.90 (14.13)	0.63 (0.79)	0.39 (27.93)	0.42 (7.55)
	Production Consd.	0.93 (15.41)	0.81 (0.80)	0.80 (28.05)	0.58 (7.65)
Z_i	Total Flow Consd.	0.90 (14.14)	0.63 (0.74)	0.40 (27.87)	0.43 (7.46)
	Production Consd.	0.93 (14.66)	0.81 (0.79)	0.63 (27.87)	0.60 (7.49)
Z_i*	Total Flow Consd.	0.90 (14.19)	0.63 (0.79)	0.39 (27.93)	0.41 (7.56)
	Production Consd.	0.93 (14.74)	0.81 (0.80)	0.62 (28.05)	0.57 (7.74)
P_{ij}	Total Flow Consd.	0.90 (13.75)	0.64 (0.78)	0.39 (27.82)	0.46 (7.32)
	Production Consd.	0.92 (15.76)	0.81 (0.80)	0.63 (27.82)	0.42 (8.60)
P_j	Total Flow Consd.	0.90 (13.97)	0.64 (0.77)	0.39 (27.91)	0.44 (7.43)
	Production Consd.	0.92 (16.31)	0.81 (0.80)	0.62 (27.920	0.59 (7.56)
P_{ij}*	Total Flow Consd.	0.91 (13.65)	0.63 (0.79)	0.45 (24.48)	0.42 (7.54)
	Production Consd.	0.93 (15.61)	0.81 (0.79)	------ ------	0.61 (7.42)

Table 16 (Con'd).

	Total Flow Consd.	0.91 (13.56)	0.63 (0.79)	0.39 (27.93)	0.42 (7.54)
P_{j*}					
	Production Consd.	0.93 (15.10)	0.81 (0.80)	------ ------	0.60 (7.42)
	Total Flow Consd.	0.90 (13.96)	0.65 (0.77)	0.39 (27.94)	0.44 (7.41)
C_j					
	Production Consd.	0.92 (16.28)	0.81 (0.80)	0.62 (28.01)	0.64 (7.09)
	Total Flow Consd.	0.91 (13.41)	0.63 (0.78)	0.39 (27.84)	0.44 (7.35)
P_{ij},P_{ij*}					
	Production Consd.	0.94 (13.74)	0.81 (0.80)	0.63 (27.83)	0.62 (7.26)
	Total Flow Consd.	0.91 (13.54)	0.63 (0.79)	0.39 (27.93)	0.44 (7.43)
P_j,P_{j*}					
	Production Consd.	0.93 (14.91)	0.81 (0.79)	0.63 (27.90)	0.61 (7.37)
	Total Flow Consd.	0.91 (13.41)	0.63 (0.78)	0.40 (27.68)	0.43 (7.40)
Z_{ij},Z_{ij*}					
	Production Consd.	0.94 (14.36)	0.81 (0.80)	0.63 (27.64)	0.61 (7.41)
	Total Flow Consd.	0.92 (12.23)	0.64 (0.78)	0.40 (27.59)	0.51 (6.85)
Z_{ij},P_{ij}					
	Production Consd.	0.94 (13.81)	0.81 (0.79)	0.63 (27.60)	0.65 (7.00)
	Total Flow Consd.	0.91 (13.09)	0.66 (0.76)	0.39 (27.87)	0.49 (7.06)
Z_i,P_j					
	Production Consd.	0.94 (14.27)	0.81 (0.78)	0.63 (27.87)	0.63 (7.15)
	Total Flow Consd.	0.91 (13.06)	0.63 (0.79)	0.39 (27.95)	0.45 (7.39)
Z_{i*},P_{j*}					
	Production Consd.	0.93 (14.64)	0.81 (0.80)	0.62 (27.99)	0.59 (7.52)
	Total Flow Consd.	0.92 (12.89)	0.63 (0.79)	0.39 (27.95)	0.44 (7.39)
Z_{ij*},P_{ij*}					
	Production Consd.	0.94 (13.88)	0.81 (0.80)	0.62 (28.00)	0.60 (7.48)

(numbers in parenthesis are standard errors of estimate)

According to the popular statistical literature of regression, a variable may not contribute to the *goodness of fit* measure if it is collinear with an already considered variable in the calibration process (Pindyck and Rubinfeld, 1980). To what extent this has been the case could be observed from a table of intercorrelations of the included variables. In this exercise, an intercorrelation of all the variables has been performed by data set (Appendix 2a-2d). Observation of these tables shows that there is a high relationship among the destination based *spatial structure* measures, specifically among P_{ij}, P_j and C_j. Variables $P_{ij}*$ and P_j* although destination based, have very low association with the other variables. As explained earlier, the pairs, P_{ij} and P_j and $P_{ij}*$ and P_j*, are strongly correlated, as are Z_i and Z_{ij}, and Z_i* and $Z_{ij}*$. This confirms that the effect of the matrix values of *spatial structure* on spatial interaction will not be very different from the impact of the vector values. Further, the correlation values indicate that the effect of *spatial structure* on individual points in a given network is similar even when some of the points are not included in computing the *spatial structure* values. However, the finding in this exercise shows that in almost all cases, the vector values improved the *specification* and *goodness of fit* of the *basic model* better than the matrix values. This implies that more complete *spatial structure* measures will have stronger impact on the performance of interaction models.

Other high correlations are also observed between the *production constraint* (A_i) and the *propulsiveness* measure (V_i); 0.93, 0.85, 0.82, and 0.88 for the telephone, bus, rail and vehicle related variables respectively. Fairly high relationships are also observed among some more variables such as between V_i, P_{ij} and Z_{ij}; and also between Z_{ij*}, P_{ij} and the constraint A_i. The relationship among the rest of the variables is in general low and inconsistent.

The SRMSE *goodness of fit* and the regression R^2 values are very different, probably for two major reasons. The first may be a problem of aggregating the calibration of the models. That is, the system wide calibration obviously generalizes and as a result may distort actual conditions. The second, and probably no less important is the R^2 value itself. It has been reported (Knudsen and Fotheringham, 1986) that the R^2 is an unsatisfactory goodness of fit measure since it is not sensitive to changes in model *specifications*. Therefore, it may not be reflecting the true effect of the *spatial structure* variables.

There has been a special encounter during the process of calibrating the *production constrained* model with the destination based distance-type *spatial structure* variables (P_{ij*} and P_{j*}) using the regression approach. The reason is that the constraint, A_i became infinitesimally small when generating it in the search process and as a result it turned out to be a constant of zero values. The beta estimated by the search process from the same *production constrained* model with variable, P_{ij*}, has particularly been large, over 100 (Table 13, p. 92).

5.6. Negative and Positive Betas

In the various origin specific calibrations, not all the betas that correctly *specify* a model or give the best *goodness of fit* are negative. Positive beta values were, initially, considered as something meaningless (Linnman, 1966). Later, Fotheringham (1984) and Fotheringham and Dignan (1984) added a new meaning to the direction of the distance exponents. The kind of relationship that exists between destinations has been stated as *competitive* if the betas are negative. If the calibrations of the models result in positive beta values, then *agglomerative* (i.e., concentration of interacting origins and destinations within short distances) forces are said to be in order. In this exercise, when a *spatial structure* variable is added to the model, not only the values of the betas change, but so also the direction of some of them. Although a clear pattern across the different models and data sets is not observable, the betas of some centers switch from negative to positive values or vice-versa when a model is calibrated with one or the other of the *spatial structure* variables and/or models.

However, there is a tendency of positive betas to associate themselves with big centers and the negative ones with smaller centers. There seems to be some kind of consistency between high proportion of negative betas and significance of models in either specification, predictive capacity, or both, especially, for the best models. The percent of centers that changed direction is given by model and data type in the following table:

The *production constrained* calibration of the model with variable P_{ij} for node 3 of the telephone data, for instance, results in a positive beta associated with it. The same model calibration with variable Z_{ij} gives a negative beta estimate for the same node.

Furthermore, the distance of the beta values from zero in either directions is different for different model

Table 17

Percent of Centers that Switched Direction
Associated with a Given Spatial Structure Variable

Model	Telephone	Bus	Rail	Vehicle
Total	92.13	13.33	40.91	52.00
Production	7.69	20.00	50.00	48.00
Double	7.69	13.33	36.36	36.00

calibrations with the various *spatial structure* measures. If the previous assumptions are right, an increasingly negative beta is a sign of the significance of distance impacting interaction. An increasingly positive beta, on the other hand, is indicative of the importance of agglomeration effect.

5.7. Consistency of the Effects of Spatial Structure Variables on Centers of a System

The correlation of the best beta estimates, from the different model calibrations with the various *spatial structure* variables, shows that the effect of the addition of *spatial structure* variable is more model specific than is data specific. Variations are observed between the *total flow constrained* and *production* constrained models across data sets. These variations seem to be consistent across the different model calibrations for all data sets. The correlation between the best betas of the *basic model* and the best betas derived from calibrating models with the various *spatial structure* variables are given in the following table.

Each value in Table 18 represents the degree of association between the best beta calibration of the *basic model* and the best beta calibration of other models with the various *spatial structure* measures. High correlation values are indicative of similar trend in the beta values of the

Table 18

Correlation between best betas of the basic model and other models with various spatial structure variables by Center of System

Model	Telephone			Bus			Rail			Vehicle		
	Tt	Pr	Db	Tt	Pr	Db	Tt	Pr	Db	Tt	Pr	Db
Z_{ij}	.28	.72	.76	.99	.98	.94	.99	.99	.99	1.0	1.0	.97
Z_{ij*}	1.0	1.0	1.0	.93	.87	.90	.99	1.0	1.0	.86	1.0	.99
P_{ij}	1.0	.91	.92	.88	.79	.99	.75	.97	.98	.90	.93	.92
P_j	.31	.92	.92	.90	.84	.92	.74	.97	.95	.92	.93	.90
P_{ij*}	.33	.67	.10	-.56	.47	.94	-.35	.79	.71	-.37	.38	.61
P_{j*}	.31	.69	.11	-.63	.46	.93	.60	.65	.29	-.67	.31	.62
C_j	1.0	.89		.89	.81		.77	.99		.91	.90	
$P_{ij}P_{ij*}$	.40	.82		-.66	.39		-.16	.76		-.58	.29	
P_jP_{j*}	.50	.42		-.12	.30		-.29	.97		-.58	.24	
$Z_{ij}Z_{ij*}$	-.23	.73		.98	.98		.98	1.0		.74	.14	

Tt = total flow constrained
Pr = production constrained
Db = doubly constrained

various model calibrations for each node in a system of interacting nodes, and low correlation values are indicative of irregular trends in beta values. In general, there is a consistently similar trend in correlation values across nodes of the various data sets and the different model calibrations. Among the exceptions is, for example, the association between the beta values of the *total flow* *constrained* model calibrated with variable Z_{ij} and the beta values of the *basic total flow constrained* model of the telephone data. The correlation value between the two is only 0.28. The correlations of the betas in the other data sets are high although the calibrated models are the same *total flow constrained* ones and the included *spatial structure* measure is also the same Z_{ij}. Similar conditions are observed for variables P_j, P_{ij*}, P_{j*} and the pairs (Z_{ij}, Z_{ij*}). Most of the betas derived from the different models are, however, highly intercorrelated. Only 19.73% of the associations have values below 0.50, and only 11.54% are negatively correlated.

CHAPTER 6.

SUMMARY, CONCLUSIONS AND IMPLICATIONS

6.1. Summary

The empirical analysis in Chapter 5 of this study has shown that the effect of a *spatial structure* measure may be generalized at different levels. It has been determined that in general:

(a) There are different ways of representing *spatial structure* which improve the *specification* of the *basic model* or give a better *goodness of fit, or both. This has been observed across the different data sets and model calibrations in this study.*

(b) The effects of the different *spatial structure* measures have not been the same in all the cases under investigation. This statement relates to the magnitude of model improvement due to the addition of a given *spatial structure* measure, and also the consistency of impact of a *spatial structure* variable on interaction across the data sets and models.

(c) When ranked in terms of frequency of improving a model, both in terms of model *specification* and *goodness of fit*, the best *spatial structure* measure of significant effect has been P_j*, the vector of the destination based distance-type *spatial structure* measure.

(d) Observation of the pattern of the betas shows that both positive and negative distance exponent estimates are possible in interaction model calibrations even without *spatial structure* variables. With additional variables or constraints, the proportion of the negative betas tend to increase although the switch in direction of the exponents does not show any appreciable pattern.

As a direct consequence of the above conditions, only some of the stated hypotheses have been accepted (Tables 19 and 20). Thus, the major hypothesis which proposed significant improvement in model *specification* and *goodness of fit* has been found true for only some of the cases. Some of the spatial structure measures improved the *specification* and *goodness of fit* of the models; some of them improved only one of the two; a good number of them showed better situations even though the improvements were not significant at 0.05% level; and others gave worse results than the *basic model*.

Sub-hypothesis 1 (p. 54) proposed significantly improved *specification* and *goodness of fit* of models when

Table 19

Significant Results of the Origin Specific
Total Flow Constrained Model Calibrations

	Telephone		Bus		Rail		Vehicle	
Model	C.V.	G.F.	C.V.	G.F.	C.V.	G.F.	C.V.	G.F.
Basic								
With Z_{ij}								
Z_{ij*}								
P_{ij}			+				+	+
P_j			+				+	+
P_{ij*}		+	+	+		+	+	+
P_{j*}		+	+	+		+	+	+
C_j			+				+	+
P_{ij}, P_{ij*}			+			+	+	
P_j, P_{j*}			+			+	+	
Z_{ij}, Z_{ij*}								

destination based accessibility measures are included in the calibration process. This hypothesis has not been accepted across data sets and model calibrations. Of the total flow calibrations of the different data sets, variables P_{ij} and/or P_j have significant impact on only the vehicle data

Table 20

Significant Results of the Origin Specific Production Constrained Model Calibrations

	Telephone		Bus		Rail		Vehicle	
Model	C.V.	G.F.	C.V.	G.F.	C.V.	G.F.	C.V.	G.F.
Basic								
With Z_{ij}	+	+						
Z_{ij*}								
P_{ij}	+	+						
P_j	+	+						+
P_{ij*}					+		+	
P_{j*}			+		+			
C_j	+	+						+
P_{ij}, P_{ij*}					+		+	
P_j, P_{j*}			+		+		+	
Z_{ij}, Z_{ij*}								

in *specification* as well as *fit* of model. The impact of these variables in the bus data has been only in improving the *specification* of the model.

The *production constrained* calibration of the four data sets with the same *spatial structure* measures shows a different picture. They have impacted the *specification* and *fit* of all of them except the railway data. However, they have not been found as best model *specifiers* in all of the data sets. except the models with variable P_j which resulted in the best *fit* for the bus and vehicle data.

Sub-hypothesis 2 (p. 56) which proposed a possible model improvement with *spatial structure* measure of only the distance variable has been partly accepted. The destination based *spatial structure* measure, P_j* has not only improved the models which it is a part, but it turned out to be the best model *specifier* and predictor for the *total flow constrained* model calibration of the telephone, bus and vehicle data. It is also an important addition for the *production constrained* calibration of the data sets except for the telephone one.

The origin based *spatial structure* measure of the distance variable (hypothesis 4, p. 59), on the other hand, showed only some insignificant impact on the *total flow constrained* calibration of the telephone and the *production constrained* calibration of the railway data.

The third sub-hypothesis (p. 58) proposed model prediction and *specification* improvements with origin based accessibility measures. This hypothesis is accepted for only the *production constrained* calibration of the telephone data and that for only giving a significantly improved *goodness of fit*. The *coefficient of variation* associated with the same model shows only a slight improvement over the basic model.

Sub-hypothesis 5 and 6 (pp. 60 & 61) which proposed better model *specification* and *goodness of fit* with *multi-directional* measures also falls short of full acceptance. Only in the *production constrained* calibration of the vehicle data has the pair variables (P_j,P_j*) resulted in the best *specified* model although the *fit* of the model is only second from the last. The destination based *bi-directional spatial structure* measures also showed significant impact in *specifying* the *total flow constrained* model for the bus data, and in giving improved *goodness of fit* for the railway data.

Sub-hypothesis 7 (p. 63) proposed that including the interacting origins and destinations while deriving the *spatial structure* measures brings a difference in model improvement. Although not significantly, variables P_j, P_j* Z_i and Z_i* gave better results than variables P_{ij}, $P_{ij}*$, Z_{ij}, and $Z_{ij}*$ in almost all cases. Significant differences between the two sets is not expected since they approximate each other. But, there have been instances where the vector values in some data sets are significant while the matrix values are not.

Sub-hypothesis 8 (p. 65) proposed improved model *specification* and/or *goodness of fit* because of including the *production* and *attraction constrained* models *balancing factors* as *spatial structure* measures. Especially, B_j (measured by C_j) has been found to improve the *specification* and *performance* of the *basic model* significantly.

The last of the sub-hypotheses (p. 66) proposed the same *goodness of fit* from models with origin or destination based *spatial structure* measures. This hypothesis has not been accepted at all. The significant *goodness of fit* measures of models with origin and destination based accessibilities are all different. Further, including destination based *spatial structure* measures seem to give better *goodness of fit* than the origin based *spatial structure* variables. Even variable C_j which at times impacted the models usefulness significantly, is a destination based measure. But, when it comes to the system-wide calibrations (the *goodness of fit* measure here being R^2), there does not seem to be any notable difference between one or the other of the *spatial structure* measures in influencing the model's *fit*. Even those variables that significantly improved the goodness of fit of the origin specific calibrations, as measured by SRMSE, hardly show any improvement in the R^2 in many of the cases (Table 21). The maximum improvement because of introducing a spatial structure variables (P_{ij} and P_{ij*} of the vehicle and railway data respectively) in the total flow constrained model calibration is only 0.06.

Table 21

Adjusted R^2 Values of the Models with the Spatial Structure Variables that Significantly Improved the Goodness of Fit of the Basic Model as Measured by SRMSE

Model	Telephone R^2	Bus R^2	Rail R^2	Vehicle R^2
Total Flow Cons.				
Basic	**0.89**	**0.63**	**0.39**	**0.40**
P_{ij}	--	--	--	0.46
P_j	--	--	--	0.44
$P_{ij}*$	0.91	0.63	0.45	0.42
P_j*	0.91	0.63	0.39	0.39
C_j	--	--	--	0.44
$P_{ij},P_{ij}*$	--	--	0.39	--
P_j,P_j*	--	--	0.39	--
Production Cons.				
Basic	**0.93**	**0.81**	**0.62**	**0.59**
Z_{ij}	0.92	--	--	--
P_{ij}	0.92	--	--	--
P_j	0.92	--	--	0.59
C_j	0.92	--	--	0.64

In the case of the production constrained calibration, the level of improvement is even lower than this. Except for the vehicle data, where the goodness of fit of the *basic model* is improved by 0.05 because of including variable C_j, the rest have virtually the same R^2 value.

6.2. Conclusions and Implications

This study has shown that different methods of including the *spatial structure* effect have different impacts on interaction modeling. But still generalizations are possible. Probably, a major concluding generalization is that the *specification* and *prediction capability* of spatial interaction models are likely to improve if some representative variable of *spatial structure* is included during calibration. But there are difficulties associated with identifying the correct variable, especially, regarding the theoretical justification that could be associated with the selection of one type of *spatial structure* over another.

The fact that there is not any conclusive support to a single *spatial structure* measure that could be applicable over all data sets and model calibrations may be an indication of one or the other, or a combination of the following:

(a) The *spatial structure* measure for different data sets are possibly different.

(b) The constraint of a model may possibly be sufficient to take care of the *spatial structure* effect; if another variable of similar definition is included, it may distort the impact of the new variable on interaction.

(c) The nature of the *spatial structure* variable needed may be different for different model calibrations, that is, *total flow, production or attraction constrained* models and/or data type.

The measurements of *spatial structure*, as exercised here and in previous studies, are not without problems. Both the population potential-type and distance-type *spatial structure* measures are not independent of the variables of the *basic* gravity model. The confounding effect is a result of both the distance and mass variables. *Spatial structure* may be an important element if it does not interfere with the already included distance and mass terms in the gravity-type spatial interaction models.

In calibrating the gravity model, with the various *spatial structure* variables, the role of the mass terms is ignored since they are forced not to vary. The adjustment for model specification and performance is thus only a partial one. Therefore, the fact that the distribution over space of the distance exponents is approaching a constant value may not guarantee that the model is also approaching full specification.

There is also a possible problem associated with coefficient of variation (C.V.) as a measure of *spatial structure.* If, after adding a *spatial structure* measure, the *mean* of the betas becomes zero, the C.V. of their distribution will be undefined. Further, if, after the inclusion of a spatial structure measure, all the beta values distributed over space become zero, it would mean that spatial structure has totally replaced the spatial separation effect. Will this be an advantage or could it be considered as an improvement over the basic model's specification?

The question of constraints is an intriguing one. Why the *basic production constrained* model performs better than the *basic total flow constrained* model, that is, without adding a *spatial structure* variable to the model statements, and like-wise why the *doubly constrained* model results in the best *specification* and *goodness of fit* of models give important clues as to whether a *spatial structure* variable is a necessary addition to interaction modeling. It is tempting to think this way because the constraints of the *doubly constrained* and *attraction constrained* models can also serve as *spatial structure* variables, be it origin or destination based accessibility. As mentioned in Chapter 4, Wilson (1967) in fact, writes that these constraints are used as *accessibility* or *competition* measures as may be required. In Fotheringham's study, and in this one, the addition of any *spatial structure* measure to the *doubly constrained* model brought insignificant or no improvement at all. In most cases, in fact, it has reduced the model's usefulness. The constraint of the attraction constrained model (B_j) has also improved the specification and goodness of fit of a good number of the origin specific models in this study. As shown in Table 21 too, the use of spatial structure variables has not been found encouraging since the R^2 values (assuming that it measures goodness of fit correctly) of the new models are not improved appreciably to justify the effort.

However, this may not mean that *spatial structure* does not influence flow at all. The question is rather, whether it is necessary to create a new variable or interpret existing constraints as *spatial structure* measures. It has been shown in Chapter 2 that after all, the origin based and destination based *spatial structure* measures are transpose values of each other. Variable C_j, which measures attraction constraint (B_j), is different because the *propulsiveness* values are used to derive it. It could not be used as the B_j of the *doubly constrained* model because, here, the balancing factors are iteratively determined. Nonetheless, including variable C_j in both the *total flow* *constrained* and *production constrained* calibration has resulted in improved model *performance*.

Although, in general, the addition of a particular type of *spatial structure* variable results in improved model *specification* and/or *goodness of fit* (SRMSE), the magnitude of contribution is dependent on the type of spatial flow and the model used to estimate the parameters. Since at times some of the *spatial structure* variables reduced the *specification* and/or *goodness of fit* of the *basic model,* the response of the specific model to the configuration of the locations under study need be examined closely. The improvement in model performance using the linear *spatial structure* measure of the railway data (Appendix 1c) has not been particularly encouraging. Thus, further research is still necessary to determine theoretically the variable effect of *spatial structure* measures on different flow systems.

The *coefficient of variation* and *goodness of fit* measures used in this study are standard measures. Therefore, they could be used across different model calibrations of different or the same data sets. Observation of the various tables of beta and *goodness of fit* values shows that the *production constrained* model calibrations are superior to the *total flow constrained* model calibrations in *goodness of fit* measures. The *coefficient of variation* however may or may not be better.

Why different points in the same network of origins and destinations have different beta values associated with the best *goodness of fit* and vic-versa even after including the *spatial structure* variable is something that needs further investigation. In other words, why does one model associated with a given node have a better *specification* and/or *fit* than another node of the same system, given that they are treated identically? If according to the assumption, the accessibility indexing role of the distance variable is eliminated by the introduction of the *spatial structure* variable, should the *coefficient of variation* and SRMSE not approach zero? If these measures are not at fault, it means that spatial interaction models are still far from becoming satisfactory representatives of reality.

A potential source of error in interaction modelling is the quality and sufficiency of the data used to calibrate the models. The data used in this kind of study represents only a sub-set of the universe of interactions. As long as we are not treating a closed system, some of the points in a network will be left out. The order of an interaction matrix is, therefore, an important consideration since the *spatial structure* measures are dependent, as are the interaction values, on the number of the points considered in the sub-system. To this effect, some evidence has been provided by Ewing (1986). While commenting on Fotheingham's airline data, Ewing reports that the distance cut-off point (160 miles) considered by Fotheringham to avoid modal competition gives a wrong impression of his findings. Ewing argues that when the cut-off point is increased to 1000 miles (which apparently decreases the number of points interacting by air), he gets a different *spatial structure* effect. According to Ewing, in fact, the improvement in model specification advocated by Fotheringham is "...the result of contamination of the data by a modal share effect. And the apparent success of his model in reducing the spatial variation is ...coincidental" (p. 547).

The implication of the various questions raised and the remarks made above is that the concept of *spatial structure* needs further refinement. What this study has shown is that the 'independent' variables of the gravity-type spatial interaction models could neither be easily measured nor defined. Thus, it may be appropriate to consider various other ways of improving the accuracy of an interaction model before settling on only including a *spatial structure* measure of one kind or the other. Plausible theoretical backing is also necessary for using a particular set of measures as correct model *specifiers, best predictors or both*. Further research is in order to be able to understand the role of the spatial attributes that influence the generation and attraction of trips. A *spatial structure* measure may not be the only missing variable.

APPENDICES

Appendix 1

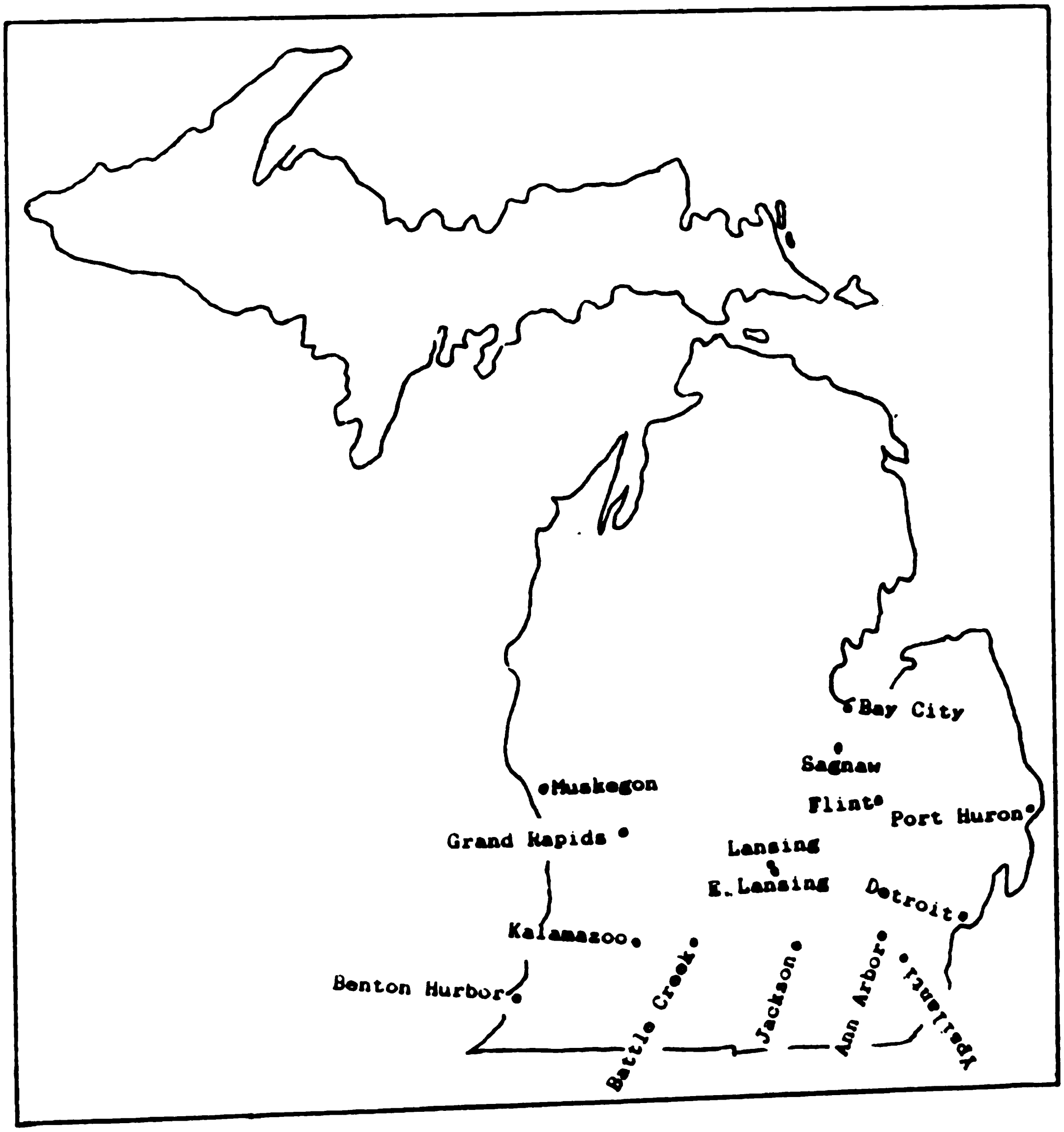
Interaction Nodes

Appendix 1 a

Michigan: Cities of Bus Passengers Flow Data

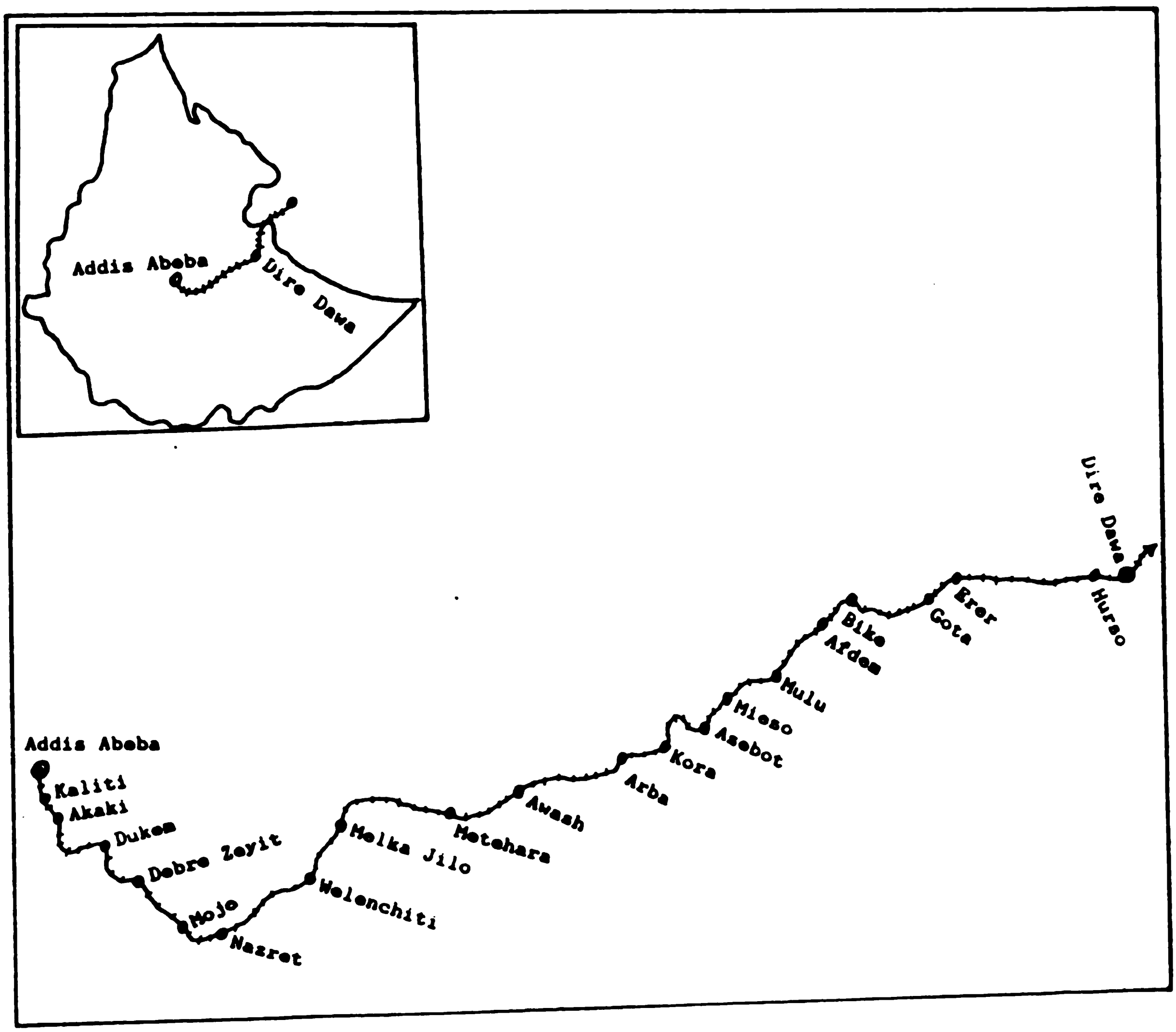
Appendix 1 b

Michigan: Cities of Vehicle Flow Data



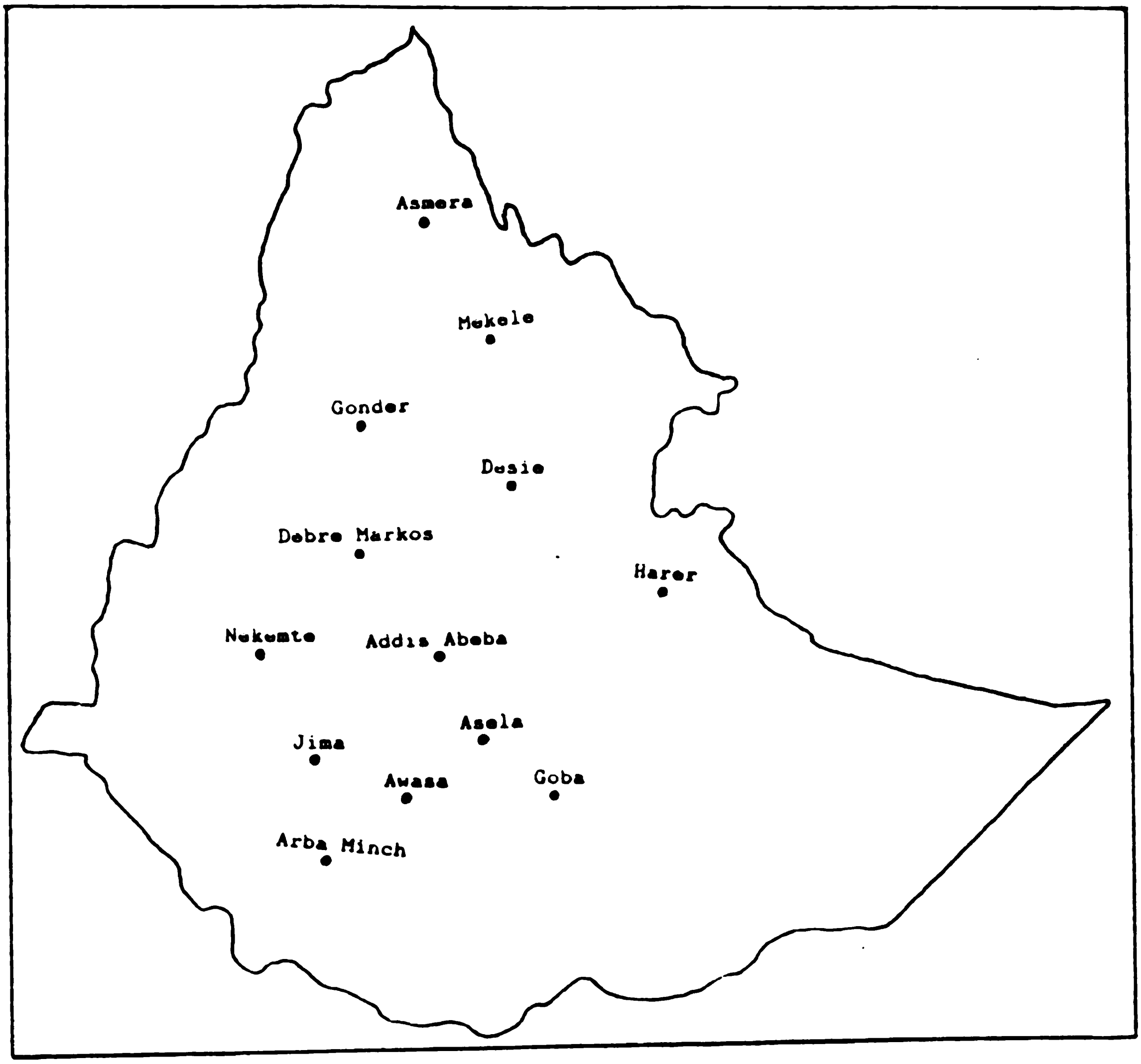
Appendix 1 c

Ethiopia: Cities of Rail Passengers Flow Data



Appendix 1 d

Ethiopia: Cities of Telephone Flow Data



Appendix 2

Correlations

Appendix 2a

Simple Correlation of the Variables of the Telephone Data

	T_{ij}	d_{ij}	V_i	W_j	C_j
T_{ij}	1.000				
d_{ij}	-0.292	1.000			
V_i	0.587	-0.171	1.000		
W_j	0.638	-0.159	-0.083	1.000	
C_j	0.171	0.368	-0.020	0.269	1.000
P_j	0.215	0.394	-0.026	0.337	0.993
P_j*	-0.430	0.153	0.054	-0.674	0.018
Z_i	0.181	0.349	0.308	-0.028	-0.083
Z_i*	-0.383	0.153	-0.652	0.056	-0.002
P_{ij}	0.599	0.055	0.637	0.225	0.725
$P_{ij}*$	-0.406	0.071	0.076	-0.670	0.018
Z_{ij}	0.602	0.055	0.203	0.631	0.052
$Z_{ij}*$	-0.357	0.071	-0.649	0.077	-0.031

Appendix 2a (Con'd).

	Pj	Pj*	Zi	Zi*
Pj	1.000			
Pj*	-0.046	1.000		
Zi	-0.083	0.004	1.000	
Zi*	0.004	-0.083	-0.046	1.000
Pij	0.729	-0.020	0.095	-0.404
Pij*	-0.047	0.994	-0.022	-0.109
Zij	0.095	-0.404	0.729	-0.020
Zij*	-0.022	-0.109	-0.047	0.994

	Pij	Pij*	Zij	Zij*
Pij	1.000			
Pij*	0.003	1.000		
Zij	0.199	-0.414	1.000	
Zij*	-0.414	-0.126	0.003	1.000

Appendix 2b.

Simple Correlation of the Variables of the Bus Data

	T_{ij}	d_{ij}	V_i	W_j	C_j
T_{ij}	1.000				
d_{ij}	-0.302	1.000			
V_i	0.570	-0.118	1.000		
W_j	0.419	-0.187	-0.069	1.000	
C_j	-0.041	0.366	-0.005	-0.098	1.000
P_j	-0.084	0.383	0.003	-0.199	0.990
P_j*	-0.180	0.242	0.024	-0.430	0.231
Z_i	-0.027	0.383	-0.047	0.014	-0.071
Z_i*	-0.189	0.242	-0.331	0.031	-0.017
P_{ij}	0.006	0.246	0.140	-0.192	0.953
$P_{ij}*$	-0.178	0.223	0.028	-0.431	0.231
Z_{ij}	0.054	0.236	-0.045	0.164	-0.124
$*Z_{ij}*$	-0.187	0.223	-0.331	0.036	-0.020

Appendix 2b (con'd).

	P_j	P_{j*}	Z_i	Z_{i*}
P_j	1.000			
P_{j*}	0.303	1.000		
Z_i	-0.071	-0.022	1.000	
Z_{i*}	-0.022	-0.071	0.303	1.000
P_{ij}	0.963	0.299	-0.138	-0.094
P_{ij*}	0.303	0.999	-0.025	-0.074
Z_{ij}	-0.138	-0.094	0.963	0.299
Z_{ij*}	-0.025	-0.074	0.303	0.999

	P_{ij}	P_{ij*}	Z_{ij}	Z_{ij*}
P_{ij}	1.000			
P_{ij*}	0.303	1.000		
Z_{ij}	-0.168	-0.096	1.000	
Z_{ij*}	-0.096	-0.076	0.303	1.000

Appendix 2c.

Simple Correlation of the Variables of the Railway Data

	T_{ij}	d_{ij}	V_i	W_j	C_j
T_{ij}	1.000				
d_{ij}	0.115	1.000			
V_i	0.260	0.126	1.000		
W_j	0.280	0.119	-0.046	1.000	
C_j	0.082	-0.137	-0.016	0.293	1.000
P_j	0.133	-0.065	-0.026	0.474	0.908
P_j*	-0.071	0.052	0.010	-0.253	-0.408
Z_i	0.139	-0.065	0.537	-0.023	-0.043
Z_i*	-0.054	0.052	-0.210	0.012	0.019
P_{ij}	0.155	-0.170	0.145	0.455	0.835
$P_{ij}*$	-0.069	0.021	0.010	-0.250	-0.405
Z_{ij}	0.165	-0.170	0.517	0.147	0.022
$Z_{ij}*$	-0.054	0.021	-0.207	0.010	0.012

Appendix 2c (Con´d).

	P_j	P_j*	Z_i	Z_i*
P_j	1.000			
P_j*	-0.326	1.000		
Z_i	-0.048	0.016	1.000	
Z_i*	0.016	-0.048	-0.326	1.000
P_{ij}	0.929	-0.300	0.056	-0.035
$P_{ij}*$	-0.323	0.991	0.015	-0.044
Z_{ij}	0.056	-0.035	0.929	-0.300
$Z_{ij}*$	0.015	-0.044	-0.323	0.991

	P_{ij}	$P_{ij}*$	Z_{ij}	$Z_{ij}*$
P_{ij}	1.000			
$P_{ij}*$	-0.290	1.000		
Z_{ij}	0.221	-0.033	1.000	
$Z_{ij}*$	-0.033	-0.038	-0.290	1.000

Appendix 2d.

Simple Correlation of the Variables of the Vehicle Data

	T_{ij}	d_{ij}	V_i	W_j	C_j
T_{ij}	1.000				
d_{ij}	-0.308	1.000			
V_i	0.268	-0.283	1.000		
W_j	0.266	-0.259	-0.037	1.000	
C_j	-0.152	0.519	0.024	-0.571	1.000
P_j	-0.154	0.518	0.024	-0.579	0.999
P_j*	-0.087	0.420	0.013	-0.327	0.724
Z_i	-0.156	0.518	-0.584	0.024	-0.042
Z_i*	-0.086	0.420	-0.322	0.014	-0.030
P_{ij}	-0.126	0.508	0.070	-0.578	0.997
$P_{ij}*$	-0.086	0.413	0.010	-0.327	0.724
Z_{ij}	-0.129	0.508	-0.582	0.073	-0.068
$Z_{ij}*$	-0.086	0.413	-0.322	0.010	-0.025

Appendix 2d (Cont´d)

	P_j	P_{j*}	Z_i	Z_{i*}
P_j	1.000			
P_{j*}	0.723	1.000		
Z_i	-0.042	-0.030	1.000	
Z_{i*}	-0.030	-0.042	0.723	1.000
P_{ij}	0.998	0.721	-0.069	-0.045
P_{ij*}	0.723	1.000	-0.025	-0.041
Z_{ij}	-0.069	-0.045	0.998	0.721
Z_{ij*}	-0.025	-0.041	0.723	1.000

	P_{ij}	P_{ij*}	Z_{ij}	Z_{ij*}
P_{ij}	1.000			
P_{ij*}	0.721	1.000		
Z_{ij}	-0.094	-0.040	1.000	
Z_{ij*}	-0.040	-0.040	0.721	1.000

LIST OF REFERENCES

LIST OF REFERENCES

Abler, Ronald; Adams, John S and Gould, Peter. 1971. Spatial Organization: The Geographers View of the World. Prentice-Hall, New Jersey.

Alonso, W. 1978. A Theory of Movement. In N M Hansen (ed), Human Settlement Systems, Cambridge, Massachusetts, pp. 197-211.

Anderson, Theodore R. 1955. Intermetropolitan Migration: A Comparison of the Hypotheses of Zipf and Stouffer. American Sociological Review, 20(3):287-291.

Anderson, Theodore R. 1956. Potential Models and Spatial Distribution of Populations. Papers and Proceedings of the Regional Science Association, vol 2.

Aufhauser, E and Fischer, M M. 1985. Log-Linear Modelling and Spatial Analysis. Environment and Planning A, 17(7):931-951.

Batty, A and Sikdar, P K. 1984. Proximity Aggregation-estimation of Spatial Interaction Models. Environment and Planning A, 16(4): 467-486.

Baxter, M J and Ewing, G O. 1979. Calibration of Production Constrained Trip Distribution Models and the Effect of Intervening Opportunities. Journal of Regional Science, 19(3):319-330.

Baxter, M J. 1985. Mis-specification in Spatial Interaction Models. Environment and Planning A, 17(5):673-678.

Baxter, M J. 1985. Mis-specification in Spatial Interaction Models: Further Results. Environment and Planning, 17:673-678.

Baxter, Michael J. 1979. The Application of Logit Regression Analysis to Production Constrained Gravity Models. Journal of Regional Science. 19(2):171-177.

Baxter, Richard. 1976. Computer and Statistical Techniques for Planners. London, Methuen.

Bennett, R S. and Others. 1985. Spatial Structure, Spatial Interaction, and their Integration: A Review of Alternative Models. Environment and Planning, 17:625-645.

Black, William B. 1972. Interregional Commodity Flows: Some Experiments with the Gravity Models. Journal of Regional Science, 12(1):107-118.

Bright, Margaret L and Thomas, Dorothy S. 1941. Interstate Migration and Intervening Opportunities. American Sociological Review, 6:773-783.

Bucklin, Louis P. 1971. Retail Gravity Models and Consumer Choice: A Theoretical and Empirical Critique. Economic Geography, 47(4):487-497.

Carroll, j Douglas. 1955. Spatial Interaction and the Urban Metropolitan Description. Traffic Quarterly, 9(2):149-61.

Carrothers, Gerald A p. 1956. An Historical Review of the Gravity and Potential Concept of Human Interaction. Journal of the American Institute of Planners, 22:94-102.

Cesario, F J. 1973. Parameter Estimation in Spatial Interaction Modeling. Environment and Planning, 5:503-518.

Cesario, Frank J. 1975. Linear and Non-linear Regression Models of Spatial Interaction. Economic Geography, 51(1):69-77.

Chisholm, M and O'Sullivan, P. 1973. Freight Flows and Spatial Aspects of the British Economy. Cambridge, Cambridge University Press.

Clark, Colin and Peters, G H. 1962. The Intervening Opportunities Opportunities Method of Traffic Analysis. Traffic Quarterly 19: 104-115.

Clarke, M. 1985. The Role of Attractiveness Functions in the Determination of Equilibrium Solutions to Production Constrained Spatial Interaction Models. Environment and Planning, 17:175-183.

Cliff, A D and Martin, R C. 1974. Evaluating the Friction of Distance Parameter in Gravity Models. Regional Studies, 8:281-268.

Cliff, A D and Others. 1975. Map Pattern and Friction of Distance Parameters: Reply to Comments by R.J. Johnston and others. Regional Studies, 9:285-288.

Coffey, William J. 1981. Geography: Toward a General Spatial Systems Approach. Methuen, London.

Colwell, Peter F. 1982. Central Place Theory and the Simple Economic Foundation of the Gravity Model. Journal of Regional Science, 22(4):541-546.

Curry, L. 1972. A spatial Analysis of Gravity Flows. Regional Studies, 6:131-147.

Dodd, Stuart C. 1955. The Interactance Hypothesis: A Gravity Model Fitting of Physical Masses and Human Groups. American Sociological Review,15(2):245-256.

Edens, H J. 1970. Analysis of A Modified Gravity Model. Transportation Research, 4:51-62.

Ewing, G. 1986. Spatial Pattern in Distance Deterrence Parameters and Fotheringham's Theory of Competing Destinations. Environment and Planning A, 18(4):547-551.

Ewing, G. O. 1974. Gravity and Linear Regression Models of Spatial Interaction. Economic Geography, 50(1):83-88.

Fisk, C and Brown, G R. 1975. The Role of Model Parameters in Trip Distribution Models. Transportation Research, 9:143-148.

Fotheringham, A Stewart and Webber. M J. 1980. Spatial Structure and the Parameters of Spatial Interaction. Geographical Analysis, 12(1): 33-46.

Fotheringham, A S. 1984. Spatial Flows and Spatial Patterns. Environment and Planning A, 16:529-543.

Fotheringham, A S. 1985. Spatial Competition and Agglomeration in Urban Modeling. Environment and Planning A, 17:213-230.

Fotheringham, A S. 1986. Further Discussion on Distance Deterrence Parameters and the Competing Destinations Model. Environment and Planning, 18:553-556.

Fotheringham, A S. 1986. Modelling Hierarchical Destination Choice. Environment and Planning A, 18(3):401-418.

Fotheringham, A Stewart and Degnan, Tony. 1984. Further Contributions to the General Theory of Movement. Annals of the Association of American Geographers, 74(4):620-633.

Fotheringham, A Stewart. 1981. Spatial Structure and Distance Decay Parameters. Annals of the Associations of American Geographers, 71(3):425-436.

Galle, Omer R. and Taeuber, Karl E. 1966. Metropolitan Migration and Intervening Opportunities. American Sociological Review, 13(1).

Getis, A. 1984. Interaction Modeling Using Second-order Analysis. Environment and Planning A, 16(2):173-183.

Glejser, H and Dramais, A. 1969. A Gravity Model of Interdependent Equations to Estimate Flow Creation and Diversion. Journal of Regional Science, 9(3):439-449.

Gordon, I R. 1985. Economic Explanations of Spatial Variation in Distance Deterrence. Environment and Planning A, 17(1):59-72.

Gould, Peter. 1975. Acquiring Spatial Information. Economic Geography, 51(2):87-99.

Gregory, S. 1963. Statistical Methods and the Geographer. Longmans, London.

Griesinger, Donald W. 1979. Reconsidering the Theory of Social Gravity. Journal of Regional Science, 19(3):291-330.

Griffith, D A. 1976. Spatial Structure and Spatial Interaction: A Review. Environment and Planning A, 8:731-740.

Griffith, Daniel A and Lea, Anthony C (eds). 1983. Evolving Geographical Structure. Martinus Nijhoff, The Hague.

Haggett, Richard. 1980. Systems Analysis in Geography. Clareden Press, Oxford.

Halder, Alan. 1970. An Alternative Approach to Trip Distribution. Transportation Research, 4:63-69.

Halperin, W C and Others. 1984. A generalized Procedure for Comparing Models of Spatial Choice. Environment and Planning A, 16(10):1289-1301.

Hansen, Walter G. 1959. How Accessibility Shapes Land Use. Journal of American Institute of Planners, 25:73-76.

Haynes, Kingsley E and Fotheringham, A Stewart. 1984. Gravity and Spatial Interaction Models. Scientific Geography Series, Vol. 2, Sage Publication, London.

Haynes, Kingsley E and Others. 1973. Intermetropolitan Migration in High and Low Opportunity Areas: Indirect Test of the Distance and Intervening Opportunities Hypotheses. Economic Geography, 49(1):68-73.

Heanue, K E and Pyers, C E. 1965. A Comparative Evaluation of Trip Distribution Procedures. Highway Research Record, NO.114, HRB, Washington D C.

Hendriks, Paul. 1985. Measurement of the attractiveness of Places of Study for Geography Students: A comparison of Compensatory and non-compensatory Combination Rules. Tydschrift, 76:1.

Hoover, Edgar M and Giarrantani. 1984.An introduction to Regional Economics. Alfred A Knopf, New York.

Hua, Chang I and Porell, Frank. 1979. A Critical Review of the Development of the Gravity Model. International Regional Science Review, 4(2):97-126.

Hurst, Michael E Eliot. 1974. Transportation Geography: Comments and Readings. McGraw-Hill, New York.

Hutchinson, B G. 1974. Principles of Urban Transport Systems Planning. McGraw-Hill, New York.

Hyman, G M. 1969. The Calibration of Trip Distribution Models. Environment and Planning, 1:105-112.

Isard, Walter. 1975. Introduction to Regional Science. Prentice-Hall, New Jersey.

Isbell, Eleanor C. 1944. Internal Migration in Sweden and Intervening Opportunities. American Sociological Review, 9(6):627-639.

Johnston, R J. 1975. Map Pattern and Friction of Distance Parameters: A Comment. Regional Studies, 9:281-283.

Johnston, R J. 1976. On Regression Coefficients in Comparative Studies of the Frictions of Distance. Tijdschrift Voor Economic en Soc Geografie, 67 Nr 1.

Kanafani, Adeb. 1983. Transportation Demand Analysis. McGraw-Hill, New York.

Kennedy, Peter. 1985. A Guide to Econometrics. MIT Press, Cambridge.

Kirby, Howard R. 1970. Normalizing Factors of the Gravity Model: An Interpretation. Transportation Research, 4:37-50.

Kitamura, Kyuichi and Kostynuik, Lind P. 1979. Aggregation in Spatial Choice Modeling. Transportation Science, 13(4):325-342.

Knippernberg, Hans and De Vos, Sjoerd. 1982. On Measurement and Explanation of Spatial Structure Effects. Tijdschrift Voor Econ. en Soc. Geografie 74, Nr. 2 pp. 120-128.

Krueckeberg, Donald A and Silvers, Arthur L. 1974. Urban Planning Analysis: Methods and Models. John Wiley, New York.

Kua, James B and Sirmans, C F. 1979. The Functional Form of the Gravity Model. International Regional Science, 4(2):127-136.

Ledent, J. 1985. The Doubly Constrained Model of Spatial Interaction: A More General Formulation. Environment and Planning A, 17(2):253-262.

Lombardo, Stand Rabino, G A. 1986.Calibration Procedures and Problems of Stability in Non-Linear Dynamic Spatial Interaction Modeling. Environment and Planning A, 18(3):341-350.

Lowe, John C and Moryadas, S. 1975. The Geography of Movement. Waveland Press, Illinois.

Long, Wesley H and Uris, Richard B. 1971. Distance, Intervening Opportunity, City Hierarchy and Air Travel. The Annals of Regional Science, 5(1):152-161.

Long, Wesley H. 1970. Air Travel, Spatial Structure, and Gravity Models. The Annals of Regional Science , 4(2):97-106.

Long, Wesley H. 1970. The Economics of Air Travel Gravity Models. Journal of Regional Science, 10(3):353-363.

Manheim, Marvin L. 1979. Fundamentals of Transportation Systems Analysis. MIT Pres, Massachusetts.

Mathur, Vijay K. 1970. An Economic Derivation of the Gravity Law of Spatial Interaction: A Comment. Journal of Regional Science, 10(3): 403-405.

Mazurkrewicz, L. 1984. Spatial Interaction Modelling: A Statistical Approach Based on the Concept of the Wave Equation. Environment and Planning A, 16(10):1387-1390.

Mazurkrewicz, L. 1985. A Statistical Model of Multiple Spatial Interaction Pattern. Environment and Planning A, 17(11):1533-1539.

Miller, Edward. 1972. A Note on the Role of Distance in Migration: Cost of Mobility Versus Intervening Opportunities. Journal of Regional Science, 12(3):475-478.

Morrill, Richard L.1970.The Spatial Organization of Society. Wadsworth Publishing Company, California.

Nakanishi, Masao and Cooper, Lee G. 1974. Parameter Estimation for Multiplicative Competitive Interaction Model-Least Squares Approach. Journal of Marketing Research, 11:303-311.

Niedercorn, J H and Bechdolt, B V. 1969. An Economic Derivation of the Gravity Law of Spatial Interaction. Journal of Regional Science, 9(2):273-282.

Niedercorn, J H and Bechdolt, B V. 1970. An Economic Derivation of the Gravity Law of Spatial Interaction: Reply. Journal of Regional Science 10(3):407-410.

O'Sullivan, Patrick.1981. Geographical Economics. McMillan Press. London.

Olsson, Gunnar. 1970. Explanation, Prediction, and Meaning Variance: An Assessment of Distance Interaction Models. Economic Geography, 46(2):223-233.

Oppenheim, Norbert. 1980. Applied Models in Urban and Regional Analysis. Prentice-Ha;;, New Jersey.

Pederson, P. 1970. Diffusion Within and Between National Urban Systems. Geographical Analysis, 2:203-254.

Pindyck, Robert and Rubinfeld, Daniel. 1981. Economic Models and Economic Forecasts. McGraw-Hill, New York.

Pollard, Ernest C and Huston, Douglas C. 1969. Physics: An Introduction Oxford University Press, London.

Roy, J R and Brotchie, J F. 1984. Some Supply and Demand Considerations in Urban Spatial Interaction Models. Environment and Planning A, 16(9):1129-1135.

Ruiter, Earl R. 1967. Toward A Better Understanding of the Intervening Opportunities Model. Transportation Research, 1:47-56.

Rushton, Gerard. 1969. Analysis of Spatial Behavior by Revealed Space Preference. Annals, 59:391-400.

Sheppard, Eric S. 1978. Geographic Potential. Annals of the Associations of American Geographers, 69(3):438-447.

Sheppard, Eric S. 1978. Theoretical Underpinnings of the Gravity Hypothesis. Geographical Analysis, X(4):386-402.

Sheppard, Eric S. 1979. Gravity Parameters Estimation. Geographical Analysis, 11(2):120-132.

Sheppard, Eric S. 1979. Notes on Spatial Interaction. The Professional Geographer, 31(1):8-15.

Shneider, Morton. 1959. Gravity Models and Trip Distribution Theory. Regional Science Association, 5:51-56.

Siegel, Jay and Woodyard, Monty. 1974. Position in the Urban Hierarchy as a Determinant of Integration. Land Economics, L(1):75-82.

Smith, Terence R. 1981. A Family of Spatial Interaction Models Incorporating Information Flows and Choice Set Constraints Applied to U.S. Interstate Labor Flow. International Regional Science Review, 6(1):15-31.

Stetzer, F. 1976. Parameter Estimation of the Constrained Gravity Model: A Comparison of Six Methods. Environment and Planning, 8:673-683.

Stewart, John Q. 1947. Empirical Mathematical Rules Concerning the Distribution and Equilibrium of Population. The Geographical Review, 37(3):461-485.

Stewart, John Q. 1948. Demographic Gravitation: Evidence and Applications. Sociometry, 11(1):31-58.

Stouffer, A S. 1940. Intervening Opportunities: A Theory Relating Mobility and Distance. American Sociological Review, 5(6):845-867.

Taaffe, Edward J. 1973. Geography of Transportation. Prentice-Hall, New Jersey.

Taylor, Peter J. 1971. Distance Transformation and Distance Decay Functions. Geographical Analysis, 3(3):221-238.

Tomazinis, A R. 1962. A New Method of Trip Distribution in an Urban Area. Highway Research Board Bulletin 347, Highway Research Board Washington, D C. pp.77-99.

U S Department of Commerce. 1965. Calibrating and Testing a Gravity Model for Any Size Urban Area. Bureau of Public Roads, Washington.

Wagner, D J and Hawkins, A F. 1970. The Calibrations of the Distribution Model for the Selnec Study. Transportation Research, 4:103-112.

Webber, M j. 1948. Explanation, Prediction and Planning: The Lowry Model. Pion Limited, London.

White, Roger W. 1976. A Generalization of the Utility Theory Approach to the Problem of Spatial Interaction. Geographical Analysis, 8(1):39-46.

Willenkens, Frans J. 1983. Specification and Calibration of Spatial Interaction Models. A Contingency Table Perspective and an Application to Inter-urban Migration in Rotterdam. Tijdschrift voor Econ.en Soc. Geografie 74, Nr.4 pp. 239-252.

Wilson, A G. 1967. A Statistical Theory of Spatial Interaction Models. Transportation Research 1:395-400.

Wilson, A G. 1970. Advances and Problems in Distribution Modelling. Transportation Research, 4:1-18.

Wilson, A G. 1971. A Family of Spatial Interaction Models and Associated Developments. Environment and Planning, 3:1-32.

Wilson, A G. 1975. Some New Forms of Spatial Interaction Models: A Review. Transportation Research, 6:167-179.

Young, E C. 1924. The Movement of Farm Population. Ithaca, Cornell Agricultural Experiment Station, Bulletin 426.