AN INCOME DISAGGREGATED MODEL
OF
URBAN SPATIAL STRUCTURE

By

© DOUGLAS MALCOLM MUNRO

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Arts

McMaster University

September 1980

MASTER OF ARTS (1980)
(Geography)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: An Income Disaggregated Model of Urban Spatial Structure

AUTHOR: Douglas M. Munro, B.A. (Honours) (University of Western, Ontario)

SUPERVISOR: Dr. M.J. Webber

NUMBER OF PAGES: vii; 208

ABSTRACT

The task of this paper is to evaluate, in some detail, the effects of income disaggregation upon the predictive and descriptive abilities of a comprehensive, information-minimising, spatial interaction model. The model is comprehensive in that it predicts the probability that an individual household chooses a particular residential location, work place and shopping trip pattern subject to expectations based upon survey data about the average cost of shopping and work trips and the average number of shopping trips per week per household. This probability is information minimising compared to a prior probability distribution which is chosen to reflect residential land availability.

The purpose of the analysis is to determine whether or not the income group disaggregated model provides a more accurate representation of observed spatial structure than does the aggregate model (treating the sample data as a homogeneous group).

It is concluded that although some descriptive advantages accrue to the disaggregated model, there is generally little predictive advantage to be gained by income group disaggregation unless the specific aim of the study is to discern the differences that exist in locational and trip making behaviour between income categories.

ACKNOWLEDGEMENTS

This thesis has drawn upon the resources of an ongoing research project; a project initiated by Dr. Michael Webber. My sincere thanks are extended to all of the people involved with the study; especially to Michael Webber and Morton O'kelly who offered me their advice and stimulated my interest in the subject. I would also like to express my gratitude to Dr. Liaw and Dr. Hall for providing their assistance and encouragement when I needed it most.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

AN INTRODUCTION TO URBAN MODELLING

The processes which influence the location of activities in urban areas have been of particular interest to planners and scientists in various disciplines for almost as long as cities have existed. However, it was not until the advent of digital computer technology and associated information systems that it became possible to even consider the construction and empirical testing of operational urban models which sought to explain the location of these activities (Putman, 1975, 187). The urban transportation studies of the 1950's marked the beginning of a period of intense urban modelling activity in geography and related disciplines which continues apace to this day.

The field of urban modelling has undergone considerable evolution and progress since its inception over twenty years ago. Within the North American context most of these modelling efforts have been in direct response to urgent planning problems experienced in urban environments: traffic congestion, conflicting land use patterns, central city out-migration, suburbanization and what is loosely referred to as urban sprawl. Voorhees' (1955) paper outlining a general theory of traffic movement served to demonstrate that theories and models of traffic flow could be used by planners to provide information about traffic problems and possible solutions to these problems. Voorhees' work marks the beginning of this period of development of systematic methods of land use and transportation planning. Subsequent large-scale transportation studies initiated the development of operational models based upon theories of traffic flow.

The relationship between traffic and land use was also an area of debate during the 1950's. As McLoughlin (1969) points out, many early transportation studies treated land use as a peripheral concern. This is not surprising since these early modelling efforts were largely in response to immediate and pressing problems of traffic (traditionally the prerogative of engineers) and precluded a complete examination of interrelated issues. The work of Mitchell and Rapkin (1954) did much to convince engineers of the importance of land use as a generator of traffic and by the late 1950's the configuration of land use was an explicit consideration in planning models. This naturally led to the idea that land use might also be modelled using appropriate theoretical structures, and in 1959 the first land use models were attempted by Hansen (1959) in Washington, Harris (1960) in Penn-Jersey and Hamburg and Creighton (1959) in Chicago. These works set the stage for the modelling efforts of the 1960's.

Since then the field of urban modelling has expanded into a multi-disciplinary science and encompasses a number of disciplines including engineering, economics, urban planning and geography. Urban model building is now firmly established within geography, reflecting the profound influence that urban analytical interests have had on the discipline. A matter of common concern among the various disciplines which deal with urban modelling and take the city as the prime system of interest is modelling methodology. This, then, is the context for a discussion of urban modelling within this paper. Of particular interest here will be the design and use of operational urban models, i.e. models which are applied using real world data. The specific focus of this thesis is the development and empirical testing of a disaggregated spatial interaction model of urban spatial structure.

This thesis attempts to examine, in as comprehensive a way as possible, the design of operational urban models, that is, mathematical models of cities designed for solution on a digital computer. It is possible to develop a perspective on urban modelling by examining the association between certain theories and models but, although such relationships will be alluded to throughout this thesis, a more appropriate focus here is upon the design of operational models. The strong practical-procedural motivation in the historical development of land use and transportation models has already been mentioned. Many of the critical issues involved in constructing and applying operational urban models relate to methodology rather than substantive theory, hence the emphasis adopted here.

From one point of view the subject area is quite restrictive in that the models and issues which will be outlined relate mainly to the operationalisation of spatial interaction and economic models; models concerned with the spatial configuration of the urban system, which say little about the aspatial aspects which influence the structure of the urban environment. Yet the restrictive nature of the discussion still permits some generalization since the difficulties faced in developing operational urban models are similar to the problems faced in developing other models. This will allow some conclusions to be drawn about the general issues involved in developing models and methods for quantitative urban analysis.

The recent advance of such models and methods of urban analysis are well known and documented (Haggett, 1965; Batty, 1976; Carrothers, 1956; Harris, 1968; Wilson, et. al: 1977). During the 1960's most such work in geography dealt with statistical methods. Lowry's model of metropolis (Lowry, 1964) marked the beginning of a stream of mathematical modelling activity. To many this indicated a new stage of development in the discipline of urban geography as represented by a shift from inductive to deductive methods (Wilson et. al. 1977, 1; Batty, 1976, 3).

Paralleling this stage of development was a concern for consistent methods of building mathematical models. Of considerable impact was the application of the concept of entropy.

Recent developments in urban modelling have demonstrated the utility of models based upon the entropy-maximising formalism. Wilson's use of the paradigm to derive a consistent family of spatial interaction models is perhaps the most widely recognized and easily understood application within the discipline (Wilson, 1970a, 1974). The entropy-maximising method provides a useful and practical model building tool which can be applied to the development of a wide class of urban models. As Wilson (1970a, 11) points out, the entropy-maximising procedure "enables us to handle extremely complex situations in a consistent way, and past experience has shown that this sort of consistency is very difficult to achieve otherwise."

The approach taken by Wilson employs a statistical mechanical definition of entropy as applied to aggregate geographical concepts. A more powerful, alternative basis for the entropy concept can be found in information theory (Webber, 1977). This alternative information-minimising approach is used to derive the model developed in this paper.

Batty (1976, 10) has criticised recent modelling efforts for their lack of theory, and points out the inherent difficulty of testing theories in the social sciences; i.e., that cause-effect mechanisms are compounded by many factors. The fundamental limitation in urban modelling is the inability to hold all variables but one constant so that isolated effects on the system can be traced. Consequently, most urban modelling research has been concerned with the development of ad hoc models of particular urban subsystems which have been tested separately, usually on different study areas.

Nevertheless, there have been several ambitious efforts which have focused on the development of comprehensive, theoretically based models; models incorporating several rather complex ideas or hypotheses embodied within a single theory. It is here that the information-minimising method is most useful, by providing a paradigm for constructing such models in an internally consistent manner.

The model presented in this paper is a comprehensive, information-minimising model which describes aggregate characteristics of an urban population. It is an operational model which predicts residential location, work trip and shop trip behaviour using data for the city of Hamilton, Ontario. The model employs the information concept to assign individuals to a set of categories in a way which is maximally noncommittal with respect to missing information and at the same time does not contradict given data. The following generalization of model types should help to clarify where this model fits within the overall context of urban models.

If we categorize a city into three basic components; people, interaction and facilities, then we may distinguish between three basic types of models. They are (i) short term spatial interaction models in which interaction varies, (ii) medium term residential location models in which people and interaction vary, and (iii) long term models in which people, facilities and interaction vary. The model developed in this study is a medium term model.

This model is an extension of the modelling efforts of Webber, et. al. (1979b) and incorporate several important improvements:

(i) better quality disaggregate data and expanded study area,

(ii) increased number of retail facilities,

(iii) improved parameter search, and

(iv) income group disaggregation.

To a large extent the research behind the present model and previous model versions stems from the theoretical work of Webber (1979a, 1979b, 1979c) and encompasses issues raised in location theory. The criteria used to evaluate the theory and its models are those of realism and veracity of predictions rather than ease and cheapness in forecasting.

Unfortunately, the previous model versions have ignored the possibility of very realistic differences in behaviour that may exist between urban consumers in different income groups. By disaggregating the model according to income group the hypothesis is being presented that people in different income groups have different expenditure on housing and travel, and exhibit different travel patterns for both work and shop trips. If this is the case then disaggregation by income category would improve model realism by providing different cost constraints for each income group.

The major focus of this thesis will be to investigate the predictive and descriptive accuracy of this disaggregated model which takes into account the differences shown to exist between various income categories. For each income category the model determines the residential location of households, the distribution of work trips and the shopping trip patterns, subject to observed work and shop trip costs and to the expected number of shopping centres visited per week per household. The distribution of workplaces and shopping facilities is exogenously given.

The important question to be addressed is; "Does the income disaggregation improve the residential location, work trip and shop trip interchange predictions ?" Hathaway (1975) has examined this problem using a conventional, doubly-constrained, trip distribution model disaggregated by: (i) age, sex and marital status, (ii) socioeconomic group, (iii) occupational classification, and (iv) standard industrial classification. Hathaway argues that the potential advantage of the disaggregate model is its ability to forecast the change in travel patterns resulting from a change in the relative proportions of different categories of trip-makers. Hathaway concluded that there was little advantage to be gained from the disaggregation using a simple trip distribution model unless considerably more detail is introduced to the model (detail in this case could be interpreted as the number of constraints).

It seems reasonable, then, to postulate that a more comprehensive model (a medium term model containing more constraints), such as the one presented in this paper, will derive greater benefit from disaggregation than the simple model used in Hathaway's study. Also, previous experience with earlier versions of the model developed in this paper has indicated that a possible point of departure for further improvements would include a finer classification of the actors within the modelled urban system. The remainder of this thesis is organised as follows.

Chapter Two discusses several aspects of urban spatial modelling, including a brief introduction to the use of entropy and information theory in urban analysis and a discussion of aggregation issues. Chapter Three describes the model in detail including the methods of household categorization , prior probability formulations and constraint specifications. A review of previous model versions and results is given. Also, a discussion of the calibration methods and solution procedure algorithms is presented. Chapter Four describes the new survey data and outlines the methods used in the difficult task of processing the raw data into a form which is compatible for input to the model.

Details of the analysis are contained in Chapter Five. The strategy for analysis is briefly outlined here. First, a preliminary investigation is conducted of the premise that significant differences in the work and shop trip distributions and shop trip frequency exist between income categories of the population. The mean value estimates and their variances are examined. A t-test is performed to establish whether or not the differences in the estimated mean values between categories are significant. At this point it may be necessary to adjust income class intervals to better separate the income categories. Next the model is calibrated using the total sample data (treating the sample population as a homogeneous group) and for each separate income category. Testing the model involves investigating the goodness of fit of the model predictions to the observed data. This includes a comparison of observed and predicted residential location, work trip interchange and shop trip interchange matrices for each income group. Since it is not known a priori which statistic is most appropriate, several goodness of fit statistics are computed for each pair of observed and predicted matrices. The final and conclusive test of the model is to examine the fit of the sum of the predicted income-specific matrices to the aggregate observed probability matrices using the same goodness of fit measures as before. Again, the important question to be addressed is; "Does the income disaggregation improve model predictions ?" If the answer is "yes", then other requirements will have to be met before it could be stated that the disaggregate model is a significant improvement over the aggregate model. Of key importance is the trade-off between the cost of using the model (time, effort, data requirements) and the gain in predictive accuracy.

Finally, Chapter Six discusses the results and implications of the modelling exercise in light of other recent research. Insights gained and possible departure points for further research are presented.

CHAPTER TWO

CONTEMPORARY ISSUES IN URBAN SPATIAL MODELLING

2.1 Introduction: Models and Theories in Urban Analysis

The strong practical and procedural motivation for the development of operational urban models has been briefly touched upon in the previous chapter. Modelling methodology and design issues are of key importance to the development of all manner of urban models including the spatial interaction model developed in this paper. This, then, provides the context for the discussion of contemporary urban modelling that follows.

The focus of this paper is upon operational urban models and more specifically the development and testing of an operational, information-minimising model of urban spatial structure applied using data for the city of Hamilton, Ontario. The model to be presented in this paper reflects an attempt to design a more realistic and relevant urban model which integrates features of the theoretical work of Webber (1979a, 1979b, 1979c) and others within an improved operational framework.

This chapter will address in detail two topics which relate directly to this model and are of significance to the model building process in general: 1) entropy and information theory, and 2) disaggregation issues. However, before proceeding to a more substantive discussion of these two topics it is useful to elaborate the model building process in a little more detail. To this end it is necessary to provide some general definitions which relate to the structure and purpose of urban models within the social science context. The model building process which may be considered an integral part of what is loosely called "the scientific method", is first discussed and the conceptual relationship between theory and model is stressed. The way in which the model building process fits within the theory building process and vice versa is then illustrated. This provides a structure in which to develop a more focussed discussion of issues pertaining directly to the model developed later in this thesis.

In a general sense the term 'model' refers to some well-defined process which may be used to develop a deeper (theoretical) understanding of some real world phenomena. Of particular concern here is the structure of models since it is this structure which is manipulated in the model building process. Model components or 'variables' represent the observable elements in the theory on which the model is based and are related in a formal sense by equations. Already it can be seen that the notions of theory and model are complementary and interrelated by the process of model building.

The model building process may be considered as a reflection of the scientific method in which theory is tested and refined (Batty, 1978, 67). In essence a theory is an abstraction from the complexity of the real world. Theories provide statements of order or pattern by simplifying and extracting the essential order from some complex real world phenomena. This is generally recognized as being useful by aiding in our understanding of these processes. Most theories are formulated by setting hypotheses from which testable propositions can be extracted. It is useful to express theories as sets of testable propositions which may or may not be refuted by comparison with reality. Here, then, is the link between theory and model, for the concept of a model hinges around the notion of a testable proposition.

Operational models are indispensable tools for testing theory especially in the social sciences where direct experimentation with the system of interest (i.e. cities) is impossible. A working definition of a model might be taken as a tool for testing theory in an artificial (symbolic, mathematical) environment (Batty, 1978, 68). In this sense Harris' (1966) definition of a model as "an experimental design based on theory" indicates one role of urban modelling, i.e. the translation of a theory into a form which is testable.

This might lead to the assumption that the prime role of operational urban modelling is in testing theory, in explaining, generalizing and understanding urban phenomena, but this need not be entirely so. Not all operational urban models are based upon theory. Many such models rely solely on statistical regularity (for example, the model developed by Chapin and Weiss, 1962). The presence or absence of theoretical foundations in urban models ultimately depends upon the particular function of the model. For example, in planning the prime role of modelling may be simply to provide some structure or starting point for solving an inherently difficult problem. In contrast to a model which is designed to test a specific theory, a planning model need not be 'correct' in any testable sense for it to be useful.

It can be seen that the purpose for which a model is constructed has a pronounced effect upon the design of that model. Model purpose also serves as a useful basis for classifying types of urban models. A complete discussion of design issues and methods of categorizing urban models is both out of place and beyond the scope of this section. A wide variety of model classification schemes exists in the literature and the interested reader is referred to the reviews by Senior (1973, 1974), Lowry (1968), Lee, D. (1973), Lee, C. (1973), Putman (1975) and Batty (1976). Methodology and design issues in urban model building are of particular concern here and several authors have provided guidelines for operational modelling research (Batty, 1976; Wilson, 1974a; Broadbent, 1970). To avoid ambiguity a discussion of such issues will be presented in later sections of this paper where the model developed in this study will serve as an example in operational model design. The important differences between theoretical and operational urban models will now be illustrated with several examples.

Examination of recent urban modelling literature reveals a strong distinction between theoretical and operational models (Wilson, et. al., 1977; Wilson, 1979; Batty, 1976, 1978). Theoretical models are essentially statements of urban phenomena, the design of which is governed by rigorous mathematical analysis (axiomatic rather than empirical) of their causal structures. These models tend to be deeply rooted in theoretical bases such as location theory (Alonso, 1964; Muth, 1969; Hoover, 1948; Weber, 1929), microeconomic theory (Isard, 1956; Greenhut, 1963) or consumer behaviour; their formulation is mathematically elegant and often quite abstract.

In contrast, operational models are designed to be tested empirically using real work data and are often regarded as statistical descriptions of urban phenomena which focus upon estimating a model for the purpose of comparing the models' predictions to observed data. A direct comparison of operational models with theoretical models can be seen by contrasting the models in the book by Batty (1976) with those in the book by Papageorgiou (1976). Although operational models may imply certain theoretically based hypotheses, they are generally incapable of incorporating all the richness of detail found in their theoretical counterparts. However, they do provide a means of verifying assumptions and underlying hypotheses via empirical testing. Examples of operational models designed for this purpose are Herbert and Stevens' (1960) use of linear programming to develop Alonso's scheme and Harris' (1962) work in estimating the utility functions involved.

The attempts at operationalizing central place theory, first postulated by Christaller (1933) and Losch (1944) is another example worth illustrating briefly. Only rudimentary testing of the theory was attempted during the 1950's since techniques of analysis had not progressed further than linear statistical models (Berry, 1967). More complete testing of the theory became possible in the 1960's with the development of retail models (Huff, 1963; Lakshmanan and Hansen, 1965). Even these models are highly simplified representations of the original theory and Batty (1978, 70) argues that a real test of the theory has not yet been performed since present-day retail models are lacking many of the notions embodied within the original theory (e.g. hierarchy, product differentiation, demand and price, externalities).

The ability to test operational models renders them useful for another purpose - as potential urban planning tools. The role of models in planning has been addressed by a number of authors (e.g. Lee, D. 1973). Perhaps the most widely recognized contribution in this area is the Lowry model (1964) and its predecessors. The line of operational models developed in the tradition of the Lowry model are too numerous to mention here. A summary of these models in the North American context is given in Putman (1975) and Batty (1976); British applications have been documented in several reviews, including Broadbent (1970) and Batty (1976). The flurry of modelling activity in urban planning over the last two decades reflects the fact that society is becoming increasingly aware of the need for effective urban planning and management. Operational models have been used to assess the impact of increased intensity of land use both before and after the fact, and to evaluate the consequences of different development schemes. If dependable statements are to be made regarding the effects of land use practices upon the urban environment then the most scientifically based methodology should be applied.

Against this background the remainder of this chapter focusses upon two selected topics which are of importance to urban modelling in general and relate directly to the operational model presented later within this paper: 1) entropy and information theory; and; 2) aggregation/disaggregation issues. Each section will progress from a general discussion of the topic to more specific matters pertaining to the model presented in Chapter Three. In so doing, the discussion will develop the rationale behind this research and elaborate upon the important methodological considerations including: modelling objectives and choices, model derivation, scale of analysis and inherent limitations and difficulties. To fix ideas regarding aggregation/disaggregation issues, section 2.4 provides a brief review of Hathaway's approach to disaggregating a conventional, doubly-constrained spatial interaction model (Hathaway, 1975). Hathaway's findings will be outlined in order to provide a comparative basis for the analysis presented in Chapter Five.

2.2 Information Theory and the Entropy-maximising paradigm

Wilson's seminal papers of the early 1970's served to spark interest in the entropy-maximising methodology (1969, 1970b, 1971, 1973). Since then the entropy concept has found its way into a variety of geographical applications, usually as a model building tool or as an objective measure of some system property (Webber, 1975; Senior, 1976; Cesario, 1973, 1975; Batty, 1976; Gould, 1972). The concept has always had some mystery associated with its use in the social sciences and as Mogridge (1972) points out the concept is difficult to define unambiguously and comprehensively; confusion has resulted from a wide range of applications in different contexts. Wilson's (1970) use of the concept of entropy has predominantly been for the development of hypotheses or theory building (according to Wilson's definition, model = hypothesis, theory = well-tested hypothesis).

Perhaps the earliest applications of entropy-maximising methodology within the field of urban modelling was Wilson's use of the technique to derive the classical gravity models (reviews of the historical development of the gravity model may be found in Olsson (1965) and Carrothers (1956). Early gravity models were a direct analogy to Newtonian physics and heuristic derivations of these models within the social science context took the "ready made" Newton model and modified it to render it internally consistent with respect to accounting constraints. Incorporated within these models was a constant of proportionality; a necessary factor which scaled the total predicted interactions to the observed total. This scaling factor was necessary to overcome the deficiency resulting from the multiplicative structure of the classical gravity model, i.e. a doubling of flows for any particular origin or destination resulted in a quadrupling of the number of trips between these zones, when only a doubling of flows was expected. Also, as Senior (1976) points out, the use of classical gravity models has been beset with other difficulities. The application of linear regression calibration techniques to intrinsically non-linear gravity models has cast some doubt upon the validity of results obtained by some authors (eg. Chisholm and O'sullivan, 1973; Cliff et. al., 1974).

Wilson's entropy-maximising derivations put a whole family of gravity models on a sound and common theoretical basis by deriving them mathematically using statistical mechanical methods. The advantages to be gained from the entropy-maximising derivation of the gravity model as opposed to the heuristically derived gravity model include the following.

(i) The entropy-maximising method builds the model from first principles expressing the model as a constrained optimisation problem. It does not "patch up" a ready made model.

(ii) The entropy-maximising derivation aids in the interpretation of perceived travel costs since constraint equations are associated with the form of the impedence function.

(iii) The balancing factor terms come directly out of the derivation.

(iv) The method facilitates the development of more complex models by allowing for the incorporation of additional constraints in an internally consistent manner. These constraints specify explicitly our information or hypothesis about the system.

As Evans (1969) points out, the consistency achieved by using entropy-maximising methods is of great benefit but the modeller is still faced with the difficult task of ensuring that the constraints rigorously describe this knowledge. Advantage (iv) is of particular importance with respect to the comprehensive model developed later in this paper. The versatility of the entropy paradigm allows the incorporation of different costs constraints for each income group within the disaggregate model.

The entropy-maximising method used by Wilson is based upon a technique in statistical mechanics known as the microcanonical ensemble. The logical principle, within a trip distribution matrix framework, concerns the most probable distribution of micro-states (element values) which are consistent with partial macro-state information (row and/or column totals). The probability distribution with the greatest number of micro-states maximises entropy (uncertainty) because it is here that we are most uncertain about the micro-states of the system (i.e. this probability distribution contains the largest number of micro-states which are assumed a priori to occur with equal probability).

By choosing $\{T_{ij}\}$ to maximize

$$W\{T_{ij}\} = \frac{T\,!}{\prod_{ij} T_{ij}!} \tag{2.1}$$

where T_{ij} = the (i, j)th element of the trip distribution matrix,

$\{T_{ij}\}$ = any matrix satisfying row and/or column totals,

$W\{T_{ij}\}$ = a measure of entropy, and

$T = \sum_{ij} T_{ij}$,

subject to linear constraints formulated from the expected value functions (column and row totals, based upon observed data) we obtain the maximum entropy of the distribution (which is monotonically related to the probability distribution). Tribus (1969) shows how different probability distributions are generated by different formulations of constraint equations. The multipliers (or functions of them) are associated with the constraint equations and can be obtained by solving for the constraint equations.

The statistical mechanics approach to entropy taken by Wilson utilizes the concept as a measure of the probability associated with a state of the system - its logarithm is proportional to the number of microstates that can give rise to that state. The fundamental assumption is that given the information available, all microstates are equally probable. This statistical mechanics definition of entropy is parallel to a technique in statistical mechanics called the microcanonical ensemble. Wilson (1968) points out the analogy to entropy-maximising methods and in a later paper (1970) describes the use of the Darwin-Fowler method (a technique found within statistical mechanics which does not rely upon the use of Sterling's approximation) to achieve virtually the same results as entropy-maximising methods used in urban and regional analysis. The main features of the statistical mechanics approach to entropy are outlined in Fast (1970); Wilson (1970) and Levine and Tribus (1979); an elementary review is given in Cesario (1975). Georgescu-Roegen (1971, Chapter 6) shows that there are difficulties associated with the statistical mechanics approach which can be avoided by using an alternative approach based upon information theory.

The entropy concept as found within information theory represents expected information. It is viewed in terms of a random variable taking various values with unkown probabilities. The aim is to make the best estimate of a probability distribution subject to given information. For example, what is the probability of workers living in zone i and working in zone j incurring work cost c_{ij}, given that $\bar{c}$ is the mean cost for all workers ? Shannon and Weaver (1949) have presented a unique and unambiguous measure of the amount of uncertainty associated with a discrete probability distribution that can be used to answer the above question. A discussion of this measure will be presented shortly.

A brief note should be made of one other (subjective) view of entropy before proceeding to a more in depth discussion of the information theoretic bases of entropy. This interpretation (Wilson, 1970) relates entropy to Bayesian statistics (personal, subjective, "degree of belief" probability). In Bayesian statistics the parameter about which an inference is made (e.g. μ, σ) is regarded as a random variable which has a distribution of its own. Also, Jaynes (1957a) has related the information theory and statistical mechanics views of entropy with a subjective rather than relative frequency view of probability.

The information theory interpretation of entropy is used in the development of the model presented in this paper. This interpretation of entropy is due to Jaynes (1957a, 1957b) and is based upon concepts of probability theory. Let X be a random variable which can take values $(x_i, \ldots, x_n)$ with a corresponding probability distribution

$$\underline{\underline{p}} = (p_1, \ldots, p_n)$$

which is unknown. p_i is the probability of X being in the state x_i. Assume that the states are a priori equi-probable. What we know is the expectation value

$$\sum_i p_i \ f(x_i) = E[f(x)] \qquad (2.2)$$

The form of (2.2) is similar to the formulation of the constraints of the entropy-maximising model discussed later in this paper. Additional information include

$$p_i > 0, \ \forall_i \qquad (2.3)$$

and

$$\sum_i p_i = 1. \qquad (2.4)$$

With respect to this information only how may we calculate $\underline{\underline{p}}$?

Jaynes (1957a, 1957b) proposes that Shannon's measure of uncertainty (Shannon, 1948) provides us with an unambiguous criterion for the amount of uncertainty in a discrete probability distribution which avoids bias and yet agrees with prior information. This measure of uncertainty (or entropy) of $\underline{\underline{p}}$ is

$$I_s = S(p_1,\ldots,p_n) = -k \sum_i p_i \ln p_i . \tag{2.5}$$

Jaynes' rationale for solving the problem was to maximise (2.5) subject to (2.2) and (2.4).

$$\max. \quad I_s = -k \sum_i p_i \ln p_i$$

$$\text{s.t.} \quad \sum_i p_i f(x_i) = E[f(x)]$$

$$\sum_i p_i = 1.$$

The solution to the problem is:

$$L = -\sum_i p_i \ln p_i + \lambda(1 - \sum_i p_i) + \mu(E[f(x)] - \sum_i p_i f(x))$$

where λ and μ are the multipliers. Differentiating we obtain

$$\frac{\partial L}{\partial p_i} = \frac{\partial}{\partial p_i} \left(- \sum_i p_i \ln p_i\right) - \lambda - \mu(f(x_i))$$

$$= - 1 - \ln p_i - \lambda - \mu \, f(x_i) = 0.$$

Absorbing -1 into λ and rearranging gives

$$\ln p_i = - \lambda - \mu f(x_i)$$

therefore

$$p_i = \exp(- \lambda - \mu \, f(x_i))$$

Substituting into (2.4) we obtain

$$e^{\lambda} = \sum_i \exp(- \mu \, f(x_i))$$

This information theoretic measure of entropy is equivalent to that derived by Wilson using aggregate concepts. Define:

$$p_{ij} = \frac{T_{ij}}{T} \qquad (2.6)$$

Using Stirling's approximation and a monotonic function of (2.1) (conversion to logs is most convenient so that terms will be additive) we know that

$$\ln W = \ln T! - \sum_{ij} T\, p_{ij}\, [(\ln p_{ij} + \ln T) - T\, p_{ij}]$$

$$= \ln T! - \sum_{ij} (T\, p_{ij} \ln T\, p_{ij} - T\, p_{ij})$$

$$= \ln T! - \sum_{ij} T\, p_{ij} \ln T - \sum_{ij} T\, p_{ij} \ln p_{ij} + \sum_{ij} T\, p_{ij}$$

$$= \ln T! - T \ln T + T - T \sum_{ij} p_{ij} \ln p_{ij}$$

Therefore the Shannon measure of uncertainty is analogous to Wilson's formulation of entropy. However, there are definite advantages to using Shannon's measure:

(i) Stirling's approximation for factorials is not relied upon,

and (ii) the formulation based upon information theory is much more flexible and has many consistent theoretical properties.

These advantages and other properties relating to the Shannon measure will not be elaborated upon here. The reader is referred to Mathai and Rathie (1975), Khinchin (1957), Kullback (1959), Cox (1961) and Tribus (1969).

According to Walsh and Webber (1977) the difficulties in generalising I_s (which was derived axiomatically) to continuous probability distributions led to an alternative measure, the Kullback Information Gain (hereafter referred to as I_k), which depends upon two probability distributions. Snickars and Weibull (1977) have criticised I_s for assuming that every micro-state is equi-probable. They argue that the assumption is sensitive to the choice of micro-state space; that is, that the probability assignment at the macro level depends upon the selection of micro-state space. They too propose the I_k as a better alternative. The information measure is formulated as follows:

$$I_k(\underline{\underline{P}};\underline{\underline{Q}}) = - \sum_i p_i \ln \frac{p_i}{q_i} \tag{2.8}$$

where $\underline{\underline{Q}}$ is the a priori most probable probability distribution (prior). It can be seen that I_s is a special case of I_k (assuming identical configuration of micro-state space) in which the a priori most probable distribution is the uniform distribution

$$\underline{\underline{P}}^\circ = \underline{\underline{Q}} = (\frac{1}{N},\ldots,\frac{1}{N}).$$

If there is no prior information then this assumption regarding $\underline{\underline{Q}}$ is warranted. However, if there is prior information, then by not including it in $\underline{\underline{Q}}$ or by not adjusting the micro-state space to make it consistent with $\underline{\underline{Q}}$, we are in fact using an entropy derivation which is biased with respect to prior information.

The functional form of the objective functions used in the model developed in this study (to be described in the next chapter) is the same as (2.8). When relating to a model which employs a measure of entropy in the form of (2.8) we should strictly talk in terms of information-minimising rather than entropy-maximising. Some confusion may result with respect to the sense of the optimisation involved but as Macgill (1975) points out, since these models are formulated as mathematical programming problems, the sense of optimisation may be reversed simply by reversing the signs on the objective function. Thus by dropping the negative sign on the entropy function (2.8) the formulation consists of minimising a convex objective function over convex constraint sets (convex because they are defined in all cases only by linear equality or inequality constraints). Evans (1973) provides a proof that the entropy function is strictly convex. Coelho and Wilson (1977) demonstrate the uniqueness property for solutions of various formulations of mathematical programming problems based upon entropy-maximising models, following directly from classical programming theory (Hadley, 1964).

The I_k form of the entropy function (2.8) has been most vigorously promoted by entropy model users because it can be argued to have more general properties and significance than (2.5). In loose terms, the crux of the argument is that information can only sensibly be defined in relative and not absolute terms, hence the need for prior values against which to relate the estimated values. In cases when we have no prior estimates we use the uniform distribution. Of course, the selection of what constitutes priors depends upon the model framework. For trip distribution modelling we might consider an interaction matrix from a previous time period. Within the location modelling context one could incorporate priors based upon land use availability or zonal area; or other information regarding a zonal system such as the level of spatial aggregation (see Batty and March, 1976b; Batty and Sammons, 1980). Detailed (non-geographical) descriptions of the theoretically significant properties of I_k are expounded in Kullback (1959), Good (1956), Hobson (1969), Tribus (1969), and Mathai and Rathie (1975). A comparison of I_k and I_s can be found in Hobsen and Cheng (1973). Renyi (1970) presents a heuristic argument illustrating the nature of the change which occurs when moving from a prior to a posterior distribution. Tribus and Rossi (1973) discuss this change with respect to path dependence in I_k.

I_k has demonstrated its usefulness in many urban modelling contexts including applications by Theil (1972), Cesario and Zerdy (1975) and Batty and March (1975). Cesario (1975) provides an excellent numerical example illustrating the case of I_k, priors formulations and the effect of varying the size of micro-state space. Perhaps the most widely adopted but unrecognized use of I_k is in bi-proportional matrix adjustments (adjusting a matrix according to known row and column totals). Macgill (1975, 1977) demonstrates the use of I_k for this purpose within a trip distribution context and also the uses of I_k in balancing factor methods.

In summary, it has been shown by several authors that entropy-maximising or information-minimising methods are useful operational tools in the development and extension of models for spatial analysis. Perhaps its most attractive feature is the ease with which complex models, such as the one presented later in this paper, can be built in an internally consistent manner. Using the entropy methodology we can build aggregate models of situations where statistical averaging over unknown micro-behaviour is involved. Entropy models have been applied to a wide variety of urban phenomena. Wilson (1969, 1970, 1973, 1975) has used the paradigm to develop improved spatial interaction models for several purposes: commodity flows and transportation models (mode and route split); distribution models for person trips (working, shopping, recreation); and, shopping and residential location models. Information theoretic measures have been developed by Walsh and Webber (1977), Batty and Mackie (1972), Batty and Sammons (1979) and Morphet (1975) for various uses in spatial analysis ranging from calibration statistics to measuring the impact of spatial aggregation on a zonal system. Entropy techniques have also been successfully applied to migration models (Stillwell, 1975) and population density models (Bussiere and Snickars, 1975; Cesario and Zerdy, 1975).

Several theoretical approaches to the concept of entropy have been presented and different formulations of entropy functions have been discussed. Although the approach taken to entropy is largely a matter of personal taste (it would suffice to work with any of the interpretations consistently), the information theoretic interpretation of entropy has been adhered to in developing the model presented in this thesis for the sake of realism, i.e. it is a more realistic approach in that real world situations contain limited information regarding systems of interest. A case has been made for the Kullback information measure, I_k. This is the form of the objective function used in the model developed later in this paper.

2.3 Aggregation Issues in Urban Modelling

In this section a discussion of aggregation issues in operational urban modelling is presented. The advantages accrued to disaggregated models, difficulties encountered and the possible types of disaggregation will be elaborated upon. This will set the groundwork for the disaggregated information-minimising model developed in this study.

The form of the model developed in Chapter Three of the thesis is essentially that of a residential location model. As Wilson et. al. (1977) point out, most operational residential location models have been based upon gravity model principles and have involved several simple assumptions. For example, the Lowry (1964) model, albeit a comprehensive model, has assumed that housing supply always meets demand; and, has also taken several variables such as basic employment as exogenously given. The model developed by Lowry has had a pronounced effect upon the development of subsequent residential location model research including the model presented in this paper. For this reason it is worth illustrating the mechanisms of the model briefly.

In the Lowry model the basic sector is defined as that portion of the economy which does not rely upon the distribution of the local population for its locational determinacy (i.e. the industrial sector). The non-basic sector (services) is considered locationally dependent upon population distribution. The model consists of several internally linked submodels which interact in broad terms as follows. First, the population supported by basic employment is calculated given the total basic employment and assuming a particular activity rate. Then total service employment and hence total population is estimated based upon a further assumption regarding population demand for services. The spatial distribution of activities is produced by a gravity model in conjunction with a unique iterative land use accounting mechanism. As a starting point the model assumes that the spatial distribution of basic employment is given, then basic sector workers are allocated to residences and work places. The demand for non-basic services is distributed in a gravity-like manner around the households. These non-basic jobs then generate more population to be located, more services and so on. At each iteration land use constraints must be met.

The Lowry model has served as a predecessor to several modelling extensions including Garin's matrix reformulation (Garin, 1966) and other British applications (Cripps and Foot, 1970; Batty, 1971; Echenique et. al., 1969). All these models have been extremely aggregative, taking no account of spatial differences that exist with respect to household incomes, occupational categories or house types.

A similar lack of disaggregation is evident in operational models developed in the tradition of Alonso (Alonso, 1960; Harris, 1962; Herbert and Stevens, 1960). Gravity-based shopping models such as those developed by Huff (1969) and Lakshmanan and Hansen (1965) also have tended to be too aggregative. Hathaway (1974) has examined disaggregation within the context of trip distribution modelling and has remarked on the lack of disaggregate models used in transportation studies.

The apparent lack of concern for disaggregation is most conspicuous in operationalized aggregate models. As Goldner (1971, 108) points out,

> "It is significant that although the conceptual and experimental modellers call for expanded disaggregation, the operational versions have not pushed far in this direction."

On the other hand, theoretical models have addressed the problem and Wilson (1970a, 1974a) has developed a framework for constructing operational models which incorporate disaggregation in a consistent way. One of the major objectives of this thesis is to fill this gap between theory and practice; to construct and to test a disaggregated, comprehensive model as an extension of the aggregate modelling efforts of Webber et. al. (1978, 1979a, 1979b). This strategy for operational research corresponds with one of the methods proposed by Wilson, Rees and Leigh (1977, 13) for making progress; i.e. to develop existing lines of work by exploring disaggregative possibilities. An alternative point of departure would be to adopt an entirely different modelling methodology. A micro-behavioural approach would employ a disaggregate demand model to deal more adequately with the above mentioned aggregation problem. However, it was decided to retain the aggregate viewpoint and to expand the existing line of research by disaggregation in an attempt to improve model realism. It was the author's opinion that the existing aggregate model research presented the most useful framework in which to employ disaggregation and that, in general, the aggregate model would benefit most from the structural modification. The following statement lends some support to this point of view.

> "While in the micro-approach an emphasis is placed on the analysis of variability in choice contexts, it should be emphasized that this style of work does not have a monopoly of 'behavioural' considerations. In both micro behavioural and what is generally referred to as the 'aggregate' approach, market (or merely population) segmentation is invariably adopted to account for dispersion in the patterns of behaviour arising from (some of) the observable differences between spatial actors. In the past this process has not always been performed particularly well, especially in the traditional aggregate style."

(Williams and Wilson, 1979, 11)

There are several reasons for the existence of dispersion in people's preferences, imperfections in markets, ignorance on the part of the analyst or lack of knowledge on the part of individuals whose behaviour is being modelled. Within the context of the model developed later in this paper, trip making "behaviour" is inferred from the constraints embodied within the model. The purpose of disaggregating the model by income group is to reduce the dispersion in patterns of behaviour by making available new information via income group-specific constraints. This stategy predicates the existence of homogeneous groups of spatial actors; actors whose spatial behaviour is influenced by their income category.

Several options exist for disaggregating aggregate urban models. Disaggregation possibilities may be viewed from a variety of perspectives and relate to important considerations of model design (eg. type of activity being modelled, scale of analysis, modelling objectives). Perhaps the most obvious consideration in disaggregation is model purpose since the type of activity being modelled will dictate those variables which may or may not be segregated for some useful purpose. A housing model may benefit greatly from a classification of houses according to construction style, cost or tenancy status. Similarly, a work trip distribution model may produce some useful information if consideration is made of various modes of transport or a simple white collar/blue collar split among occupational categories of trip-maker. Of course, the specific objectives of the modelling exercise will preclude certain options in each case.

Basically any model dealing with spatial interaction can be disaggregated by purpose, no matter how the flows are manifest in the model: i.e. people, goods, information, money. The model presented in this paper is comprehensive in that it is concerned with the distribution of residences and it makes the fundamental distinction between two trip purposes - shopping and working. This separation of trips reflects the belief that trip-making behaviour is affected by purpose. This idea is explored further in the analysis.

The potential benefits of disaggregation in this model are most pronounced due, in part, to the variety of spatial activities being modelled. The types of disaggregation considered for this model follow from preliminary work, the focus of which was to identify relevant subsystems which were vital to the meaningful application of the model. This was done for 2 reasons: 1) to ensure that the interaction behaviour which is of interest (i.e., residential location, shop and work trip distributions) is completely described by the equation systems of the model, and; 2) to subdivide the model components in such a way as to classify the problem. This is an important aspect of the rationale behind disaggregation - i.e. the model is focussed more acutely thereby making it possible to model the significant interactions more realistically. As Broadbent (1970, 473) points out, structural simplicities are unavoidable in the design of experimental models. Disaggregation is one method of ensuring that the model apparatus has maximum sensitivity in the required directions.

Viewed from the perspective of comprehensive models, disaggregation into subcomponents (by purpose) may be considered as an organisational and operational necessity. The advantage of the comprehensive model structure is that the model embodies more realistic components of urban spatial structure than simple spatial interaction models and avoids oversimplified assumptions. The simplest hypothesis with respect to gravity type residential location models is that households locate around workplaces, or more accurately, according to accessibility to work places (e.g. the residential component of the Lowry model). This hypothesis considers spatial interaction between residence and workplace as the only determinant of residential location (in contrast, the model developed in this thesis takes account of work and shop trip interaction as well as residential land availability). An obvious weakness of this assumption is that there is no provision for households which have no member in the workforce (i.e. unemployed or retired people).

This brings us to another type of disaggregation which is potentially useful, especially in residential location modelling - household or person type classification. Disaggregation according to working status is a useful means of characterizing distinct spatial location behaviour among households and is closely linked to disaggregation according to income group. The problem of unemployed/retired households is addressed in following chapters where it is shown that such households constitute an appreciable portion of the urban population of Hamilton (especially for low income groups).

Disaggregation by commodity type is a familiar form of accounting in shopping models. In this study, shopping trips have been classified according to type of goods purchased. For modelling purposes, grocery shopping trips alone have been considered. Support for this is given by Wilson (1970, 66) who makes the case for disaggregation by class of good, the assumption being that grocery shopping is the most frequently observed type of shopping behaviour and that the trip-making behaviour generated by convenience goods is distinct from shopping for other consumer items.

The level of spatial aggregation is another design consideration within the context of model disaggregation. The description of any trip-distribution pattern by a spatial interaction model is ultimately dependent upon the adequacy of the zonal system used to describe the observed flows. The choice of zone size is an important consideration. If zones are small and the majority of trips are between zones rather than within them, then reducing zone size will do little to improve the description of interaction. Conversely, the description of inter-zonal interaction will be poor if the zones are so large that most interaction takes place within the zones. The implications of zonal resolution to model accuracy are discussed in Wilson (1974b, 169), Broadbent (1970, 473) and Openshaw (1976). The model developed in this paper avoids the problem inherent in large zones by measuring interaction on a neighbourhood basis thereby providing a high level of resolution (181 × 181 matrix). These flows are then aggregated to the zonal scale for operational convenience.

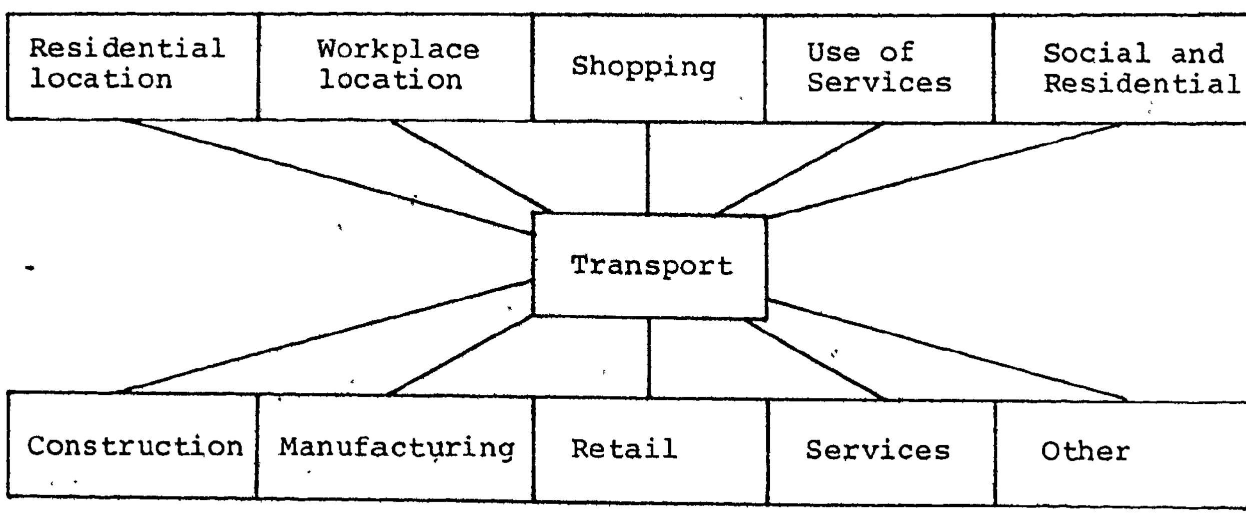
Disaggregation does present some difficulties. Two closely related problems are particularly severe: data requirements and calibration. The topic of data requirements will be addressed here. A discussion of calibration difficulties is presented in the next chapter. Wilson (1970, 1974) has proposed several spatial interaction model formulations which take advantage of disaggregation but he readily admits that the data problems associated with calibrating these models are immense. To a large extent lack of suitable data has held back research in this area. Traditionally, one of the advantages of the aggregate modelling approach has been the ease with which data requirements for aggregate models could be met by conventional published sources (government agencies, census data). However, disaggregated models usually demand data at a finer level of detail than are available from published sources. This usually necessitates collecting data directly via sample survey. High quality data may be obtained in this way although the cost of doing so is great since a data collection framework must be designed when the appropriate level of aggregation is governed by the research objectives and project resources constraints. The data used in the study have been obtained directly from a survey conducted in the Hamilton area in 1978. These data include information on multi-purpose, multi-stop trips and are of sufficient detail to meet the specifications of the model presented in the following chapter.

In contrast to operational difficulties incurred by disaggregation, conceptual problems may complicate matters. Webber (1980, 130) admonishes researchers to identify more rigorously the first order objects of study. One of the difficulties of modelling residential location and trip making behaviour within the framework of a single model is the distinction between households and individuals. The complexities of household structure have been simplified in the model developed in Chapter Three, for example, by assuming one worker per household. The identification of the behaving objects is a necessary though often implicit prerequisite in constructing theory (and for that matter, models of the theory).

A closely related issue concerns the level of aggregation or disaggregation which is implicit within a model and is a function of the scale of analysis. Unlike some disciplines which are defined by scale (e.g. microbiology, biology, ecology) the discipline of geography operates over a wide variety of scales. Reductionism may be a useful driving force in research but it does not follow that all useful results can be obtained at the finest scale (Wilson, 1979). Therefore we must address one final question: "Where do we draw the line on disaggregation ?". With respect to the model presented in this thesis the focus is restricted to the intra-urban scale and the line has been drawn at the point where we distinguish between residential location and trip making behaviour by purpose for different income groups. In so doing we recognize the existence of subsystems (defined by income group and trip purpose) and also the interrelatedness of these subsystems (residential location takes into account observed work trip and shop trip behaviour).

In order to simplify the analysis of a complex system such as a city, it is useful to define subsystems as in Figure 2.1. The model develped in this study interelates shopping, residential location and journey to work components within an interacting systems framework. These are the three main types of spatial interaction model as defined by Cordey-Hayes and Wilson (1971). As illustrated in Figure 2.1, the common link between these model types is transportation. The fundamental importance of the use of travel characteristics to distinguish between these basic types of spatial interaction will become clear in the analysis that follows.

Figure 2.1 Intra-urban models



2.4 Hathaway's Approach to Disaggregation

Hathaway (1975) has disaggregated a doubly constrained trip distribution model in order to investigate the gain in descriptive accuracy which results from taking account of differences shown to exist between various categories of trip maker. Four classifications of trip makers were considered: age, sex and marital status, socio-economic group, occupational classification and standard industrial classification. A trip distribution model was calibrated for each category of trip maker and an "overall" distribution model was calibrated which treated the population as a homogeneous group.

The "overall", homogeneous trip distribution model was of the form:

$$t_{ij} = A_i O_i B_j D_j \exp(-\lambda c_{ij}), \qquad (2.9)$$

$$\text{subject to} \quad \sum_i t_{ij} = D_j, \; i = 1,N, \qquad (2.10)$$

$$\sum_j t_{ij} = O_i, \; j = 1,N \qquad (2.11)$$

$$\text{and} \quad \sum_i \sum_j t_{ij} c_{ij} = \bar{c}, \qquad (2.12)$$

Where t_{ij} are the number of trips from origin zone

i to destination zone j, O_i are the total number of trips from origin zone i, D_j are the total number of trips to destination zone j, N is the total number of traffic zones within the study area, c_{ij} is the cost (time) of travel from zone i to zone j, $\bar{c}$ is the total expenditure on travelling and A_i and B_j are the balancing factors of the familiar form:

$$A_i = [\sum_j B_j D_j \exp(-\lambda\ c_{ij})]^{-1} \tag{2.13}$$

$$\text{and} \quad B_j = [\sum_i A_i O_i \exp(-\lambda\ c_{ij})]^{-1} \tag{2.14}$$

The disaggregated model was of the form:

$$t^k_{ij} = A^k_i\ O^k_i\ B^k_j\ D^k_j \exp(-\lambda^k\ c_{ij}),\ k=1,\ K \tag{2.15}$$

where the superscript k refers to category k. A^k_j and B^k_j are found from

$$\sum_j t^k_{ij} = O^k_i,\ i = 1,N \text{ and } k = 1,K \tag{2.16}$$

$$\text{and} \quad \sum_i t^k_{ij} = D^k_j,\ j = 1,N \text{ and } k=1,K \tag{2.17}$$

The investigation followed a number of stages. The first step calculated average trip lengths for each group of trip makers. Significant differences were shown to exist in the average lengths between different sections of the population. The second step calibrated and ran the 'overall' distribution model and the disaggregate model to see how well each reproduced the survey data. The survey data were obtained for a sub-area of the London Transport Survey area from a 1966 10% sample census. Travel times were computed by the TRANSITNET program for all observed work trips. Full details of this are given in Hathaway (1975). Using the disaggregate model, the inhomogeneity of travel behaviour between different categories of trip makers is represented firstly by differences in the decay constant in equation (2.15), and secondly by the different spatial distribution of trip ends. Thus, the possibility exists that an origin-destination pair of zones might have attractions or generations nearly all of which correspond to persons of type k. Even though the two zones may be far apart, one might expect a considerable interchange of persons of type k between them. Consequently, the trip interchange between these two zones may be even more under-estimated by a conventional, homogeneous trip distribution model.

Hathaway's analysis revealed the following. The first step of the analysis reveals that significant differences existed in the mean trip length between categories, most notably that trip makers in "higher" socioeconomic groups had on average higher work trip journey times. In order to determine the significance of small differences in mean trip length a t-test was employed to establish confidence intervals. Most of the observed differeces between categories were found to be significant. Next, both the 'overall' and disaggregated trip distribution models were calibrated using the method of maximum likelihood. The difference in decay constants between categories was only marginal and did not appreciably aid in the interpretation of the differences in travel behaviour between groups. The question of the significance of small differences in the decay constant between categories was addressed and the method of relative likelihood was used to estimate the probabilities that categories of trip makers had the same 'real' underlying decay constants. In most cases the difference in decay constants between categories was found to be significant. As might be expected from the form of the deterrence function, $\exp(-\lambda)$, the decay constants displayed a strong negative correlation with average trip length. In general, as the age of the trip maker increased the decay constant demonstrated a gradual increase whilst the average trip length decreased. Also, broadly speaking, males demonstrated a lower decay constant value (greater propensity to travel further for work trips) then did females.

Several goodness of fit measures were used to test the match between observed and predicted matrices; the chi-squared statistic, root mean square error and mean % error. The study revealed serious drawbacks with the use of the chi-squared statistic. Due to the sparsity of some synthesized and observed matrices, attempts were made to aggregate zones until most of the elements exceeded five. These aggregated matrices were then compared and the same measures of comparison were calculated. Aggregating the zones into larger spatial groups did little to improve the overall results. Finally each set of synthesized category matrices were summed and the resultant trip matrix was compared to the total trip matrix.

The 'overall' trip distribution model given by equation (2.9) reproduced the observed trip interchange with an average error of 42.5%. The fit obtained by calibrating, synthesizing and summing separate matrices was only marginally better. For example, the synthesized socioeconomic group matrices summed to give a matrix which differed from the observed matrix by an average of 38.9%. Hathaway presented four major conclusions as a result of this study:

(i) Trip length characteristics do differ significantly from trip making category to trip making category;

(ii) A simple trip distribution model will not reproduce an observed trip interchange situations at all well unless considerably more detail is introduced to the model;

(iii) Different sections of the trip-making population possess significantly different deterrence function parameters; and,

(iv) The calibration of separate trip-distribution models leads only to a marginal increase in the ability of that model to reproduce the observed situation

As a final note Hathaway (1975, 87) added "... the error involved in the assumption, in the trip distribution model, of the homogeneity of trip making behaviour is small in comparison with the error involved in the form of the model itself."

The model developed in this paper will pursue the topic of disaggregation in more detail than the trip distribution model used in Hathaway's study. The comprehensive model used here for analysis seeks to extend the efforts of Hathaway, principally by following through on the second of Hathaway's conclusions and incorporating more detail into the model structure. It is hoped that the form of this model will contain less error than the model used by Hathaway, thereby accruing greater benefit from disaggregation.

2.5 Summary

This chapter has established the operational modelling context and has examined in some detail two topics which impact directly upon the model to be presented in the next chapter. A consideration of entropy and information theory is essential for an understanding of the information theoretical principles which constitute the mathematical bases of the model. The discussion of aggregation issues is especially important with respect to the objectives of the analysis which follows. It should serve as a preliminary introduction to the design features of the disaggregated model. The brief review of Hathaway's (1975) study and results should shed some light upon the nature of the disaggregation problem and will provide a comparative basis for the analysis of Chapter Five.

A conspicuous lack of disaggregation has been shown to exist in recent urban modelling efforts, in particular for theoretically-based, operational models. The problem stems from the inherent difficulty of translating the theory into an empirically testable form of model. Although disaggregated urban economic location models are logically complex and more intuitively acceptable, the results of this testing have been disappointing (Batty, 1978; Senior, 1977; Hathaway, 1975). Recent results of empirical testing have revealed that a careful evaluation of the benefits and disadvantages accruing to disaggregation must be made.

Aggregate macromodels stand a better chance of statistical validation, however, in terms of explanation, disaggregated models are to be preferred. From the point of view of data requirements, macromodels are more feasible, but as Batty (1978, 75) points out, the planning process may require models which are predictively useful at a disaggregated scale. The model outlined in the next chapter incorporates disaggregation in an attempt to operationalize more realistic and relevant components of urban spatial theory.

CHAPTER THREE

A DESCRIPTION OF THE MODEL AND ASSOCIATED DEVELOPMENTS

3.1 Introduction

This chapter discusses the model developed in this study, its components and some issues which are directly related to these components. For clarity the aggregate version of the model (treating the sample data as homogeneous) will be referred to as CALBTN2.1. The income group disaggregated version will be called CALBTN2.2.

The chapter is organised as follows. First, a brief review of the important developments in previous model versions will be presented in order to place CALBTN2.1 and CALBTN2.2 within the proper context. The following sections will then outline the methods of categorisation, the form of the model and the derivation of the model. During the development of this model considerable effort has been devoted to improve the operational efficiency of the solution procedure. The calibration method used in this model is a hybrid technique based upon Powell's algorithm (Powell, 1970). This procedure will be briefly summarised; also a method will be presented which separates the solution procedure into work and shop trip subcomponents thereby substantially reducing the cost of using the model.

3.2 A Brief Review of Previous Model Versions: Problems and Developments

This section will briefly outline some of the major developments in previous model versions in order to convey how this on-going modelling effort has progressed to the present stage. The discussion will focus upon interesting qualitative conclusions that have been drawn from the experience rather than providing an elaborate account of former models and their results. For complete details refer to Webber et. al. (1978, 1979a, 1979b) and Okelly (1978).

Previous model versions were information-minimising spatial interaction models, as are CALBTN2.1 and CALBTN2.2; however, they were long run models which predicted shopping facility location as well as modelling the location of residents journey to work and shop trip working behaviour. These models experimented with different constraint specifications and prior formulations. In each case the objective was to test model realism; to show that the assignment produced by the model compared with observed patterns. A series of revisions took place, each learning from the errors of previous models. The nature of these revisions and the insights gained are now presented.

The earliest model version employed a generalized trip cost constraint which did not distinguish between work trips and shop trips. This proved to be unsatisfactroy. As a result, separate work trip and shop trip constraints were incorporated to more realistically represent the differences underlying these two types of urban trip making behaviour.

Different residential land use prior probability distributions were attempted. These included priors based upon zone size and various combinations of developable land. The land use accounting machanism presently employed in CALBTN2.1 and CALBTN2.2 was found to produce the best results. A prior distribution based upon the proximity of shopping centres to one another was used to test ideas about agglomeration economics in the retail sector (one of the problems ignored by the Lowry model). It used a negative exponential function of the average distance between centres. This modification did not significantly improve model predictions.

A problem which became apparent in previous versions was how to deal with households in the "no shop trip" and "no work trip" categories. These included households in the unemployed or retired category and those that were observed to make no shop trips during the period of the survey. Model results showed that it was too easy to assign individuals to these categories since zero costs associated with these categories allowed the model to spread individuals out and still meet the average cost constraints. Therefore to prevent overpredictions in the "no work trip" and "no shop trip" categories it was necessary to add a constraint on the average number of shop trips per week per household and to include an unemployed/retired category for work trips.

A sensitivity analysis conducted on one of these prior model versions revealed some interesting results (Webber, et. al., 1979a). This previous long-term model version was run with different values of the mean cost estimates in order to determine the sensitivity of the model to these parameters (parameters in this case referring to the mean cost estimates). As the mean cost of work trips (time, in minutes) was increased, the observed/predicted work trip matrix corelations (measured in terms of R^2) increased and the proportion of the population living downtown decreased. The effect of varying the three parameters (mean cost of work trip, mean cost of shopping trip pattern and mean number of shop trips per week per household) was in the expected direction insofar as work trip distribution, shop trip pattern and population distribution were concerned; however, in terms of correlations between observed and predicted matrices, there was little improvement gained by varying the mean cost of work trips or the mean cost of shopping trip pattern. The results indicated the insensitivity of model predictions to parameter variations; variations of as much as twenty percent yielded only minor changes in observed/predicted correlations and predicted probability distributions. This cast some doubt upon the accuracy of the observed values of the parameters. It was concluded that the coarser aspects of the city's spatial structure are quite insensitive to parameter changes.

This review has outlined the major results of previous modelling efforts. It should provide a better understanding of the rationale behind the improvements incorporated into the present model versions and should also indicate the heuristic approach taken in the development of these operational models. Unfortunately all of the previous model versions have relied upon rather poor data to identify the constraints upon aggregate behaviour and have contained a small number of categories to which people could be assigned. By allowing a distribution of only 2 or 5 shopping facilities it was found that there was a high degree of locational indeterminacy for facilities, and model solutions often resulted in which most facilities were allocated to one of the central zones. Also, the assignment of shopping trips to shopping centres was quite lumpy; consequently the predicted shop trip patterns were crude as compared to observed patterns. The increase to 20 shopping facilities in CALBTN2.1 and CALBTN2.2 is certainly more representative of the observed distributions. It is hoped that by expanding the study area from 14 to 19 zones and by providing 20 shop centres a more realistic assignment of individuals to (a larger number of) categories will be achieved. Also, the new survey data are a significant improvement over the data used in previous model versions. To a large extent these data should help to overcome the inadequacies mentioned above.

The results of the sensitivity analysis have raised some questions regarding the accuracy of the observed mean cost estimates for work trips and shop trip patterns and the mean number of shop trips per week per household. The income group disaggregation used in CALBTN2.2 should provide a more realistic and accurate basis for the mean cost estimates by segregating the heterogeneous sample population into homogeneous income groups.

3.3 Categories

The model assigns individuals to place of residence, place of work and shopping trip pattern. The details of these categories are now given.

(1) Place of residence (i = 1,...,17)

The 17 residential categories correspond to zones 1 to 17 shown in Figure 3.1. These zones are based upon neighborhood and planning districts used by the City of Hamilton Planning Department.

(2) Place of work (k = 1,...,20)

Work place categories 1 to 19 correspond to zones 1 to 19 shown in Figure 3.1. The last work place category (20) is reserved for households observed to make no work trip because of unemployment or retirement. This category has a pronounced effect upon model solutions, especially for CALBTN2.2. It will be shown that significant differences exist in the proportion of households within this category for different income groups.

(3) Shopping trip pattern (j = 1,...,440)

These patterns are generated by allowing households to make zero, one, two or greater than two trips per week to a set of shopping facilities. Households are classified as follows:

(i) 0,1 shop trips per week j = 1,...,20

(ii) 2 shop trips per week $j = 21,\ldots,230$

(iii) >2 shop trips per week $j = 231,\ldots,440$.

The combinatorial expression for shop trip categories (i) and (ii) is handled in the following manner. Suppose there are x facilities. We can choose:

1 from x in x ways,

2 from x (both different) in $\frac{x(x-1)}{2}$ ways,

2 from x (both same) in x ways.

The total number of ways is then

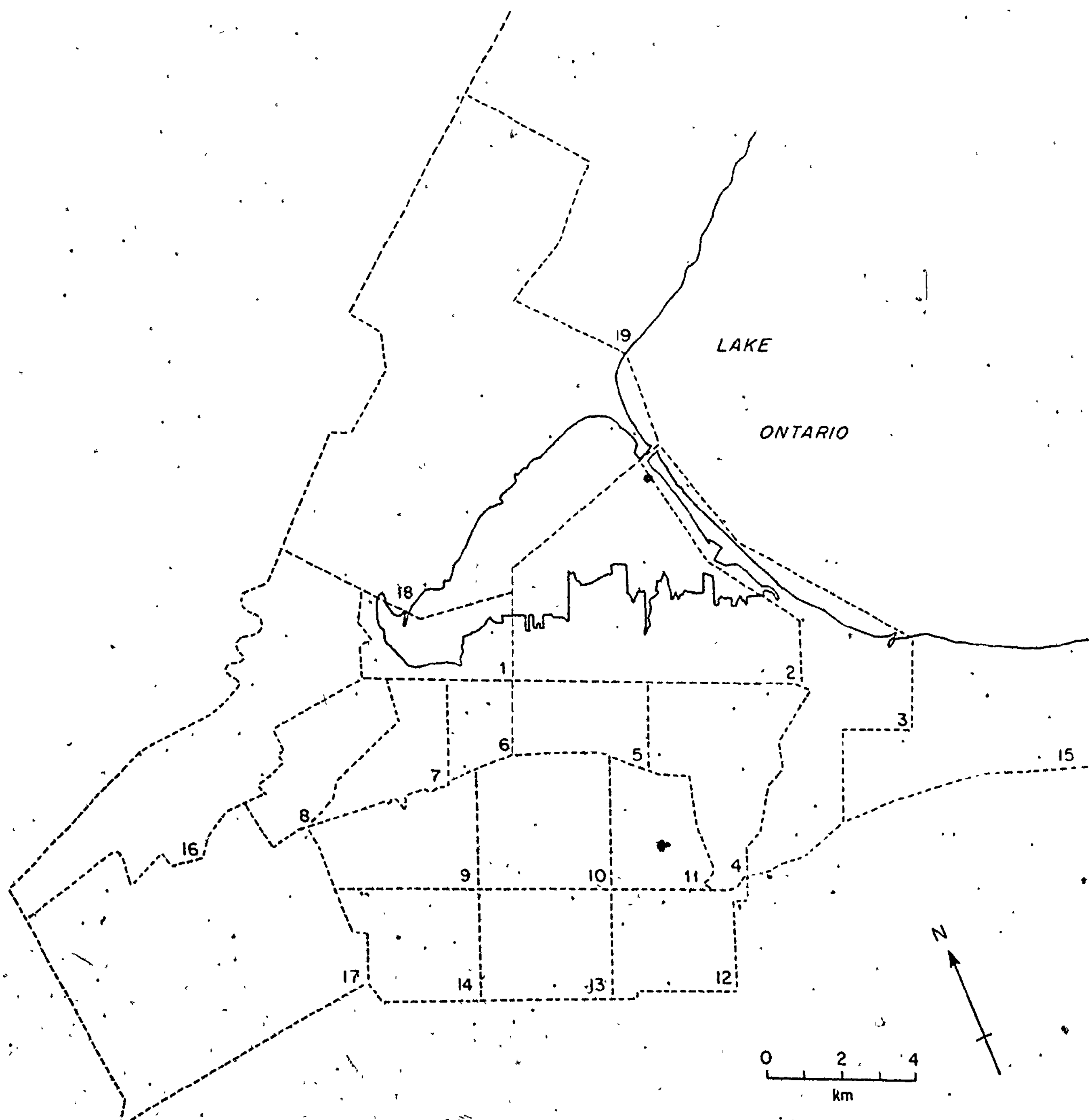
$$\frac{2x}{2} + \frac{x(x-1)}{2} + \frac{2x}{2}$$

$$= \frac{x^2 - x + 4x}{2}$$

$$= \frac{x(x+3)}{2} .$$

In this model there are 20 shopping facilities (x=20), so the number of trip types for the first two shop trip categories is 230. Category (iii) has been conceptualized as 2 shop trips per week, only at a higher frequency than category (ii). Therefore, the third category introduces an additional 210 trip types which brings the total to 440. This classification scheme provides a realistic link between the observed distribution of shopping facilities and household shopping behaviour by the calculation of costs associated with each pattern.

Figure 3.1 The Study Region Zonal Network



3.4 CALBTN2.1: The Aggregate Model

The model works in the following manner. For any given spatial distribution of retail facilities and fixed work places the model calculates the probability that an individual household chooses a particular location, workplace and shopping trip pattern subject to expectations based upon survey data about the average cost of shopping and work trips and the average number of shopping trips per week per household. This probability is information minimising compared to a prior probability distribution which is chosen to reflect residential land availability.

Let $\underline{\underline{p}} = (p_{ijk})$ where p_{ijk} is the joint probability of a household residing in zone i, exhibiting shopping pattern j and the head of household working in zone k. Define:

17 zones of residence in Hamilton $(i = 1,\ldots,17)$

440 shop trip patterns $(j = 1,\ldots,440)$

20 work trip destination zones (17 zones in Hamilton, 2 in Burlington and 1 unemployed category) $(k = 1,\ldots,20)$

Therefore, $\underline{\underline{p}}$ has 149,400 elements.

Model 2.1 is a singly-constrained information-minimising model of the form:

Choose $\underline{\underline{p}}^*$ (the asterisk denotes the solution) to minimise

$$I_k = \sum_{ijk} p_{ijk} \ln \frac{p_{ijk}}{\tilde{r}_i \; s_j \; w_k} \qquad (3.1)$$

subject to

$$\sum_{ij} p_{ijk} = d_k, \; \forall_k \qquad (3.2)$$

$$\sum_{ijk} p_{ijk} \tilde{c}_{ij} = \bar{c}_s \qquad (3.3)$$

$$\sum_{ijk} p_{ijk} c_{ik} = \bar{c}_w \qquad (3.4)$$

$$\sum_{ijk} p_{ijk} x_j = \bar{x} \qquad (3.5)$$

where the above variables are defined as follows:

d_k = observed work trips ending in zone k,

$\tilde{c}_{ij}$ = cost (time, in minutes) of shop trip pattern j to a resident in i,

$\bar{c}_s$ = the observed mean cost (time, in minutes) of shopping trips per week per household,

c_{ik} = cost (time, in minutes) of a work trip from zone i to k,

$\bar{c}_w$ = the observed mean cost (time, in minutes) of a work trip,

x_j = the number of shop trips (per week, per household) associated with trip pattern j.

$\bar{x}$ = the observed mean number of shop trips (per week, per household),

and the prior probabilities are,

$\tilde{r}_i$ = residential prior; the proportion of (Hamilton's residential + vacant + retail land in this zone) - (the amount of retail space consumed by shopping facilities allocated to this zone),

s_j = shopping pattern prior (uniform),

w_k = work trip prior (uniform).

The residential prior, $\tilde{r}_i$, is an accounting mechanism which compensates for the fact that retail facilities consume space. The amount of land available for residential development in each zone is reduced according to the number of shopping facilities allocated to that zone. There are twenty facilities of equal size which, in total occupy 100% of Hamilton's retail land space. The details of the procedure are as follows:

$$\tilde{r}_i = \begin{cases} \dfrac{L_i}{R_s+V+R_t} & \text{if } i \text{ has no centre} \\[2ex] \dfrac{L_i}{R_s+V+R_t} - \dfrac{x}{N}\left[\dfrac{R_t}{T}\right] & \text{if } i \text{ has } x \text{ centres} \end{cases}$$

where $\tilde{r}_i$ = residential prior

L_i = the sum of residential + vacant + retail land use in i

R_s = total residential land in Hamilton

V = total vacant land in Hamilton

R_t = total retail land in Hamilton

T = total land in Hamilton

N = number of facilities (20)

The retail priors $\tilde{r}_i$ are then normalised so that $\sum_i \tilde{r}_i = 1$. This prior incorporates land use competition within the model by reducing the prior probability for residential land use in a zone according to the amount of land consumed by the facilities present in that zone.

The shopping pattern prior and work trip prior are uniform and introduce no prior information regarding shop trips or work trips into the model. They are computed as,

$$S_j = \frac{1}{440}, \forall_j, \quad \text{and}$$

$$w_k = \frac{1}{20}, \forall_k.$$

The constraints are built into the model by maximising the following Lagrangean form:

$$L^* = \sum_{ijk} p_{ijk} \ln \frac{p_{ijk}}{\tilde{r}\; s_j\; w_k} - \sum_k \lambda_1^k \left(d_k - \sum_{ij} p_{ijk}\right) - \lambda_2\left(\bar{c}_s - \sum_{ijk} p_{ijk}\tilde{c}_{ij}\right)$$

$$- \lambda_3\left(\bar{c}_w - \sum_{ijk} p_{ijk} c_{ik}\right) - \lambda_4\left(\bar{x} - \sum_{ijk} p_{ijk} x_j\right)$$

where λ_1^k $(k = 1,\ldots,20)$, $\lambda_2, \lambda_3, \lambda_4$ are the multipliers.

To find the minimum of L^* it is necessary to solve the above equation by differentiating and setting the partial derivatives equal to zero.

Hence,

$$\frac{\partial L^*}{\partial p_{ijk}} = p_{ijk} \frac{\partial}{\partial p_{ijk}} \ln \frac{p_{ijk}}{\tilde{r}_i\; s_j\; w_k} + \ln \frac{p_{ijk}}{\tilde{r}_i\; s_j\; w_k} + \sum_k \lambda_1^k + \lambda_2\, \tilde{c}_{ij} + \lambda_3 c_{ik}$$

$$+ \lambda_4\, x_j = 1 + \ln \frac{p_{ijk}}{\tilde{r}_i\; s_j\; w_k} + \sum_k \lambda_1^k + \lambda_2\, \tilde{c}_{ij} + \lambda_3\, c_{ik} + \lambda_4\, x_j$$

$$= 0.$$

Rearranging yields

$$\ln \frac{p_{ijk}}{\tilde{r}_i \, s_j \, w_k} = -1 - \lambda_1^k - \lambda_2 \, \tilde{c}_{ij} - \lambda_3 \, c_{ik} - \lambda_4 \, x_j .$$

Therefore, by absorbing -1 into λ_1^k we get

$$p_{ijk} = \tilde{r}_i \, s_j \, w_k \exp(-\lambda_1^k - \lambda_2 \, \tilde{c}_{ij} - \lambda_3 \, c_{ik} - \lambda_4 \, x_j). \tag{3.6}$$

Using (3.2),

$$\sum_{ij} \tilde{r}_i \, s_j \, w_k \exp(-\lambda_1^k - \lambda_2 \tilde{c}_{ij} - \lambda_3 \, c_{ik} - \lambda_4 \, x_j) = d_k$$

and

$$\exp(-\lambda_1^k) = \frac{d_k}{\sum_{ij} \tilde{r}_i \, s_j \, w_k \exp(-\lambda_2 \, \tilde{c}_{ij} - \lambda_3 \, c_{ik} - \lambda_4 \, x_j)} \tag{3.7}$$

Substituting (3.7) into (3.6) and rearranging gives

$$p_{ijk} = \tilde{r}_i \, s_j \, w_k \, b_k \, d_k \exp(-\lambda_2 \, \tilde{c}_{ij} - \lambda_3 \, c_{ik} - \lambda_4 \, x_j) \tag{3.8}$$

where

$$b_k = \frac{1}{\sum_{ij} \tilde{r}_i \, s_j \, w_k \exp(-\lambda_2 \, \tilde{c}_{ij} - \lambda_3 \, c_{ik} - \lambda_4 \, x_j)}$$

The best estimates of the model's parameters are found by solving (3.8), and the solution chooses λ_1^k $(k = 1, \ldots, 20)$, λ_2, λ_3, λ_4 such that (3.8) and (3.2),...,(3.5) are satisfied.

3.5 CALBTN2.2: The Income Group Disaggregated Model

The form of CALBTN2.2 is essentially that of CALBTN2.1 but now $\underline{\underline{p}}^v = (p_{ijk})$ where p^v_{ijk} is the joint probability of a household in income group v residing in zone i, exhibiting shopping pattern j and the head of household working in zone k. The definitions of zone of residence i, shop trip pattern j and work trip destination zones k are as before. In addition, define four income categories ($v = 1, \ldots, 4$) as follows.

	Gross Household Income Per Annum
v = 1	\$ 0 - 10,000
v = 2	\$10,001 - 20,000
v = 3	\$20,001 - 30,000
v = 4	exceeds \$30,000

CALBTN2.2 is therefore an attraction-constrained, information-minimising model of the form:

Choose $\underline{\underline{p}}^{v*}$ to minimise

$$I^v_k = \sum_{ijk} p^v_{ijk} \ \ln \frac{p^v_{ijk}}{\tilde{r}_i \ s_j \ w_k} \qquad (3.9)$$

subject to

$$\sum_{ij} p^v_{ijk} = d^v_k, \forall_k \qquad (3.10)$$

$$\sum_{ijk} p^v_{ijk} \tilde{c}_{ij} = \bar{c}^v_s \qquad (3.11)$$

$$\sum_{ijk} p^v_{ijk} \dot{c}_{ik} = \bar{c}^v_w \qquad (3.12)$$

$$\sum_{ijk} p^v_{ijk} x^v_j = \bar{x}^v \qquad (3.13)$$

where the above variables are defined as before except for the following:

d^v_k = observed work trips for income group v ending in zone k,

$\bar{c}^v_s$ = the observed mean cost (time, in minutes) of a shopping trip (per week, per household) in income group v,

$\bar{c}^v_w$ = the observed mean cost (time, in minutes) of a work trip for income group v,

x^v_j = the number of trips (per week, per household) associated with trip patterns j for income group v,

$\bar{x}^v$ = the observed mean number of shop trips (per week, per household) for income group v.

Note that the work trip and shop trip cost matrices, c_{ik} and $\tilde{c}_{ij}$, are the same for both aggregate and disaggregate model versions. The Lagrangean form of CALBTN2.2 and its solution are virtually identical to the Lagrangean form of CALBTN2.1 and its solution except that the above income group specific variables replace their counterparts within the aggregate model formulation.

3.6 Calibration

Calibration, or parameter estimation, is an integral part of the development of spatial interaction models. It is a fundamental stage of model design in which the modeller develops a basic understanding of the model by exploring the sensitivity of the model variables and the model structure. It is a useful phase in developing operational models since it provides tests for evaluating the relevance and limitations of the model with respect to the research or planning application objectives.

Previous model versions employed the classical Newton-Raphson iterative method for calibration (i.e., finding the 'best' values of the models parameters). Although the method proved to be robust, it did present difficulties of slow convergence and excessive iterations (O'kelly, 1978). This problem has been encountered in other operational modelling studies (e.g. Openshaw, 1976, 28; Batty and Mackie, 1972) and is an inherent difficulty of the Newton-Raphson procedure. The problem relates to the stability of the Newton-Raphson method. When provided with good starting values the method is rapidly (quadratically) convergent; however, when the starting values are not good initial estimates (i.e. not in the vicinity of the root) the iterative process may diverge due to error propagation.

This was the problem experienced by the predecessors of the current model, since the choice of parameter starting values was often arbitrary (due to a lack of knowledge of the parameter space). Despite this fact the Newton-Raphson method did suffice for calibration in previous model versions. The experience gained, however, indicated the need for more efficient calibration methods, especially for larger, more complex models.

This need was reaffirmed during the development of CALBTN2.1 and CALBTN2.2 when considerable difficulties were encountered with respect to data requirements, calibration and excessive use of computer resources. To a large extent these technical difficulties have resulted from the comprehensive nature of the model (i.e. the large number of categories to which individuals are assigned) and have been compounded by the income group disaggregation. Due to the increased dimensions in the current model versions (the solution matrices $\underline{\underline{P}}^{*}$ and $\underline{\underline{P}}^{V*}$ have 149,400 elements), the practical limitations of computer storage and execution time were difficulties that had to be faced.

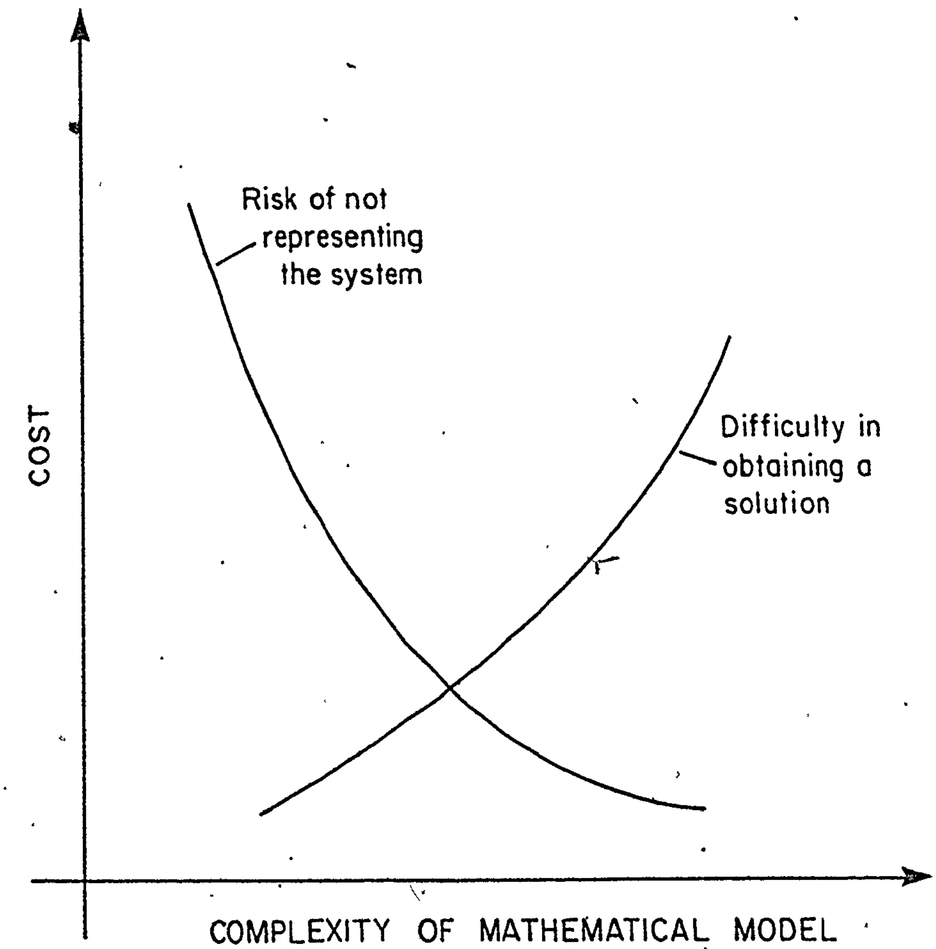
The decision to develop a comprehensive model in this study impacts directly upon the problem of model calibration. Figure 3.2 illustrates the consequences of the choice of model complexity. If a simplified model is used the risk of not representing the system of interest will be maximized but the difficulty in obtaining a solution will be minimized. On the other hand, a highly complex model will reduce the of not representing the system but will maximize the difficulty of obtaining a solution. For the purposes of this study it was decided that a relatively complex model should be used to provide optimum realism. It was realized, however, that data requirements, computer time and programming effort would be large.

Basically two approaches to solving these technical problems were chosen; (1) it was decided to use a calibration algorithm based upon Powell's method, and (2) the model was reformulated into two submodels for the purpose of calibration.

(1) Powell's algorithm

This algorithm employs a hybrid method for numerically solving a system of non-linear equations. The method is outlined and justified in Powell (1970). Batty and Mackie (1972) have made a case for hybrid calibration techniques and argue that they can be more effective with respect to convergence and stability. For these reasons hybrid combinations are to be preferred because they are based upon the demonstrated capabilities of each method at different convergence limits. The following outline of Powell's algorithm ignores some details of the program in order to provide a simplified picture of the method so that the reader may relate the following description to the general strategy. Complete details are found in Powell (1970). A rudimentary understanding of the convergence characteristics of the Newton-Raphson method is

Figure 3.2 The Trade-Off Diagram



assumed. Appendix 1 provides a basic outline of the Newton-Raphson iterative scheme.

A number of algorithms have been developed (usually based upon the Newton-Raphson method) which try to provide reliable convergence when close initial estimates are not available. A common strategy retains the direction but restricts the length of step predicted by classical methods in the following manner:

$$\underline{\underline{x}}^{k+1} = \underline{\underline{x}}^{k} + \delta^{k} \tag{3.14}$$

becomes

$$\underline{\underline{x}}^{k+1} = \underline{\underline{x}}^{k} + \psi^{k} \delta^{k} \tag{3.15}$$

where $\underline{\underline{x}} = (\lambda_2, \lambda_3, \lambda_4)$ in equation (3.8)
and ψ^k prevents the estimate $\underline{\underline{x}}^{k+1}$ from being worse than $\underline{\underline{x}}^{k}$. In practice the sum of squares of residuals is calculated as

$$F(\underline{\underline{x}}) = \sum_i [f_i(\underline{\underline{x}})]^2$$

where f_i is equation (3.8)
in order to calculate ψ^k to achieve the improvement

$$F(\underline{\underline{x}}^{k+1}) < F(\underline{\underline{x}}^{k}). \tag{3.16}$$

The hybrid method proposed by Powell uses successive values of $f_i(\underline{\underline{x}}^k)$ to build up a numerical approximation to the Jacobian matrix by the method of (3.15). In this way revisions to the Jacobian (J) require far fewer computational evaluations on each iteration. This procedure for the economisation of the total amount of computation is consistent with Broyden's scheme (1965).

Powell's algorithm solves the system of non-linear equations

$$f_i(x_1, \ldots, x_n) = 0. \qquad (3.17)$$

The algorithm works when other existing methods fail, in particular Newton iterative methods which fail if the Jacobian becomes singular (this is the main problem). Let $\underline{\underline{x}} = (x_1, \ldots, x_n)$ be the vector of unknowns (in this case the Lagrangean multipliers $\lambda_2, \lambda_3, \lambda_4$ in (3.8). The classical Newton approximation to $\underline{\underline{x}}$ of the solution of system (3.17) calculates the Jacobian

$$J_{ij} = \frac{\partial}{\partial x_j} f_i(x).$$

Next obtain a correction vector δ by solving the system of linear equations

$$\sum_j J_{ij} \delta_j = - f_i(\underline{\underline{x}}), \ \forall_i \qquad (3.18)$$

and replace vector $\underline{\underline{x}}$ by vector $(x \pm \underline{\underline{\delta}})$.

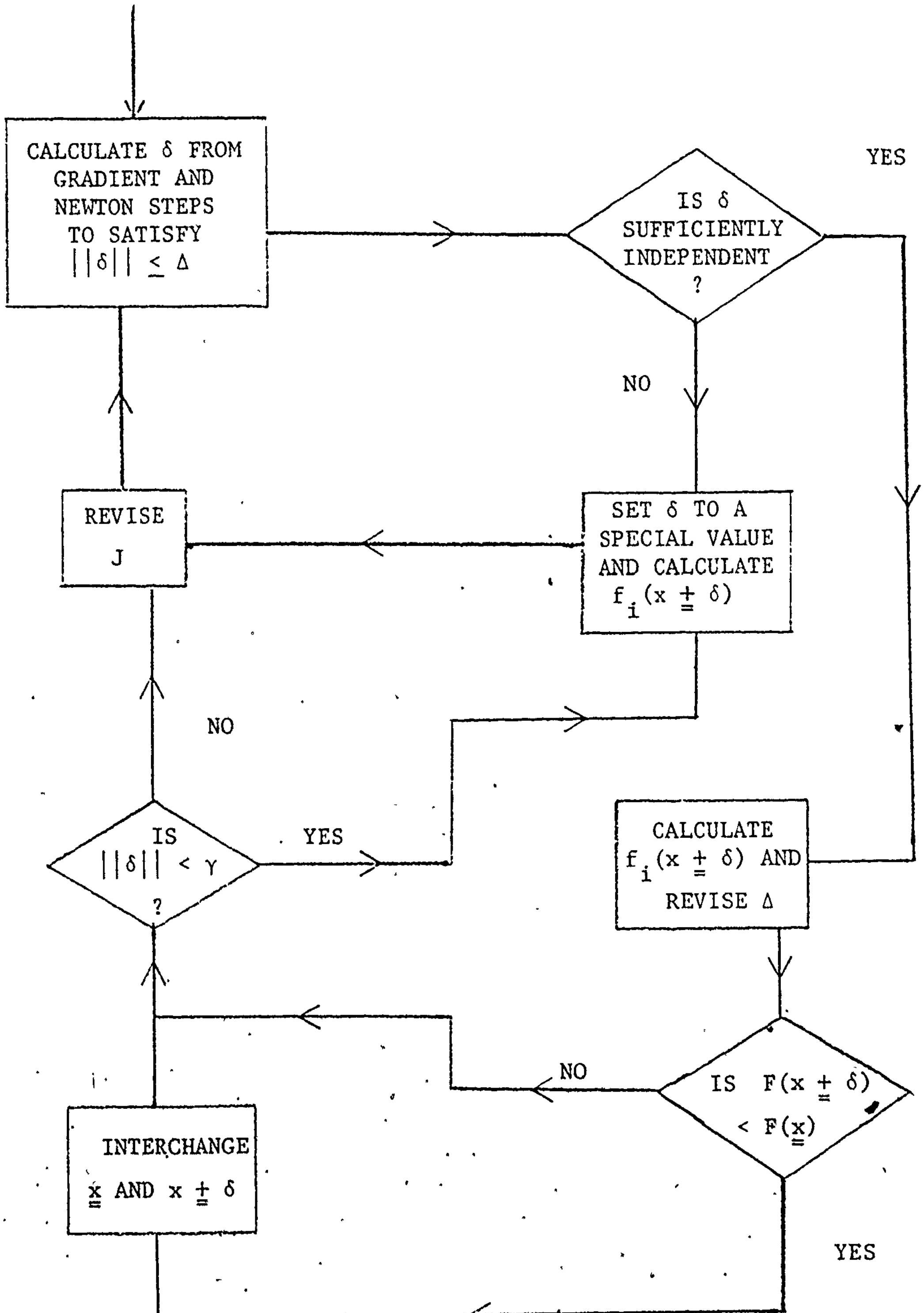
Powell's algorithm retains the fast ("second-order" or quadratic) convergence of the Newton-Raphson method but is modified so that the iteration is progressive even if the guess $\underline{\underline{x}}$ is far from the solution. This is done by using the correction vector δ calculated in (3.18) as a search direction, i.e. subsequent steps are taken along the steepest descent direction of $F(\underline{\underline{x}})$ if the Newton-Raphson iteration diverges. The correction vector δ incorporates both Newton-Raphson iteration and steepest descent characteristics. This formulation of the correction vector has its good and bad points. The steepest descent qualities yield some appealing convergence theorems (see Powell, 1970, 122); however the method may converge to a stationary point of $F(\underline{\underline{x}})$ even if it is not a solution to the equations. This will come about if the gradient of $F(\underline{\underline{x}})$ becomes very small. Thus, the method cannot overcome the familiar failing, i.e. the inability to recognize whether or not a solution is a global minimum of $F(\underline{\underline{x}})$. The usual solution to this problem is to use different starting estimates of $\underline{\underline{x}}$. Figure 3.3 illustrates a typical iteration of Powell's algorithm. The following paragraphs summarize the steps involved.

Input data include $\underline{\underline{x}}$ the vector of Lagrangean multipliers; an approximation to the Jacobian J; its inverse J^{-1}; a matrix of directions in the parameter space Ω and an associated vector of integers ω (which provide conditions for finishing the iterative process); and step length Δ.

First calculate a correction δ to apply to $\underline{\underline{x}}$. As previously mentioned, the correction is a compromise between Newton-Raphson and the method of steepest descent applied to the sum of squares $F(\underline{\underline{x}})$. The step length Δ, which can be changed on each iteration, controls the balance between the two methods. The purpose of Δ is to restrict the length of displacement $(\underline{\underline{x}}^{k+1} - \underline{\underline{x}}^{k})$ in order that $F(\underline{\underline{x}})$ be decreased. If Δ is larger, the correction δ is a pure Newton step (fast convergence); if Δ is small the vector δ is a multiple of the predicted gradient of $F(\underline{\underline{x}})$. In all cases δ satisfies the sum of squares criterion (3.16). This estimate may at first be incorrect because of the non-linear behaviour of the functions $f_i(\underline{\underline{x}})$, i.e. $f_i(\underline{\underline{x}})$ does not vary linearly between $\underline{\underline{x}}$ and $(x \underline{\underline{\pm}} \delta)$. The criterion which maintains linear independence of δ depends on the elements of ω and Ω and will not be elaborated upon here. For details, see Powell (1970, 133).

Usually, δ passes the test of linear independence in which case $f_i(x \underline{\underline{\pm}} \delta)$ is calculated and Δ is updated. The step length Δ is revised, if need be, such that it is small enough to satisfy the inquality (3.16) yet Δ must be greater than γ (the minimum step-length to approximate the first derivatives). If $F(\underline{\underline{x}})$ can be decreased substantially, then in order to prevent an extreme number of iterations (due to Δ being too small) the program may increase Δ subject to the

Figure 3.3   Summary of an Iteration



following criteria which checks the accuracy of the approximations.

$$f_i(\underline{\underline{x}} + \delta) - f_i(\underline{\underline{x}}) \simeq \sum_j J_{ij} \delta_j, \forall_i.$$

If inequality (3.16) holds then $\underline{\underline{x}}$ and $(x \pm \delta)$ are interchanged as well as $f_i(\underline{\underline{x}})$ and $f_i(x \pm \delta)$, $\forall_i$. Thus the new $\underline{\underline{x}}$ provide better approximations to the solution of the equations in the next iteration.

(2) Modified calibration procedure: shop and work trip submodels

To overcome the problem of excessive execution time the model has been reformulated into two submodels. This derivation redefines $\underline{\underline{P}}^*$ (and $\underline{\underline{P}}^{v*}$) into two matrices of dimensions (17 × 440) and (17 × 20). The two submodels are briefly outlined below (the terms correspond with definitions previously given for CALBTN2.1)

Shop trip submodel:

$$\text{minimise} \quad \sum_{ij} p_{ij} \ln \frac{p_{ij}}{s_j}$$

$$\text{s.t.} \quad \sum_{ij} p_{ij} = 1$$

$$\sum_{ij} p_{ij} \tilde{c}_{ij} = \bar{c}_s$$

$$\sum_{ij} p_{ij} x_j = \bar{x}.$$

The solution is

$$p_{ij} = \frac{s_j \exp(-\alpha c_{ij} - \beta x_j)}{\sum_{ij} s_j \exp(-\alpha \tilde{c}_{ij} - \beta x_j)}$$

Work trip submodel:

$$\text{minimise} \qquad \sum_{ik} p_{ik} \ln \frac{p_{ik}}{\tilde{r}_i \, w_k}$$

$$\text{s.t.} \qquad \sum_i p_{ik} = d_k, \ \forall_k$$

$$\sum_{ik} p_{ik} \, c_{ik} = \bar{c}_w$$

The solution is

$$p_{ik} = \frac{r_i \, d_k \exp(-\gamma \, c_{ik})}{\sum_i \tilde{r}_i \exp(-\gamma \, c_{ik})}$$

The combined model is now obtained from the two submodels using a renormalization:

$$p_{ijk} = p_{ij} \cdot p_{ik} \cdot \frac{[\sum_{ij} s_j \exp(-\alpha \, \tilde{c}_{ij} - \beta x_j)][\sum_i r_i \exp(-\gamma \, c_{ik})]}{\sum_{ij} r_i \, s_j \exp(-\alpha \, \tilde{c}_{ij} - \lambda \, c_{ik} - \beta x_j)}$$

Thus the model is now solved in two stages. The shop trip submodel is a two parameter live-shop model which employs Powell's algorithm to numerically solve for the parameters α and β. The work trip submodel is a single parameter model, the first derivative of which may now be derived analytically. The solution is then achieved iteratively according to the modified Newton-Raphson method in Jarrat (1970).

The revised calibration procedure takes full advantage of the separability properties of the model in order to avoid excessive numerical evaluations. As a result of the submodel modification core storage is reduced but more important, execution time is reduced by an order of magnitude. This modified calibration procedure helps alleviate the inherent combinatoric difficulties in this model and in so doing reduces the cost of using the model.

3.7 Summary

This chapter has presented a detailed description of the aggregate and disaggregate versions of the model used in this study. The model structure, its derivation and the methodology used in constructing the model have been outlined. A review of the problems, shortcomings and insights gained from previous model versions has shed some light upon how and why the current model versions have evolved and improved upon past experience. Perhaps the most acute problems associated with earlier models were the paucity of the data used to specify the constraints and the relatively small number of categories to which people could be assigned. A strategy for model improvement has been devised which takes these problems into account. It is hoped that these difficulties will be remedied by using the high quality survey data, by increasing the number of categories within the model and by providing income group specific results to more accurately represent the aspects of urban spatial structure under study.

From a technical point of view the inclusion of a more sophisticated and effective calibration procedure has rendered this operational model more cost-efficient as an analytical tool. This is a definite asset in practical terms. A substantial portion of any modelling project involves a heuristic process of modification and re-evaluation of results. By improving the efficiency of model calibration the consumption of project resources (in terms of money, time, effort) is reduced and the overall modelling process is streamlined.

Another key element within the broader framework of this modelling effort is the process of data estimation. This time consuming process is a prelude to actual modelling analysis. The major modifications made to the model structure as a result of this study and the utilization of entirely new survey data necessitates the construction of methods which will "massage" those data into an appropriate form for input to the model. The following chapter will elaborate upon the details of the data estimation.

CHAPTER FOUR

DATA REQUIREMENTS, DATA AVAILABILITY AND DATA ESTIMATION

4.1 Introduction

This chapter provides a detailed account of the survey data used in the analysis and the procedures and methods used to process the raw data. The primary aim of the survey data is to provide input for estimating the parameters of the model. The following section will describe the survey data and its organisation. This is essential for an understanding of how the data are manipulated into a form which is directly input to the model. Subsequent sections will describe how the data are converted into the final form, as represented by the variables within the model specifications in Chapter Three. In so doing the explicit interpretations and definitions of the model variables will become apparent. Estimation procedures for the aggregate model, CALBTN2.1, and the income group disaggregated model, CALBTN2.2, are essentially the same, however some differences do exist (these differences are primarily a result of differentiating between households within different income groups and resulting problems of small sample size). The data processing procedures described in the following pages apply to both CALBTN2.1 and CALBTN2.2. For simplicity these procedures will be described using the variables specified in the aggregate model version.

Unless stated otherwise, the data processing procedures used for the aggregate version are the same as those used for the disaggregate version.

The remainder of this chapter is organized as follows. First, details of the survey data are given in section 4.2. Section 4.3 will then discuss how the shop trip pattern and work trip cost matrices ($\tilde{c}_{ij}$ and c_{ik}) are computed and the manner in which trip making characteristics are derived from the survey data. Finally, details regarding the observed data (to be used in testing the model) are presented.

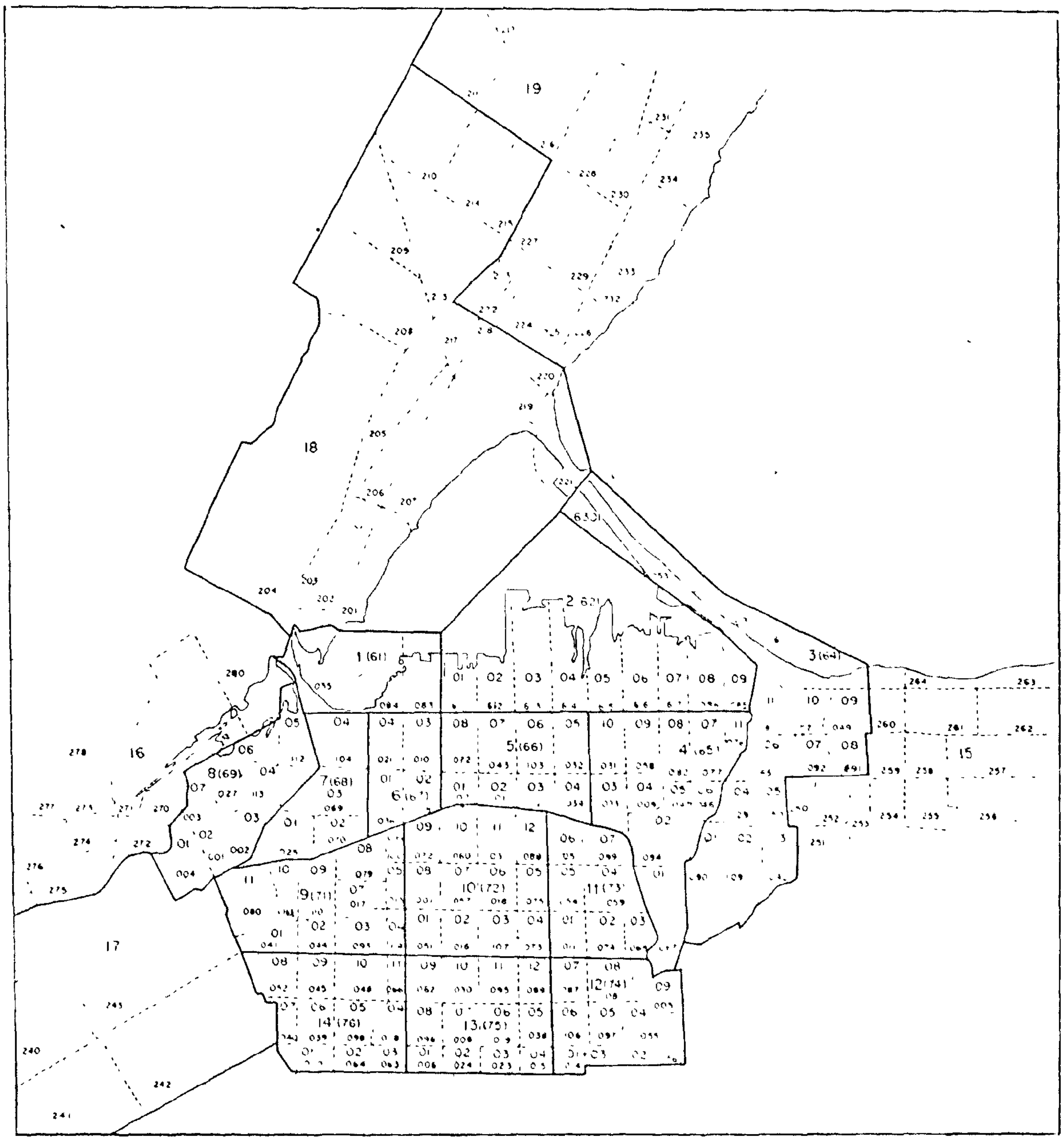
4.2 The 1978 Survey Data

The survey data collected for CALBTN2.1 and CALBTN2.2 (and subsequent versions) consist of volumes of household trip diaries collected over the summer of 1978 in Hamilton, Ontario. They include detailed records for members of households, disaggregated by household characteristics and shop and work trip characteristics. Due to the large number of these characteristics it is impractical to elaborate upon all of them within the context of this paper. Explicit details of the data may be found in Webber (1979c). This section will present an overview of the data used by the model and, equally important, will describe how the data base is organised.

(1) The study area

The study area is shown in Figure 4.1. The city of Hamilton is the focus of the study. The study area is the Hamilton census metropolitan area, which includes neighboring municipalities. To avoid problems with respect to closure, consideration is given only to internal travel. The region is divided into 19 zones; zones 1 to 14 are in the city of Hamilton. These zones are based upon neighborhood and planning districts used by the Planning Department in the city of Hamilton and attempt to encompass more or less homogeneous areas. The municipalities of Ancaster, Dundas and Stoney Creek each comprise one zone (zones 17, 16 and 15

Figure 4.1 The Study Area: Neighbourhood and Zonal Network



respectively while the city of Burlington is divided into two zones of approximately equal population (zones 18 and 19).

Much of the data required in the study can only be obtained from individual households; some are published. Particulars of the household data will be given first.

(2) The travel diary and questionnaire

An areally stratified random sample of households was drawn in May, 1978. Respondent households were asked to keep a travel diary for a period of 2 weeks and to complete a questionnaire which elicited information about the characteristics of each household. From the original sample, 784 households completed both the diary and the questionnaire. Household information gathered from the questionnaire included the number of residents, their place and type of employment, their age and vehicle ownership, the length of residence at this address, household tenancy status (own or rent), housing costs and gross family income.

The diary was designed to determine the trip making behaviour of residents in the study region. A sample page of the diary is shown in Appendix 2.1. Each household was asked to record all trips made by adult members of the household during the two week period and to describe their means of transport, purpose and expenditure on each trip. Trips made by children in the household were not recorded unless they were accompanied by an adult.

Trips were defined to be a string of stops beginning and ending at home. In this way information about multiple purpose and multiple stop trips could be gathered. The most complicated types of trips to record are those in which two (or more) persons begin a trip together but then go their separate ways after some point. Generally, a separate trip is recorded for each adult who makes at least one stop independently of other adults on the trip.

The information from the diaries is organized into four types. For each household in each zone the information is coded in the following way: one record describes the household characteristics; one record per resident describes the characteristics of that person; one record per trip is required to describe that trip; and finally, for each stop made on that trip, one record describes the details of that stop. The information which is coded by each of these four categories is illustrated in Appendices 2.2 to 2.5. The entire data base consists of 19 sequential files; one for each zone in the study region.

The household characteristics data are shown in Appendix 2.2. Each household is identified by a unique four digit number. The first two digits are the zone number while the last two digits are sequential household identifiers within the zone. It is important to note that all locations within each zone are described by neighborhood numbers (188 in total). The distribution of neighborhoods within the study region is shown in Figure 4.1.

Each person residing in the household at the time of the interview is described on the person characteristics card. The explicit details of this information are shown in Appendix 2.3. For each resident information is recorded about their family status, employment status, occupation, industry of employer, location of work and vehicle ownership status. Lists of industrial and occupational categories are shown in Appendix 2.6.

Finally, details of the information gathered for trips and stops are illustrated in Appendices 2.4 and 2.5. Note that the stop information includes means of travel, first, second and third purposes of stop, types of goods bought and expenditure. A variety of categories is available for describing shopping, recreational and social stops (purposes 1, 2, 3). Appendix 2.7 lists the classification of stops according to social and recreational purposes, and lists the types of shopping goods.

The fundamental link between the stop information and the trip information is used to define trips according to purpose. Trips are defined to be work, shopping, recreational or social trips if and only if work, shopping, recreational or social activities respectively constitute the main (first) purpose at the first stop. This method of categorising trips by purpose may result in underestimates since the first stop may be only incidental to the main purpose of the trip.

Although O'Kelly (1980) has shown that complex trip chains do occur, this paper assumes that the main purpose of the first stop can be used as the criterion to define trip purpose. This assumption will be explored further in the next two chapters.

(3) Published Data

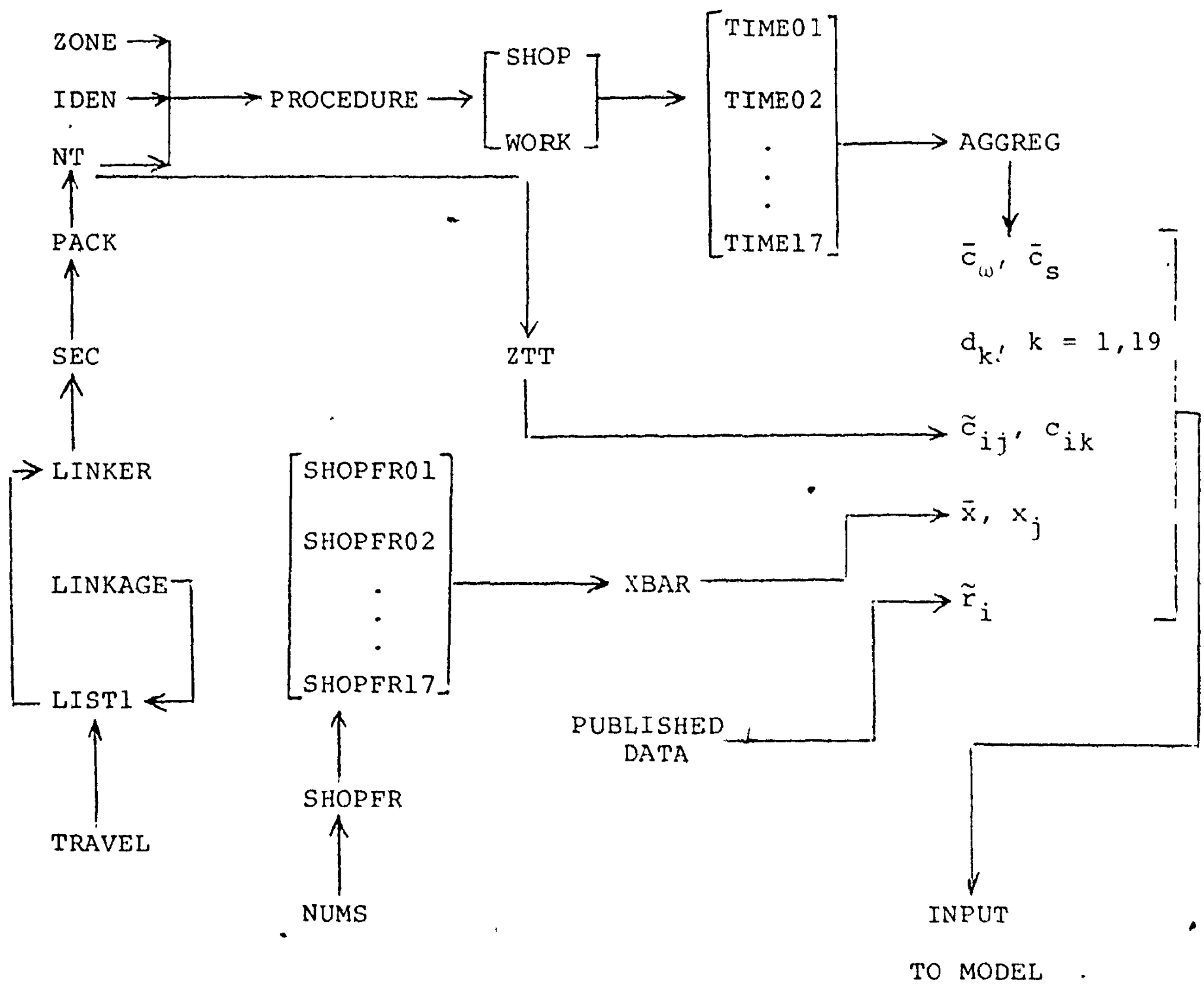
An assortment of published data were collected in order to discretize the characteristics of each neighborhood. Most of the information was obtained from recent RMHWPDD sources. The precise nature of these data are given in Webber (1979c). A cursory description of these data, which have been stored in a data file called IDEN, is given in Appendix 2.8.

4.3 Travel Times, Shop Trip and Work Trip Cost Matrices

The entire data processing procedure is schematically represented in Figure 4.2. The link responsible for computing the travel times matrices begins with program TRAVEL in the lower left corner and ends with the shop trip and work trip travel times matrices, $\tilde{c}_{ij}$ and c_{ik}, on the right. Basically two types of travel times matrices are computed: interneighborhood and interzonal. All household locations, shop and work trip origin-destination matrices and travel times calculations are computed on the neighborhood level, then aggregated to the zonal level. In this way the scale of resolution for trip making behaviour is fine enough to capture detail which otherwise would be lost at the intrazonal level. This is an improvement over methods used to compute travel times in previous model versions. The neighborhoods used in this study are small enough to account for virtually all of the potential interneighborhood interaction thereby avoiding one of the problems of zone size mentioned by Batty (Batty et al., 1973, 353).

Between two and six timed journeys by car were made over each of the streets used as a neighborhood boundary. The journeys were made off the peak daily travel periods. These neighborhood boundary travel times were converted into a matrix of interneighborhood travel times (SEC) computed using a manhattan metric technique (programs TRAVEL, LINKAGE,

Figure 4.2 Data Processing Flowchart



LINKER). The travel times between non-adjacent neighborhoods were computed using a minimum path algorithm. The upper triangular portion of the interneighborhood travel times matrix, SEC, is stored in NT. Interzonal shop trip and work trip travel times are then computed from NT using the program ZTT. Complete details of this scheme for computing travel times may be found in Webber (1979c).

Within the model the shopping trip pattern cost matrix, $\tilde{c}_{ij}$, is calculated on the basis of the above-mentioned travel times matrix and the matrix of the expected number of shopping trips per household category (x_j). Specifically, for $j = 1, \ldots, 20$, $\tilde{c}_{ij}$ is computed as the product of [the time taken to travel from zone i to the shopping facility located in zone j] and [the expected number of shopping trips per week made by households observed to make 0 or 1 shop trip per week]. For shopping patterns involving two facilities $\tilde{c}_{ij}$ $(j = 21, \ldots, 230)$ is computed as the product of [the sum of the travel times to the two centres] and [the expected number of trips per week made by households observed to make two shop trips per week]. For $j = 231, \ldots, 440$, $\tilde{c}_{ij}$ is calculated as the product of [the sum of the travel times to the two centres] and [the expected number of trips per week made by households observed to make > 2 shop trips per week]. The subscripting scheme used for $\tilde{c}_{ij}$ is also used for the solution matrix $\underline{\underline{p}}^*$ (or $\underline{\underline{p}}^{v*}$ in CALBTN2.2) and is illustrated in Figure 4.3.

1,..., 20,21,...	...40
(1,1)...	...(1,20)
41...	...59
(2,2)...	...(2,20)
60...	...77
(3,3)...	...(3,20)
78...	...94
(4,4)...	...(4,20)
95...	...110
(5,5)...	...(5,20)
111...	...125
(6,6)...	...(6,20)
126...	...139
(7,7)...	...(7,20)
140...	...152
(8,8)...	...(8,20)
153...	...164
(9,9)...	...(9,20)
165...	...175
(10,10)...	...(10,20)
176...	...185
(11,11)...	...(11,20)
186...	...194
(12,12)	...(12,20)
195...	...202
(13,13)...	...(13,20)
203...	...209
(14,14)...	...(14,20)
210...	...215
(15,15)...	...(15,20)
216...	...220
(16,16)...	...(16,20)
221...	...224
(17,17)...	...(17,20)
225...	...227
(18,18)...	...(18,20)
228...	...229
(19,19)...	...(19,20)
230	
(20,20)	

Figure 4.3 Subscripts for $\tilde{c}_{ij}$ and $\underline{P}^*$ for 20 Facilities

(6,10) LOCATION OF FACILITIES

41 SUBSCRIPT IN ARRAY

NOTE: COMBINATION OF FACILITY LOCATIONS FOR SUBSCRIPTS 231,..,440 IS IDENTICAL TO 21,...,230.

The work trip cost matrix, c_{ik}, is computed as a function of the interzonal travel times. Costs associated with the no work trip category (unemployed/retired) present some difficulties. Examination of past literature invites speculation as to the most appropriate manner in which to deal with this problem. The functional form of the current model versions is essentially that of an attraction-constrained, gravity type residential location model which allocates households around work places according to known travel cost relationships. The simplest hypothesis is that the distribution of residences is determined by the distribution of work places, in particular by workers' accessibility to work places (this was the residential component of the Lowry model). As Wilson (1970,71) points out, there are obvious weaknesses in this assumption. Increasing numbers of households, for example, of old people, contain no workers. Preliminary analysis using the sample data collected for this study have revealed that a large proportion of households in lower income groups contain no workers. Senior (1973) has proposed the use of balancing factors to correct errors due to non-trip makers. However, in light of findings by Webber (1979b), Wilson (1970a) and Kain (1962) it is felt that using balancing factors to "patch up" these errors is unsatisfactory. In this model a more appropriate way to incorporate these individuals has been devised.

The method of assigning costs to the no work trip category is to compute the weighted mean cost of a work trip to all other zones using the employment vector d_k as a weight:

$$c_{i,20} = \frac{\sum_{k=1}^{19} c_{ik} d_k}{\sum_{k=1}^{19} d_k} .$$

Methodologically, this formulation for $c_{i,20}$ can be justified by postulating that locational decision-making criteria for the no work trip households are still influenced by the present location of work places, assuming that those households in the unemployed/retired category have previously been employed.

4.4 Trip Making Characteristics

Data describing several characteristics of the trip making behaviour of survey households are required for input to the model. Included in these data are the spatial distribution of work trip ends, origin-destination matrices and estimates of average travel times. The following pages will describe the sequence of operations followed to produce these input data according to the data processing flowchart shown in Figure 4.2. A procedure is used to drive the programs WORK and SHOP which assemble the data for one zone at a time. After this procedure has been run for each zone program AGGREG processes the data from each zone to produce results for the Hamilton region. The data processing programs are now described.

Program WORK

Program WORK investigates the journey to work for heads of household (up to one per household) who work outside the home. As mentioned in section 4.2, a work trip is so defined if and only if the main purpose of the first stop is for working. In this study the assumption of one worker per household implies that only the work trip behaviour of the head of the household is relevant to the residential location decision. A possible alternative would be to assume that all work trips made by working members of the household are relevant; however, as Wilson (1970,80) points out, this presents additional difficulties in specifying the model, to include work trip costs incurred by the entire household.

Program WORK excludes from the analysis all non-permanent residents of the household, residents who are not presently employed outside the home, and work trips made by heads of household whose place of work is outside the region. Burlington destinations (zone 18 and 19) are included in the analysis.

This program computes several origin-destination matrices (neighborhood-neighborhood, zone-neighborhood, zone-zone), and the sum and sum of squares of travel times. The sum of work trip travel times is the sum of times for one-way journeys to work and is computed as:

$$\text{sum} = \sum_i \sum_j OD(i,j) \times TIME(i,j) \ ,$$

where, sum = the sum of travel times for one zone

OD(i,j) = the number of permanent residents who live in neighborhood i and work outside the home in neighborhood j , and

TIME(i,j) = the interneighborhood travel times matrix, from NT.

Program SHOP

This program examines home-based grocery shopping trips. It is assumed that grocery shopping goods are homogeneous with respect to the shopping behaviour they generate, and that this behaviour (shopping for convenience goods) is distinct from shopping for other classes of goods (e.g. durable goods). The following definitions distinguish between what is meant by a home-based grocery shopping trip and, more generally, what is considered to be a home-based shopping trip.

A trip is considered home-based shopping if any of the possible purposes at the first stop is shopping and if the trip begins at the respondent's home. Trips where no shopping takes place on the first stop but some shopping does occur on a subsequent stop would not be considered home-based shopping according to this definition.

Home-based grocery shopping trips are a subset of home-based shopping trips as defined above, i.e. there must be some shopping on the first stop. If no groceries are bought on the first stop a trip may still be designated home-based grocery if groceries are bought on one of the succeeding stops. The only stops probed by program SHOP are those which have a grocery shopping purpose or which are returns home after a grocery shopping stop. All stops which are in the Burlington neighborhoods or outside the study region do not contribute to the sum of travel times. The program computes origin-destination matrices, the sum and sum of squares of travel times (cumulated for round trips). An example of how these arrays and the sum of travel times are computed is given with reference to Figure 4.4 .

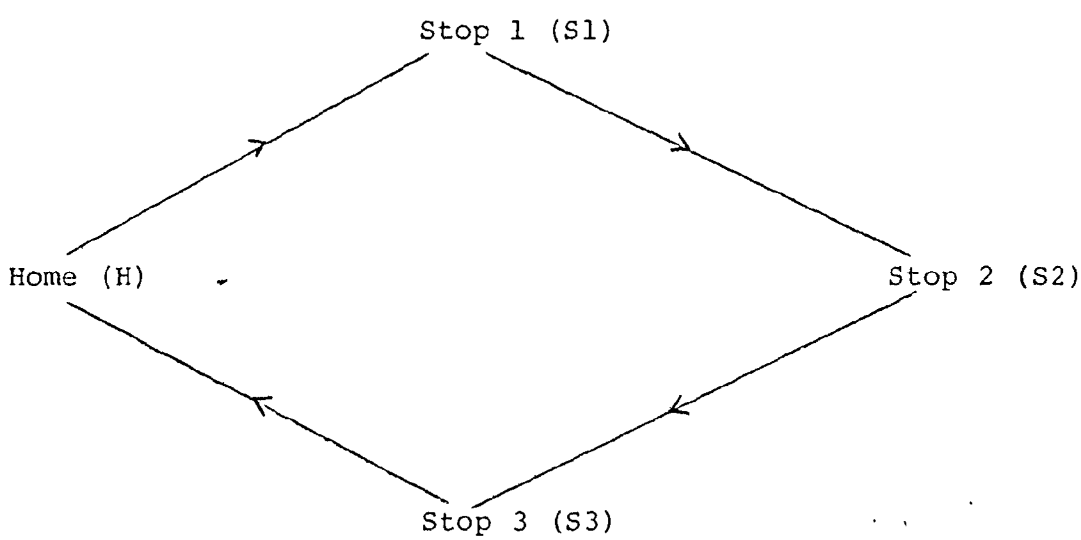
If any one of the three possible purposes for the first stop, S1 , is shopping, and if groceries are purchased on any one of the stops (S1, S2 or S3) then the trip is defined to be a grocery shopping trip; otherwise the trip is ignored. For example, if S1 is not a grocery shopping stop but S2 and S3 are, then the origin-destination array entries (H,S2), (H,S3) are incremented and the sum of travel times is increased by the travel times (H,S2), (S2,S3), (S3,H) .

Program AGGREG

AGGREG produces the following Hamilton region results: origin-destination matrices for shop and work trips; mean cost estimates $\bar{c}_s$, $\bar{c}_s^v$, $\bar{c}_w$, $\bar{c}_w^v$ and their variances, and; the work trip destination vectors d_k and d_k^v. The origin-destination matrices output from AGGREG are used to formulate the observed work and shop trip matrices for testing the model's predictions.

The mean cost estimates are computed as

Figure 4.4: A Hypothetical Grocery Shopping Trip



population weighted averages in the following manner.

For CALBTN2.1 the weighted averages are

$$\bar{c}_s = \frac{\sum_{i=1}^{17} \left(\frac{s_i}{y_i}\right) \cdot p_i}{\sum_{i=1}^{17} p_i} ,$$

and

$$\bar{c}_w = \frac{\sum_{i=1}^{17} \left(\frac{w_i}{n_i}\right) \cdot p_i}{\sum_{i=1}^{17} p_i} ,$$

For CALBTN2.2 the weighted averages are

$$\bar{c}_s^v = \frac{\sum_{i=1}^{17} \left(\frac{s_i^v}{y_i^v}\right) \cdot p_i}{\sum_{i=1}^{17} p_i} , \qquad v = 1, \ldots, 4 ,$$

and

$$\bar{c}_w^v = \frac{\sum_{i=1}^{17} \left(\frac{w_i^v}{n_i^v}\right) \cdot p_i}{\sum_{i=1}^{17} p_i} , \qquad v = 1, \ldots, 4 ,$$

The above variables are defined as:

s_i = total shop trip travel time for households in zone i,

s_i^v = total shop trip travel time for households in income group v in zone i,

w_i = total work trip travel time for heads of households in zone i,

w_i^v = total work trip travel time for heads of households in income group v in zone i,

y_i = the number of sample households in zone i,

y_i^v = the number of sample households in income group v in zone i,

n_i = the total number of work trips by heads of households in zone i,

n_i^v = the total number of work trips by heads of households in income group v in zone i, and

p_i = the population of zone i.

There are several reasons for computing the mean cost estimates in the above manner. It is the purpose of this model to predict Hamilton, not just the sample. Accordingly, the zonal averages have been multiplied by the proportion of the population in those zones. This method of weighting should result in more reliable model predictions since the mean distribution of travel in the Hamilton area has been estimated solely on the basis of survey data collected by a single random sample of households. It is then justifiable to scale up the zonal samples according to population size. The variability of zonal populations and the distribution of sample households in the four income groups is shown in Table 4.1. For example, in zone 5 there are a total of 35 households representing a population of 45,824 people, but in zone 13 the same number of households represent only 2,361 people. In order to realistically predict the region we must take these zonal population differences into account.

It should be mentioned that the raw data for household characteristics restricts the classification of households to 6 income groups as shown in Appendix 2.2. For this analysis it was decided to condense the last three of these categories to form a single > $30,000 category due to the very small sample sizes or complete lack of observations in the highest income groups for certain zones. Only those households which provided income information are included in Table 4.1 and subsequently only those households have been included in the analysis for CALBTN2.1 and CALBTN2.2.

Table 4.1

Population Distribution and Distribution of Sample Households According to Income Group

ZONE	POPULATION	GROSS HOUSEHOLD INCOME PER ANNUM $0-10,000	$10,001-20,000	$20,001-30,000	> $30,000	TOTAL HOUSEHOLDS
1	6,043	12	11	3	2	28
2	8,704	3	11	9	1	24
3	33,625	5	17	12	6	40
4	43,475	8	24	12	6	50
5	45,824	10	14	8	3	35
6	26,915	11	18	4	1	34
7	17,132	8	17	7	6	38
8	19,498	10	11	13	4	38
9	28,998	1	24	12	8	45
10	45,013	1	21	14	1	37
11	22,785	3	19	12	6	40
12	9,608	0	16	12	4	32
13	2,361	1	16	11	7	35
14	2,414	0	14	21	3	38
15	19,252	2	16	14	10	42
16	21,920	2	12	20	10	44
17	10,284	5	7	15	13	40
TOTAL	363,850	82	268	199	91	640

The fundamental hypothesis behind income disaggregation is that residential location and trip making behaviour are affected by income. A closer look at Table 4.1 reveals that there are marked zonal variations in the distribution of sample households within the four income groups. The largest portion of low income households (group 1) are found in the centrally located zones (1, 5, 6, 7, 8). Zones 9 to 16 (the "Mountain" and Dundas) have the lowest numbers of households in income group 1. Zones 15, 16, and 17 (Stoney Creek, Dundas and Ancaster) contain high proportions of households in income group 4. It follows that zonal differences in travel behaviour are to be expected, hence the use of population weighted averages is deemed suitable.

One difficulty was encountered with respect to the population weightings used for the income group mean cost estimates, $\bar{c}_w^v$ and $\bar{c}_s^v$. Unfortunately the zonal populations used as weights are not disaggregated by income group. An extensive search of locally available data sources was conducted in an attempt to obtain income-specific zonal population figures, however, no appropriate data could be found. Accordingly, the zonal population proportions for the total population (i.e. all income groups) were the most unbiased weights available for computing $\bar{c}_s^v$ and $\bar{c}_w^v$.

In addition to the mean cost estimates, AGGREG uses the journey to work data to compute the relative frequency work trip end arrays d_k and d_k^v. These work trip destination arrays are population weighted in accordance with the mean cost estimates.

For example, the calculation for CALBTN2.1 is as follows.

$$d_k = \frac{\sum_{i=1}^{17} OD(i,k) \cdot p_i}{\sum_{i=1}^{17} p_i} \Bigg/ \sum_{k=1}^{19} \frac{\sum_{i=1}^{17} OD(i,k) \cdot p_i}{\sum_{i=1}^{17} p_i}, \quad k = 1,\ldots,19$$

For CALBTN2.2 d_k^v $(v = 1,\ldots,4)$ is computed in the same manner except that the origin-destination matrix would include only households in income group v. The no work trip category is incorporated in d_{20} (and d_{20}^v) as the observed proportion of households which are unemployed or retired; d_k $(k = 1,\ldots, 20)$ is then normalized to sum to 1.

Program XBAR

This program computes $\bar{x}$, $\bar{x}^v$, x_j, x_j^v and the standard deviations for $\bar{x}$, $\bar{x}^v$. These variables are used in the constraint on the average number of shop trips (per week, per household). The procedure is identical for both CALBTN2.1 and CALBTN2.2.

Recall that the shopping patterns $(j = 1,\ldots, 440)$ are designed as follows:

pattern 1,..., 20 = one shopping trip per week (shopping class 1)

21,..., 230 = two trips per week (shopping class 2)

231,..., 440 = more than two trips per week (shopping class 3)

x_j ($j = 1, \ldots, 20$) is the expected number of shopping trips per week for households in shopping class 1; x_j ($j = 21, \ldots, 230$) is the expected number of times a household goes on two shopping trips per week if it is in shopping class 2; x_j ($j = 231, \ldots, 440$) is the expected number of times a household goes on more than two shopping trips per week if it is in shopping class 3. The data used to compute $\bar{x}$ and x_j are derived as follows (refer to Figure 4.2).

Program NUMS selects the relevant household and shop trip information from the raw data to be input to SHOPFR. For each zone the SPSS frequencies program SHOPFR provides data on the number of households in each zone which had 0, 1, 2, etc. home-based grocery shopping trips in the two week period of the diary. Output from SHOPFR are 17 data files corresponding to each zone. These data are then input to program XBAR which computes the Hamilton region frequency distribution of grocery shopping trips per week, per household. Extracted from this regional frequency distribution are the x_j and $\bar{x}$ estimates. A brief example will serve to illustrate how this is done.

SHOPFR produces the following data for the total sample of households in zone 2:

0	households made	0	shopping trips during the two week period,		
5	"	"	1	"	"
3	"	"	2	"	"
6	"	"	3	"	"
3	"	"	4	"	"
3	"	"	5	"	"
2	"	"	6	"	"
1	"	"	7	"	"
0	"	"	8	"	"
1	"	"	9	"	"

If a household made 0 trips in two weeks then it made 0 trips in the first week and 0 trips in the second week; if a household made 1 trip then it made 0 trips in one week and 1 trip in the other week; if it made two trips then it was assumed to make 1 trip in each of the two weeks. In general, a household which made 2n trips was assumed to make n trips in each week, and a household which made 2n + 1 trips was assumed to make n trips in one week and n + 1 trips in the other week. Thus for the zone 2 data given above the frequency distribution of shopping trips <u>per week</u> is:

Trips per week	Number of households	Number of households × proportion of population
0	5	.11959
1	17	.4066
2	15	.35878
3	8	.19135
4	2	.0478
5	1	.0239
6	0	0.0
7	0	0.0
	48	1.14812 = .0239 × 48

Each of the 24 households in the zone yields two observations, one for its first week and one for its second.

A similar frequency distribution is computed for each zone. The regional frequency distribution is computed in XBAR by multiplying the zonal frequency distributions by the population weights as was done for the mean cost estimates. For example, the third column of the above table shows the product of the frequencies multiplied by the proportion of the region's population in that zone (zone 2 contains .0239 of the region's population). The regional weighted proportion of households which fall into class j is then

$$\sum_{i=1}^{17} \left(\frac{f_{ij}\ p_i}{\sum_{j=0}^{8} f_{ij}\ p_i} \right) \left(\frac{n_i}{\sum_{\ell=1}^{17} n_\ell} \right) ,$$

where, f_{ij} = the number of households in zone i which made j trips per week,

p_i = the proportion of the regions population in zone i, and

$$n_i = \sum_{j=0}^{8} f_{ij}$$

The table of these values is given below.

Trips per week	Proportion of households		Average number of trips per week by group
0	.1094	.4103	.7333
1	.3009		
2	.2921	.2921	2.0
3	.1781		
4	.0736		
5	.0337	.2981	3.6169
6	.0087		
7	.0032		
8	.0008		
	2.0		

XBAR then classifies households into groups: group 1 make 0 or 1 trips per week; group 2 make 2 trips per week; and group 3 make 3 or more trips per week. The proportion of households in each of these groups is given in the third column of the above table. For households in group 1 the average number of trips per week is

$$\frac{(0 \times 0.1094) + (1 \times .3009)}{.4103} = 0.7333$$

Similar averages are given in the fourth column of the table for the other two groups. These numbers are interpreted as follows. Group 1 households make 1 trip per week (shopping patterns 1,..., 20) with a frequency of .7333 times per week. Group 2 households make 2 trips per week (shopping patterns 21,..., 230) with a frequency of 1 time per week. Group 3 households make 2 trips per week (shopping patterns 231,..., 440) with a frequency of 1.808 times per week. These are the elements of x_j $(j = 1,..., 440)$.

The observed average number of home-based grocery shopping trips per week, $\bar{x}$, is computed from the above table as:

$$(0 \times .1094) + (1 \times .3009) + \ldots + (8 \times .0008) = 1.9633$$

The Location of Shoping Centres

In order to model the observed pattern of shopping "centres" within the Hamilton region, it is necessary to locate 20 "centres" in the]7 zones (with the possibility of several "centres" in any one zone). This can be done by locating centres in proportion to the actual retail floorspace, or in proportion to the observed shopping trip destinations in the total sample. These are not very similar, but the latter was chosen. The distribution is shown in Table 4.2.

Table 4.2 Distribution of Shopping Centres

Zone of Allocation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
number of facilities			2	3	2	3	1	1	1	2	2				1	2	

4.5 Endogenous Distributions and the Observed Data

In order to test the accuracy of model predictions it is necessary to extract from $\underline{\underline{P}}^*$ the endogenous distributions which are of interest, and to compare the goodness of fit of these predicted distributions with observed zonal populations, shop trip and work trip matrices. The solution probability matrix $\underline{\underline{P}}^*$ is structured as P_{ijk}: for $k = 1, \ldots, 20$; $i = 1, \ldots, 17$; $j = 1, \ldots, 440$ (refer to Figure 4.3). The relevant predicted distributions are of the following simplified form:

$\sum_k \sum_j P_{ijk}$ is population of zone i,

$\sum_j \sum_i P_{ijk}$ is the number of work trips ending in zone k,

$\sum_{j=1}^{20} \sum_k P_{ijk}$, is live in i and make 0, 1 shop trips per week,

$\sum_{j=21}^{230} \sum_k P_{ijk}$, is live in i and make 2 shop trips per week,

$\sum_{j=231}^{440} \sum_k P_{ijk}$, is live in i and make > 2 shop trips per week.

The predicted shop trip pattern matrix is then converted into a relative frequency destination matrix which is normalized to sum to 1. Similarly, the predicted population array and the predicted work trip origin-destination matrix are both normalized to sum to 1.

The observed zonal population array is assembled from the published neighborhood population data contained in IDEN. As previously mentioned, there are no separate observed zonal population distributions corresponding with each income group. Instead, the observed population distribution is for the total population in each zone.

The observed regional shop trip origin-destination matrix is computed in AGGREG using the zonal shop trip origin-destination matrices output from program SHOP. This raw flows, regional origin-destination matrix is converted into a population weighted, relative frequency matrix in the following manner.

$$S^*_{ij} = \frac{S_{ij}}{\sum_{j=1}^{17} S_{ij}} \times \frac{p_i \times \left(\frac{\sum_{j=1}^{17} S_{ij}}{h_i}\right)}{\sum_{i=1}^{17} p_i \times \left(\frac{\sum_{j=1}^{17} S_{ij}}{h_i}\right)}$$

$$= S_{ij} \times \frac{\frac{p_i}{h_i}}{\sum_{i=1}^{17} p_i \times \left(\frac{\sum_{j=1}^{17} S_{ij}}{h_i}\right)}$$

where,

S^*_{ij} = population weighted, observed relative frequency of a grocery shopping trip from zone i to zone j,

S_{ij} = observed number of grocery shopping trips from zone i to zone j,

p_i = population of zone i , and

h_i = number of sample households in zone i.

S^*_{ij} is then normalized so that $\sum_{ij} S^*_{ij} = 1.$

The observed regional work trip origin-destination matrix is obtained from AGGREG, and includes the observed proportion of households within the no work trip category. This raw flows origin-destination matrix is population weighted as follows.

$$W^*_{ij} = \frac{W_{ij}}{\sum_{j=1}^{20} W_{ij}} \times \frac{p_i}{\sum_{i=1}^{17} p_i},$$

where, W^*_{ij} = population weighted, observed relative frequency of a work trip from zone i to zone j,

W_{ij} = observed number of work trips from zone i to zone j , and

p_i = population of zone i.

W^*_{ij} is then normalized so that $W^*_{ij} = \dfrac{W^*_{ij}}{\sum_{ij} W^*_{ij}}.$

4.6 Summary

This chapter has provided an account of the procedures used to refine the raw data into the form used by the model. This data estimation strategy has involved substantial decision-making since the rationale behind explicit variable definitions has a pronounced effect upon the model's predictions and the interpretation of these predictions. The methods are justified by theoretical and logistical considerations. Also, efforts have been made to retain objectivity and consistency within and between each component of the data processing system.

The elaboration of data estimation methods presented here sets the stage for the model analysis which follows.

CHAPTER FIVE

AN EMPIRICAL INVESTIGATION INTO DISAGGREGATION

5.1 Introduction

This chapter presents a detailed description of the analysis, the purpose of which is to determine whether or not the disaggregated model provides a more accurate representation of observed spatial structure (in terms of residential location, work trip and shop trip behaviour) than does the aggregate model. The analysis is in two parts - a preliminary inductive investigation followed by deductive analysis.

The strategy for analysis and the sequence of presentation is as follows. First the results of the data estimation process are subjected to a preliminary investigation which examines the premise that significant differences exist in the work trip and shop trip characteristics and shop trip frequency between the different income categories. The mean value estimates ($\bar{c}_w^v$, $\bar{c}_s^v$, and $\bar{x}^v$) are examined and the significance of the between-group differences in these mean values is determined using a t test. The two model versions, CALBTN2.1 and CALBTN2.2, are then calibrated and the goodness of fit of the model's predictions are tested with respect to observed data using several statistical measures. Following this is a discussion of the final results of the analysis.

5.2 Preliminary Analysis

The average cost results for the four income categories and the total sample are given in Table 5.1. Shop trip cost is defined to be grocery purpose round trip travel time (in minutes) per journey, and work trip cost is one-way journey to work travel time (in minutes). The highest and lowest average shop trip costs are for income groups 4 and 2. A noticeable gap exists in the mean shop trip values between income groups (1,2) and (3,4). Similarly, the mean work trip values for the four income groups indicate a distinctive difference in work trip lengths between income categories (1) and (2,3,4). In general the results show that the mean shop trip and work trip costs increase from the lowest to the highest income groups, demonstrating the correlation between mean trip length and income. This provides some confirmation of the premise that considerable differences exist in travel characteristics between income categories and is in accordance with Hathaway's findings (which examined work trips only). The question of the significance of these differences in mean trip length will be addressed shortly.

Table 5.1

Average Costs and Standard Deviations

income group	$\bar{c}_s$	s.d.	$\bar{c}_w$	s.d.	$\bar{x}$	s.d.
1	18.485	1.358	6.004	0.434	1.928	1.104
2	18.410	0.669	9.004	0.334	1.932	0.457
3	21.26	0.859	8.924	0.362	1.970	0.368
4	22.143	0.856	8.773	0.339	2.204	1.100
aggregate	19.57	0.730	8.819	0.344	1.963	0.241

Work trip and shop trip travel times are of approximately equal duration (considering the one-way/round trip difference); grocery shopping trips are slightly longer on average than work trips. The standard deviations indicate that the largest variability in work trip and shop trip length occurs in the lowest income group.

The mean number of shop trips per week for the four income groups does follow the same general trend found in the average travel costs. The results show that $\bar{x}^v$ is highest for income group 4, followed by 3, 2, and 1. Again, dispersion around the mean is highest for income category 1 (and second highest for income category 4).

An obvious question arising from a close examination of Table 5.1 concerns the significance of small differences in average cost values. A discussion of this question is found in Appendix 3 where, it is found that: the differences in mean shop trip cost between income groups are significant except between groups 1 and 2; for mean work trip cost the only significant differences exist between group 1 and all other groups (i.e. groups 2, 3, 4 are not significantly different); and for the mean number of shop trips per week no significant differences are found between the income groups. This latter result is not surprising if one closely examines Table 5.1. There is virtually no difference in the mean number of shop trips per week per household between the four income groups. The overall mean number of shop trips per week is about 2.

Due to the low magnitude of these mean values and the fact that shop trips occur in discrete units, the variability between income groups is bound to be relatively large compared with the mean values, hence it is difficult to discern significant differences between income categories. As mentioned in the preceding chapter, the selection of four income group intervals was largely dictated by the existing income group classes in the raw survey data and by a consideration of adequate sample sizes. It would appear that the four categories do distinguish the inhomogeneity in work trip and shop trip travel behaviour with respect to differences in household income. It is doubtful whether another externally imposed grouping between categories could have bettered the distinctions of shop trip frequency between income categories.

Table 5.2 gives the observed work trip destination arrays for the four income groups (d_k^v; k=1,20) and for the total sample. Each of these arrays sums to 1000. As expected, a significant proportion (31.4% to 42.6%) of households in income groups 2, 3 and 4 have work trips ending in zone 2, the industrial heart of Hamilton. This reflects upon one of the idiosyncracies of the Hamilton region labour force; i.e. the relatively large blue collar/labourer component. The largest portion of working households in income group 1 have work trip destinations in zone 11, Hamilton 'mountain'.

By far the most striking differences between income groups is the variation in the proportion of households that fall into the no work trip category. The percentages of households in income groups 1 through 4 which are either unemployed or retired are, 50.2, 27.9 20.3 and 23.8. In the aggregate (treating the total sample as a homogeneous group) unemployed or retired households constitute 32.4% of the total number of households. This reveals one of the major underlying weaknesses of the aggregate model, i.e. it does not accurately represent the (lack of) work trip characteristics of households in the $0 - 10,000 income groups. It is not realistic to assume that low income households conform to the mode of trip making behaviour which is representative of the total populace. More adequate ways of dealing with this segment of the population must be devised. Further attention will be devoted to this problem in the final chapter.

TABLE 5.2

Work Trip Destination Arrays: d_k and d_k^v

Destination Zone

Income group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	7	29	38	0	15	62	33	79	95	0	128	0	0	6	0	0	0	0	6	502
2	6	314	31	49	46	99	48	26	10	27	0	0	0	0	14	15	6	17	11	279
3	19	321	8	36	57	142	34	62	9	31	0	0	0	0	35	9	10	20	4	203
4	0	426	0	24	82	59	13	42	12	51	10	0	7	0	5	7	8	6	10	238
aggregate	9	245	19	27	47	101	32	75	21	34	9	0	2	2	17	8	7	11	11	324

The preliminary analysis has produced several interesting results. The majority of the differences that exist in the mean shop trip costs and mean work trip costs between income groups have been shown to be significant. More specifically, mean travel times for grocery shopping and working purposes increase as household income increases. This indicates that distance (or travel time) is less of a deterrent for households in higher income brackets, i.e. members of higher income households are willing to travel further to shop and to work. Hathaway (1975) has obtained similar results for work trips using a disaggregated trip distribution model. Also, Senior (1977, 302) in an empirical study using a disaggregated information-minimising residential location model, found that mean work trip costs gradually increased on a social class scale of (unskilled manual, semi-skilled manual, skilled manual, intermediate skilled manual, professional), thus lending support to the findings of this preliminary analysis.

Observed differences in shop trip frequency (mean number of shop trips per week per household) between income groups have proven to be not significant. There is essentially no difference in the mean number of grocery shopping trips per week per household for the different income groups. The results of this study and those of a sensitivity analysis conducted on an earlier version of the model (Webber, et. al., 1979a) have cast some doubt upon the effectiveness of the constraint which employs this mean value. The problem may be a result of estimating this mean, $\bar{x}^v$, on a 'per week' basis (which captures little potential variation in grocery shopping frequency) or may be a result of small sample sizes for income groups. As can be seen in Table 5.1 the standard deviation of the mean number of shop trips per week for the aggregate sample is far smaller than the corresponding standard deviations for the four income groups. This would imply that it is the small sample sizes for income groups which is creating the problem. In either case it would be useful to more thoroughly examine this constraint. The mean number of shop trips estimate may be more effective if it were based upon observations over a longer period, say two weeks or one month.

The preliminary analysis shows that the least dispersion around the mean values occurs for the two middle income groups. An examination of Table 5.1 indicates that the mean cost estimates for the total sample most accurately reflect the behaviour of income groups 2 and 3. Income group 1 shows consistently more variations about the mean values than any other income group. This is probably due to the extraordinarily high proportion of unemployed/ retired households within this group, which likely results in more sporadic and less frequent shop and work trips.

5.3 Calibration and Testing

Following the preliminary analysis both the aggregate and disaggregate model versions were calibrated and run. Initial runs with both model versions revealed problems with respect to excessive execution time. Convergence required between 20 and 30 iterations of Powell's algorithm before reaching a final solution; however, by adjusting the convergence criterion and by using improved starting values the model most often converged in less than half the initial number of iterations.

The following pages disclose the results of model testing. The preliminary analysis has verified the premise that inhomogeneity of travel behaviour exists between different income groups with respect to observed differences in the mean costs $\bar{c}_w^v$, $\bar{c}_s^v$, $\bar{x}^v$ and worktrip destination arrays d_k^v. What remains to be seen is to what extent these differences are reflected in the predicted spatial distribution of population, work trips and shop trips and the accuracy of model predictions; i.e. a comparison of observed/predicted residential location, work trip and shop trip origin-destination matrices for each income category and for the aggregate sample.

A commonly used measure of departure between two sets of stochastic variables is the chi-squared test. The chi-squared statistic is defined as

$$\sum_i \sum_j \frac{(p_{ij} - p_{ij}^*)^2}{p_{ij}^*}$$

where p_{ij} and p_{ij}^* are the two matrices to be compared (for example, p_{ij} may be the observed matrix; p_{ij}^* may be the synthesized matrix). As the number of degrees of freedom increases, the χ^2 distribution approaches the normal distribution. In spatial interaction modelling there are usually a large number of degrees of freedom due to a relatively large number of cells in the origin-destination matrices, therefore the use of this statistic arbritrarily imposes a normal density function which may be inappropriate.

The chi-squared statistic is not as easily applied or interpreted as other statistics since many cells in the compared matrices may be zero (Morphet, 1975). These zero cells must be removed by aggregation or exclusion from the analysis, thereby measuring correspondence between matrices with reduced information. There is much controversy over acceptable methods of aggregating cells, especially in the geographical context. One such method is proposed by Pitfield (1978). Pitfield, however, admits that there are difficulties of meaningfulness of the resulting statistic and that this limits the usefulness of the test. Empirical findings by Hathaway (1975, 84) and Openshaw (1976, 30) have revealed other drawbacks (e.g. scale) and inconsistencies in the use of the statistic. For these reasons it was decided not to use the chi-squared statistic in this study.

The following statistics have been used to compare each pair of observed and predicted matrices in this analysis. In the following, p^*_{ij} is an element of the predicted matrix, p_{ij} is an element of the observed matrix and n is the number of cells in the matrix. Note that in the case of the population distribution the summation is over one subscript.

(1) The coefficient of determination:

$$R^2 = 1 - \frac{\sum_i \sum_j (p_{ij} - p^*_{ij})^2}{\sum_i \sum_j (p_{ij} - \frac{1}{n} \sum_i \sum_j p_{ij})^2} .$$

(2) The root mean square error:

$$\sqrt{\frac{1}{n} \sum_i \sum_j (p_{ij} - p^*_{ij})^2} .$$

(3) The mean absolute error:

$$\frac{1}{n} \sum_i \sum_j |p_{ij} - p^*_{ij}|.$$

(4) The total absolute error:

$$\sum_i \sum_j |p_{ij} - p^*_{ij}|.$$

(5) The phi statistic:

$$\sum_i \sum_j p_{ij} \left| \ell n \frac{p^*_{ij}}{p_{ij}} \right|$$

The coefficient of determination is the usual measure of departure between observed and predicted matrices. However, its use in spatial interaction studies has been criticised given that inconsistencies may arise in its application out of the linear regression context (Senior, 1977). Owing to these conditions associated with regression analysis the R^2 measure has been supplemented by more straight forward measures of absolute errors between observations and predictions and by the phi statistic.

The results of model testing are given in Tables 5.3, 5.4 and 5.5. The first column of each table gives the goodness of fit values for predictions from CALBTN2.1 (treating the sample data as an homogeneous group). Columns 2 to 5 correspond with CALBTN2.2 predictions for the four income groups. The sixth column gives the results of comparing the sum of the observed and the sum of the predicted matrices for the four income groups (when summing the income-specific matrices, each separate observed and predicted matrix was multiplied by the proportion of the total sample households in that income group then normalized to sum to 1). The remaining discussion will focus upon a comparison of the observed/predicted residential location, work trip and shop trip distribution matrices.

(1) Population distributions

The goodness of fit results given in Table 5.3 indicate the following ranking (from best to worst): sum of income groups 1 to 4, aggregate, group 3, group 2, group 4, group 1. The summed income group results marginally outperform the results obtained from the aggregate model version. The predictions for the sum of income groups 1 to 4 are clearly the best on all measures except root mean square error. The remainder of the population predictions show a consistent level of agreement with the observed population distribution, except for income group one. The goodness of fit results for this income group reveal a pronounced lack of correspondence between observed and predicted distributions. This finding reveals the weakness with respect to observed population data which was alluded to in the previous chapter; i.e. failing to obtain a satisfactory published

Table 5.3 Comparison Between Observed and Predicted Population Distributions

Statistic	Aggregate Sample	Income Groups 1	2	3	4	Sum of 1 to 4
R^2	.67195	.18881	.65816	.66578	.58277	.73302
R.M.S.E.	.00009	.00394	.00060	.00050	.00073	.00040
M.A.E.	.02010	.04791	.02020	.02006	.02409	.01532
T.A.E.	.31475	.81450	.34342	.34098	.40958	.26039
PHI	.31919	1.53037	.36093	.33018	.41132	.27724

Table 5.4 Comparison Between Observed and Predicted Shop Trip Distributions

Statistic	Aggregate Sample	Income Groups 1	2	3	4	Sum of 1 to 4
R^2	.28175	.10678	.32022	.21750	.09729	.27080
R.M.S.E.	.00019	.00021	.00009	.00010	.00019	.00010
M.A.E.	.00534	.00507	.00406	.00437	.00507	.00412
T.A.E.	1.27899	1.46478	1.17328	1.26230	1.46445	1.18940
PHI	1.35119	2.12215	1.22597	1.32939	1.72458	1.30036

Table 5.5 Comparison Between Observed and Predicted Work Trip Distributions

Statistic	Aggregate Sample	Income Groups 1	2	3	4	Sum of 1 to 4
R^2	.66410	.09202	.60663	.6639	.29894	.62909
R.M.S.E.	.00002	.00032	.00002	.00002	.00009	.00002
M.A.E.	.00191	.00412	.00198	.00229	.00327	.00193
T.A.E.	.60176	1.39915	.67313	.77741	1.11174	.65573
PHI	.67478	5.09622	.79687	.79389	1.25344	.76387

Note: R.M.S.E. Root Mean Square Error
M.A.E. Mean Absolute Error
T.A.E. Total Absolute Error

source of information for the observed distribution of population in each income category it was decided that the observed total zonal population distribution provided the most accurate and unbiased estimate available. This observed total population distribution (obtained from several RMHWPDD sources) has been used in goodness of fit comparisons for all income group predictions and for the aggregate predictions. Consequently, little may be inferred from the results given in Table 5.3 about the accuracy of model predictions for separate income groups. A more appropriate focus here is upon the results obtained from the aggregate model version and the sum of the separately synthesized income group predictions.

Figures 5.1 and 5.2 display the specific differences in zonal population predictions between the four income groups, the aggregate and the summed income group predictions. The Figures are based upon data contained in Appendix 4.1. The observed zonal population proportions shown in both figures are based upon the total population distribution, as mentioned above. Figure 5.1 reveals that fluctuations between observed and predicted population distributions are most pronounced for the lowest income group. Predictions for income groups 2, 3 and 4 are fairly consistent. Figure 5.2 shows that the results of summing the four income groups compare very closely with the aggregate predictions. Both are closely correlated with the actual zonal distribution of population. The overall results of Figures 5.1 and 5.2 indicate that there is some regularity in observed/predicted discrepancies

Figure 5.1 Observed and Expected Population Distributions for the Four Income Groups

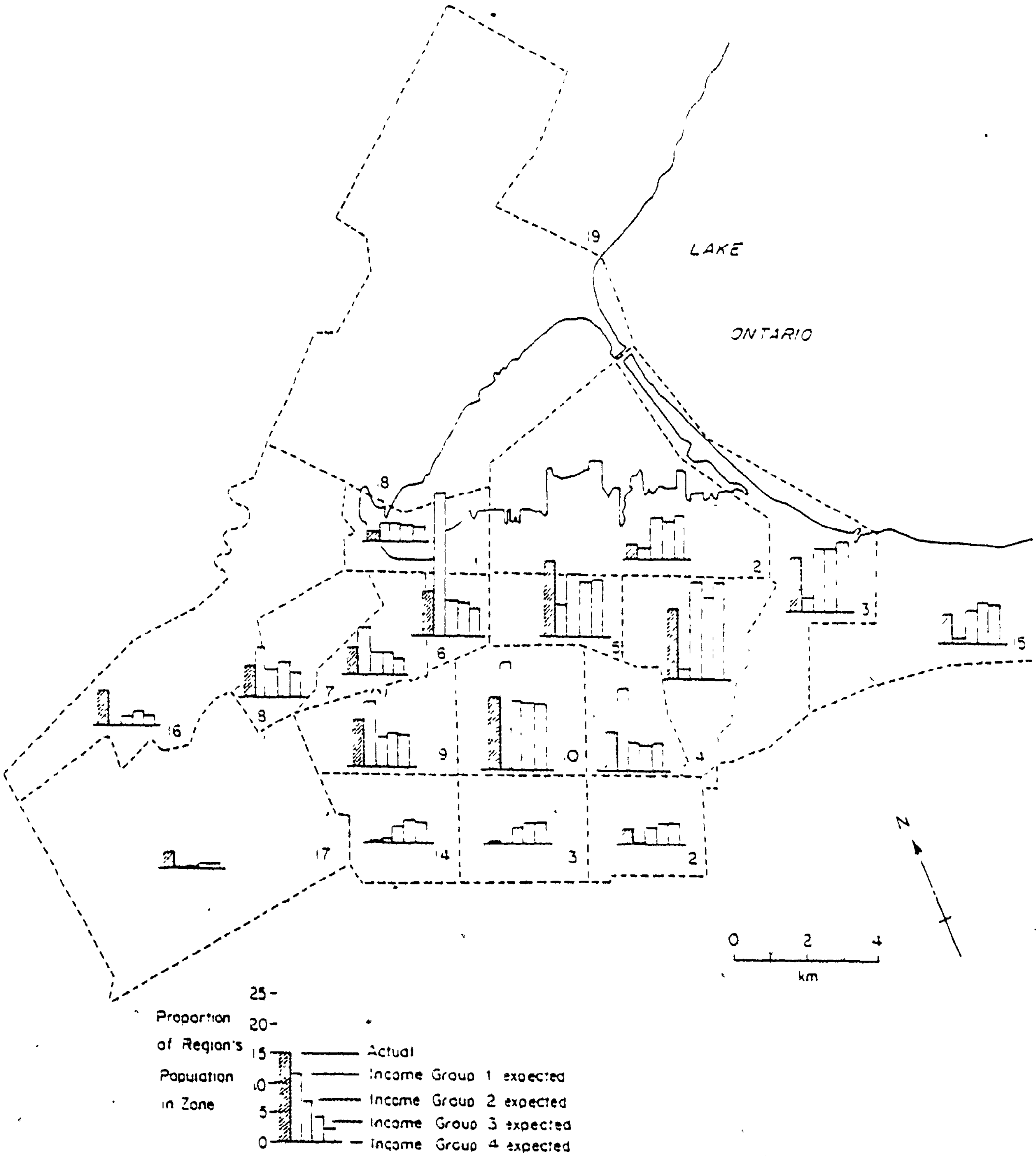
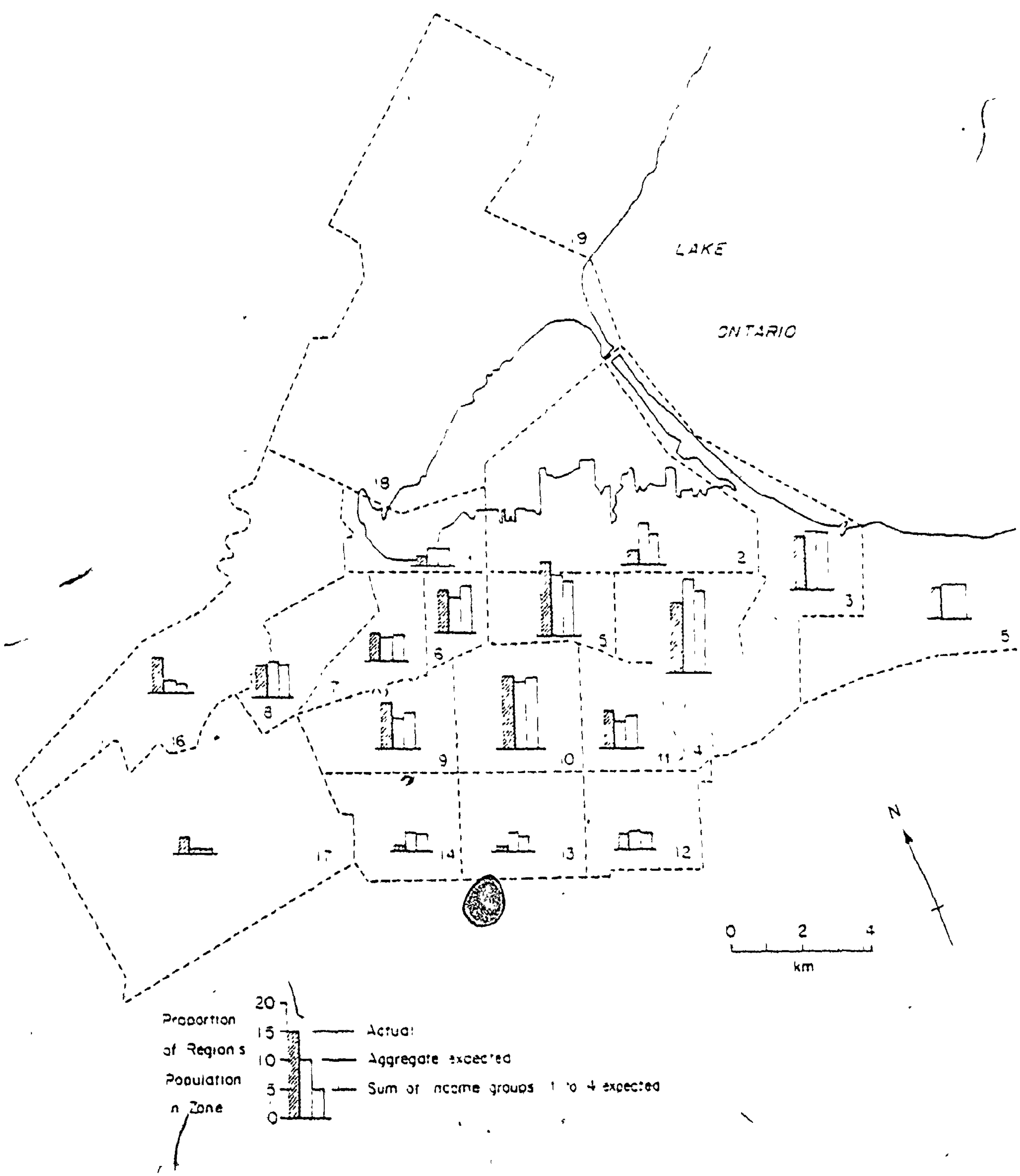


Figure 5.2 Observed and Expected Population Distributions for the Aggregate Sample and for the Sum of Income Groups 1 to 4.



shared by all the predictions. The centrally located zones (1,2,3 and 4) and zones 13 and 14 (Hamilton "Mountain") tend to be overpredicted in each case whereas zones 16 and 17 (Dundas, Ancaster) are consistently underpredicted. The underpredictions in Dundas and Ancaster may be a result of modelling these communities as part of the Hamilton urban area when, in fact, they have developed as separate communities on the periphery of Hamilton.

(2) Shop trip distributions

In general shop trip interchange predictions are the least accurate component of the model. This has also been the case for previous model versions. Possible explanations for the continuing existence of this problem will be presented in the final section of this chapter.

The goodness of fit results shown in Table 5.4 reveal that income group 2 provides the closest fit between actual and expected shop trip distribution based upon all statistical metrics (recall that the actual or observed shop trip matrix is population weighted according to the method outlined in section 4.5 of the previous chapter). The remaining predictions are difficult to rank because of the lack of agreement between the values of the various statistics, however the majority of the goodness of fit values indicate that predictions for income groups 1 and 4 fare the worst. It can be seen that a considerable gap exists, in terms of closeness of fit, between the results for (income groups 2, 3, the sum of income groups 1 to 4, the aggregate sample) and the results for (income categories 1 and 4). The relative performance of model predictions for shop trips may be ranked (from best to worst) as follows: income group 2, sum of income groups 1 to 4, income group 3, aggregate, income group 4, income group 1.

The distribution of the twenty shopping facilities is based upon observed patronage for the total sample. The correspondence between the shopping facility allocation and the predicted shop trip interchange destinations (column totals) may be seen by comparing Table 4.2 of the previous chapter with the shop trip distribution matrices output from the model (Appendix 4.2). One of the difficulties in assigning shop trips to a discrete set of service centres will become apparent upon examination of Figures 5.3 and 5.4, and the matrices in Appendix 4.2. Although the distribution of shopping centres captures most of the observed patronage, inevitably there are residuals present in the predicted matrices. Most zones which are not assigned facilities do display a small residue of shop trips made to them. The total predicted patronage (column totals) for most zones closely approximates the observed patronage, however, in the majority of cases the predicted shop trip origin zones are too evenly spread out as compared with the observed origin zones. An example of this may be seen in Appendix 4.2 (income group 2) for shop trip destination zones

Figure 5.3 Observed and Predicted Shop Trip Distributions for the Four Income Groups.

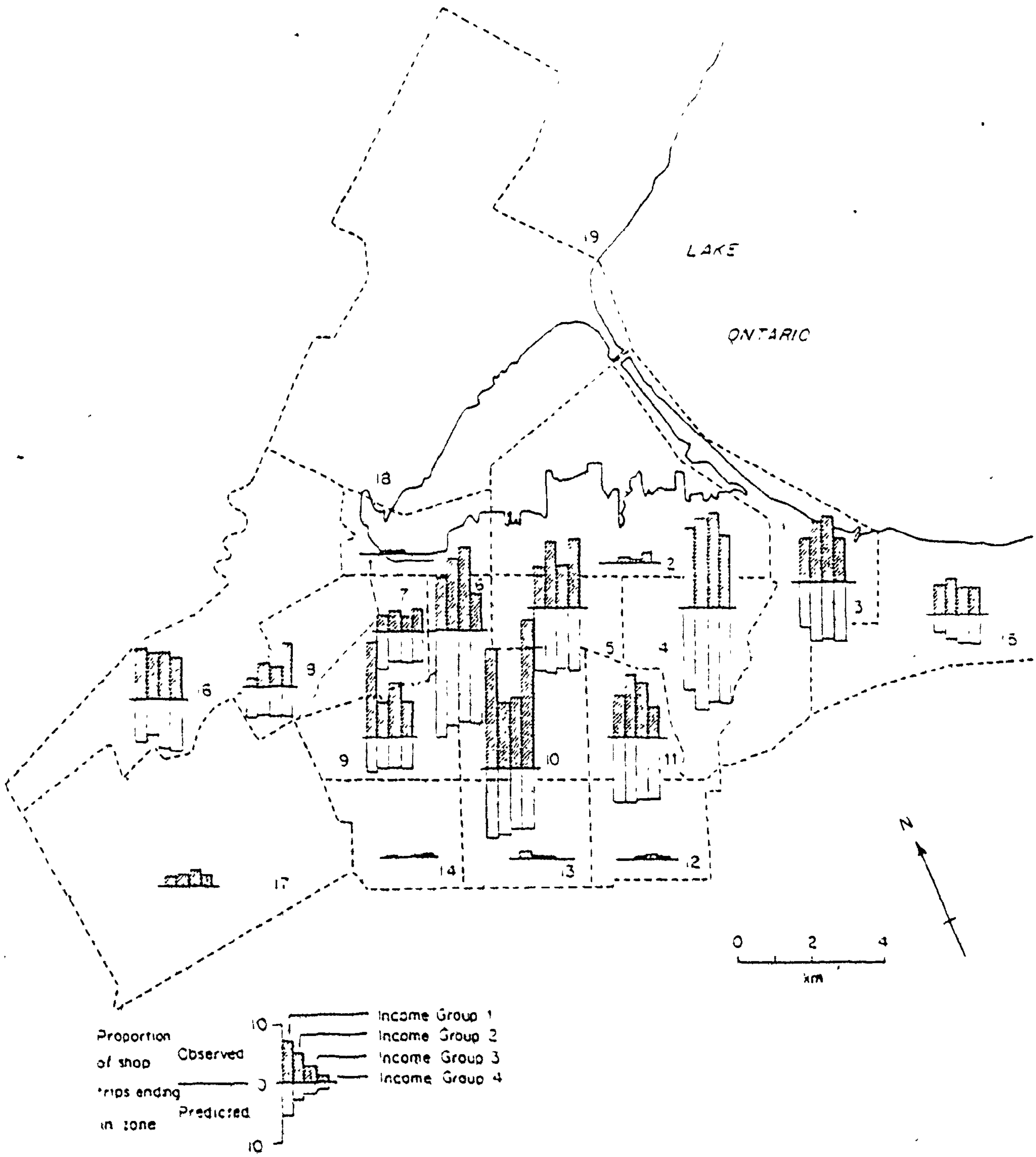
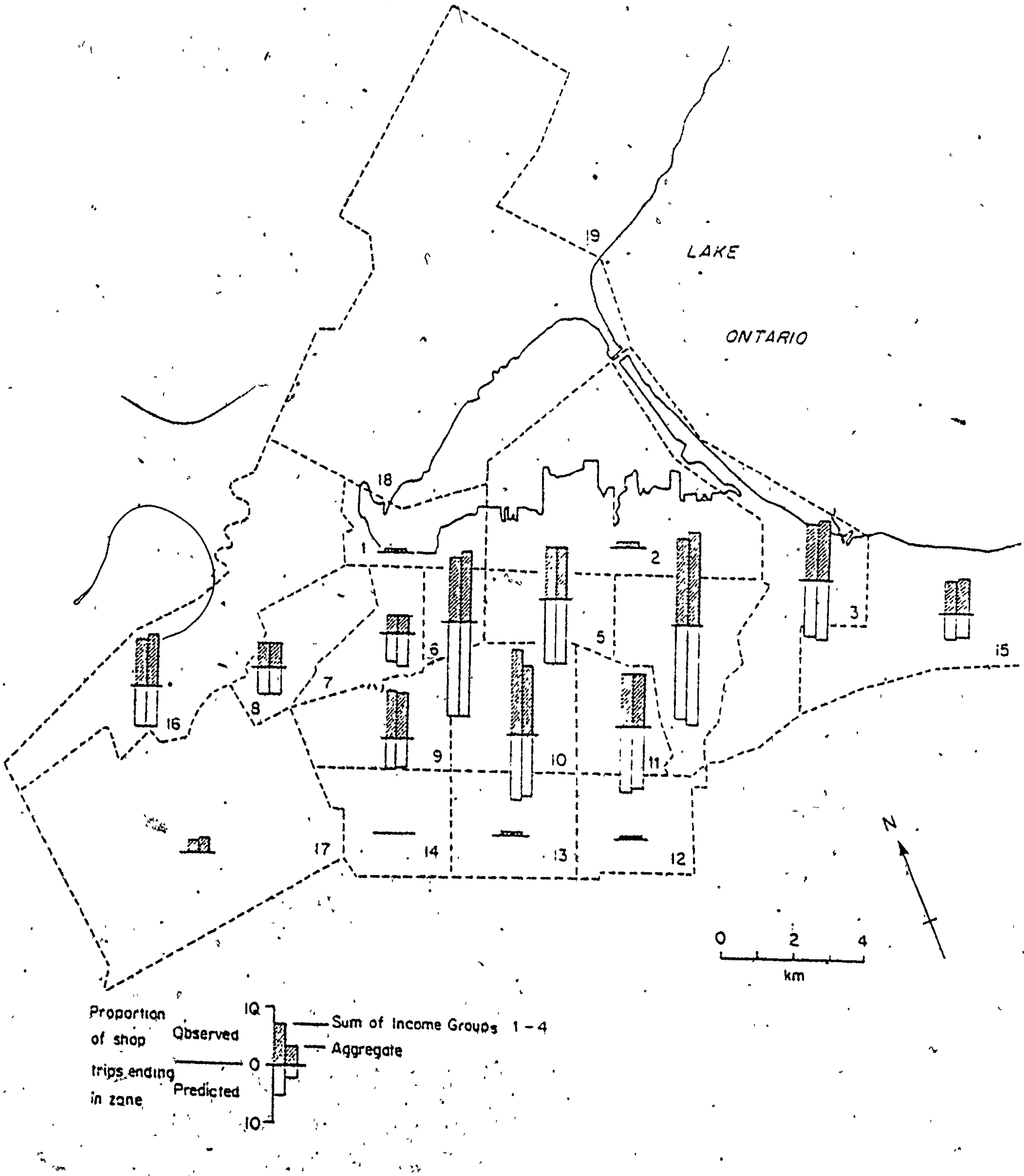


Figure 5.4 Observed and Predicted Shop Trip Distributions for the Aggregate Sample and for the Sum of Income Groups 1 to 4.



(columns) 3, 4, 5, 7, 15, and 16. This is probably an effect induced by the uniform shopping pattern prior.

(3) Work trip distributions

A comparison of the goodness of fit of observed and predicted work trip distributions is shown in Table 5.5 (note that the observed work trip matrix is population weighted - see section 4.5 of the previous chapter). The concensus of statistical evidence reveals that the aggregate model version provides the best results, followed by the sum of income groups 1 to 4, income group 3 and income group 2. Again, the outcomes for income categories 1 and 4 indicate the poorest fits.

The observed and predicted origin - destination work trip matrices are illustrated in Tables 5.6 to 5.11. Recall that model predictions are constrained to match the observed work trip destination end arrays, d_k and d_k^v, hence column totals for observed and predicted matrices are the same. A close look at these tables indicates that the predicted work trip matrices suffer the same difficulty experienced by the predicted shop trip matrices; i.e. work trip origins are distributed too equitably by the model. For example, in the case of work trip distributions for income group 2 (Table 5.7) some destination zones (eg. 9, 10, 16) have work trips which are observed to originate from a few zones, whereas the model predicts smaller numbers of work trips originating from many zones. This problem probably could have been lessened had the work trip prior probability, $\hat{w}_k$, been other than uniform. To accomplish this, however, one would require access to some reliable source of information regarding work trips that was not available at the time of this analysis (perhaps an income group specific origin-destination matrix from some previous time period).

Table 5.6 Income Group 1 Work Trip Matrices

ACTUAL WORK
TRIP INTERCHANG MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.017
2	0.000	0.000	0.000	0.000	0.000	0.000	.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.017
3	0.000	0.000	.038	0.000	0.000	0.000	.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.038
4	0.000	0.000	0.000	0.000	.015	.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.093
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.026	.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.091
6	.007	.007	0.000	0.000	0.000	.007	0.000	.014	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.042
7	0.000	0.000	0.000	0.000	0.000	.012	.006	.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.024
8	0.000	0.000	0.000	0.000	0.000	.006	0.000	.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.006	.011
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.082	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.123	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	.022	0.000	0.000	0.000	.022	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.022
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.007
14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.055
16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.062
17	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.006	0.000	0.000	0.000	0.000	0.000	.073
TOTAL	.007	.029	.038	0.000	.015	.062	.033	.079	.095	0.000	.124	0.000	0.000	.006	0.000	0.000	0.000	0.000	.006	.502

PREDICTED WORK
TRIP INTERCHANG MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	.007	.000	.000	0.000	.000	.011	.000	.000	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.001	.010
2	.000	.019	.000	0.000	.000	.000	.000	.000	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.002
3	.000	.000	.024	0.000	.000	.000	.000	.000	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.000
4	.000	.010	.003	0.000	.000	.000	.000	.000	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.005
5	.000	.000	.000	0.000	.015	.000	.000	.000	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.037
6	.000	.000	.000	0.000	.000	.048	.009	.000	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.184
7	.000	.000	.000	0.000	.000	.002	.023	.001	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.003	.042
8	.000	.000	.000	0.000	.000	.000	.002	.078	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.002	.003
9	.000	.000	.000	0.000	.000	.000	.000	.000	.092	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.028
10	.000	.000	.000	0.000	.000	.000	.000	.000	.001	0.000	.005	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.173
11	.000	.000	.000	0.000	.000	.000	.000	.000	.000	0.000	.123	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.014
12	.000	.000	.000	0.000	.000	.000	.000	.000	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.002
13	.000	.000	.000	0.000	.000	.000	.000	.000	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.001
14	.000	.000	.000	0.000	.000	.010	.000	.000	.002	0.000	.000	0.000	0.000	.006	0.000	0.000	0.000	0.000	.000	.001
15	.000	.000	.011	0.000	.000	.000	.000	.000	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.000
16	.000	.000	.000	0.000	.000	.000	.000	.000	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.000
17	.000	.000	.000	0.000	.000	.000	.000	.000	.000	0.000	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.000
TOTAL	.007	.029	.038	0.000	.015	.062	.033	.079	.095	0.000	.128	0.000	0.000	.006	0.000	0.000	0.000	0.000	.006	.502

Table 5.7 Income Group 2 Work Trip Matrices

ACTUAL WORK
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	.001	.006	0.000	0.000	0.000	.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.005
2	0.000	.011	.002	.002	.002	0.000	.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.002	0.000	.002
3	0.000	.022	.011	.005	.005	.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.005	0.000	0.000	0.000	.011	.027
4	0.000	.040	.010	.015	.005	.005	0.000	0.000	0.000	.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.005	0.000	.035
5	0.000	.027	0.000	.009	.009	.027	.009	.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.036
6	0.000	.012	0.000	0.000	.004	.012	.008	.004	.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.004	0.000	0.000	.025
7	0.000	.008	0.000	0.000	.006	.011	.006	0.000	.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.003	0.000	.011
8	0.000	0.000	.005	.005	0.000	.005	.010	.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.005	0.000	.019
9	.003	.030	0.000	0.000	0.000	.003	.003	.003	.003	.010	0.000	0.000	0.000	0.000	.003	0.000	0.000	0.000	0.000	.020
10	0.000	.054	0.000	.006	0.000	.006	.006	0.000	0.000	.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.035
11	0.000	.023	.003	.003	.003	.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.016
12	0.000	.005	0.000	0.000	.003	.003	0.000	.002	0.000	0.000	0.000	0.000	0.000	0.000	.002	0.000	.002	.002	0.000	.008
13	0.000	.001	0.000	0.000	.001	.000	.000	0.000	0.000	.000	.000	0.000	.000	0.000	0.000	0.000	0.000	.000	.000	.002
14	0.000	.001	0.000	.000	.000	0.000	.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.000	0.000	.003
15	0.000	.023	0.000	.003	.007	0.000	.003	.003	0.000	0.000	0.000	0.000	0.000	0.000	.003	0.000	0.000	0.000	0.000	.010
16	0.000	.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.015	0.000	0.000	0.000	.020
17	0.000	.020	0.000	0.000	0.000	.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.004
TOTAL	.006	.313	.031	.049	.046	.039	.044	.026	.010	.027	.000	0.000	.000	0.000	.014	.015	.006	.047	.011	.279

PREDICTED WORK
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	.001	.008	.000	.001	.001	.005	.002	.001	.000	.000	0.000	0.000	0.000	0.000	.000	.000	.000	.001	.001	.009
2	.000	.034	.002	.003	.003	.004	.001	.001	.000	.000	0.000	0.000	0.000	0.000	.001	.000	.000	.001	.001	.020
3	.000	.048	.003	.008	.003	.003	.001	.000	.000	.001	0.000	0.000	0.000	0.000	.004	.000	.000	.001	.001	.028
4	.001	.078	.007	.014	.007	.006	.002	.001	.000	.001	0.000	0.000	0.000	0.000	.002	.000	.000	.001	.001	.046
5	.001	.035	.001	.004	.009	.011	.003	.001	.001	.002	0.000	0.000	0.000	0.000	.000	.000	.000	.001	.001	.033
6	.001	.010	.000	.001	.003	.014	.006	.002	.001	.001	0.000	0.000	0.000	0.000	.000	.001	.000	.002	.001	.017
7	.000	.006	.000	.001	.001	.007	.005	.002	.000	.001	0.000	0.000	0.000	0.000	.000	.001	.000	.002	.001	.010
8	.000	.006	.000	.001	.001	.006	.006	.007	.000	.001	0.000	0.000	0.000	0.000	.000	.004	.001	.003	.002	.011
9	.000	.008	.000	.001	.003	.008	.005	.002	.002	.003	0.000	0.000	0.000	0.000	.000	.001	.001	.001	.001	.014
10	.001	.025	.002	.005	.006	.016	.007	.002	.002	.009	0.000	0.000	0.000	0.000	.001	.001	.011	.002	.001	.036
11	.000	.014	.001	.003	.002	.003	.002	.001	.000	.002	0.000	0.000	0.000	0.000	.000	.000	.000	.000	.000	.014
12	.000	.006	.000	.001	.001	.003	.002	.001	.001	.002	0.000	0.000	0.000	0.000	.000	.000	.000	.000	.000	.009
13	.000	.006	.001	.001	.001	.004	.002	.001	.001	.002	0.000	0.000	0.000	0.000	.000	.000	.000	.000	.000	.009
14	.000	.004	.000	.001	.001	.004	.002	.001	.001	.002	0.000	0.000	0.000	0.000	.000	.000	.001	.000	.000	.008
15	.000	.025	.006	.003	.001	.001	.000	.000	.000	.000	0.000	0.000	0.000	0.000	.005	.000	.000	.001	.001	.013
16	.000	.002	.000	.000	.000	.002	.002	.003	.000	.000	0.000	0.000	0.000	0.000	.000	.005	.030	.001	.000	.003
17	.000	.000	.000	.000	.000	.000	.000	.001	.000	.000	0.000	0.000	0.000	0.000	.000	.000	.011	.000	.000	.001
TOTAL	.006	.314	.031	.049	.046	.039	.048	.026	.010	.027	0.000	0.000	0.000	0.000	.014	.015	.006	.017	.011	.279

Table 5.8 Income Group 3 Work Trip Matrices

ACTUAL WORK
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.000	.006	0.000	0.000	0.000	.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.006
2	0.000	.011	0.000	.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.011
3	0.000	.054	.008	0.000	0.000	.008	0.000	0.000	0.000	.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.015
4	.010	.040	0.000	0.000	0.000	.010	0.000	0.000	0.000	.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.010
5	0.000	.016	0.000	0.000	.016	.031	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.016	0.000	.047
6	0.000	.019	0.000	.019	0.000	.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.019
7	.007	.013	0.000	0.000	0.000	0.000	0.000	.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.013
8	0.000	.004	0.000	0.000	.004	.012	.004	.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.004	0.000	0.000	.008
9	0.000	.033	0.000	.007	.013	0.000	0.000	.007	.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.013
10	0.000	.044	0.000	0.000	.009	.018	.018	.009	0.000	0.000	0.000	0.000	0.000	0.000	.009	0.000	0.000	0.000	0.000	.018
11	0.000	.005	0.000	.005	.010	.010	.005	0.000	0.000	.010	0.000	0.000	0.000	0.000	.005	0.000	0.000	0.000	0.000	.010
12	.002	.011	0.000	0.000	0.000	.009	.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.002	0.000	0.000	0.000	0.000	0.000
13	0.000	.002	0.000	0.000	0.000	.001	.001	0.000	.001	.001	0.000	0.000	0.000	0.000	0.000	0.000	.001	0.000	0.000	.002
14	0.000	.001	.000	0.000	.001	.000	.001	.000	.000	.000	0.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.001	.002
15	0.000	.011	0.000	0.000	.004	.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.019	0.000	0.000	.004	.004	.008
16	0.000	.006	0.000	.003	0.000	.015	.003	.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.009	0.000	0.000	0.000	.012
17	0.000	.006	0.000	0.000	0.000	0.000	0.000	.004	.002	.002	0.000	0.000	0.000	0.000	0.000	0.000	.006	0.000	0.000	.009
TOTAL	.019	.171	.008	.036	.057	.142	.034	.062	.009	.031	0.000	0.000	0.000	.000	.035	.009	.010	.020	.004	.203

PREDICTED WORK
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	.002	.008	.000	.000	.001	.007	.001	.002	.000	.000	0.000	0.000	0.000	0.000	.000	.000	.000	.001	.000	.006
2	.001	.032	.000	.002	.004	.005	.001	.001	.000	.000	0.000	0.000	0.000	0.000	.001	.000	.000	.001	.000	.013
3	.001	.053	.002	.007	.004	.005	.001	.001	.000	.001	0.000	0.000	0.000	0.000	.010	.000	.000	.001	.000	.020
4	.002	.071	.001	.009	.008	.008	.001	.002	.000	.001	0.000	0.000	0.000	0.000	.004	.000	.000	.001	.000	.028
5	.002	.032	.000	.003	.010	.014	.002	.003	.000	.002	0.000	0.000	0.000	0.000	.001	.000	.000	.001	.000	.021
6	.002	.009	.000	.001	.003	.017	.003	.004	.000	.001	0.000	0.000	0.000	0.000	.000	.000	.000	.002	.000	.011
7	.001	.006	.000	.000	.002	.010	.003	.005	.000	.001	0.000	0.000	0.000	0.000	.000	.000	.000	.002	.000	.007
8	.002	.007	.000	.000	.002	.010	.004	.017	.000	.001	0.000	0.000	0.000	0.000	.000	.002	.001	.003	.001	.010
9	.001	.009	.000	.001	.004	.013	.004	.005	.002	.003	0.000	0.000	0.000	0.000	.000	.000	.002	.001	.000	.012
10	.002	.025	.000	.003	.007	.022	.005	.005	.002	.009	0.000	0.000	0.000	0.000	.001	.000	.001	.002	.000	.026
11	.000	.014	.000	.002	.003	.005	.001	.001	.000	.002	0.000	0.000	0.000	0.000	.001	.000	.000	.000	.000	.010
12	.001	.008	.000	.001	.002	.006	.002	.002	.001	.002	0.000	0.000	0.000	0.000	.000	.000	.001	.001	.000	.008
13	.001	.008	.000	.001	.002	.006	.001	.001	.001	.003	0.000	0.000	0.000	0.000	.001	.000	.001	.000	.000	.008
14	.001	.006	.000	.001	.002	.008	.002	.002	.001	.002	0.000	0.000	0.000	0.000	.000	.000	.002	.001	.000	.007
15	.000	.030	.002	.003	.002	.002	.000	.001	.000	.000	0.000	0.000	0.000	0.000	.014	.000	.000	.001	.000	.011
16	.001	.002	.000	.000	.001	.003	.001	.008	.000	.000	0.000	0.000	0.000	0.000	.000	.004	.001	.001	.000	.003
17	.000	.001	.000	.000	.000	.001	.000	.002	.000	.000	0.000	0.000	0.000	0.000	.000	.000	.002	.000	.000	.001
TOTAL	.013	.171	.008	.036	.057	.142	.034	.062	.009	.031	0.000	0.000	0.000	0.000	.035	.009	.010	.020	.004	.203

Table 5.9 Income Group 4 Work Trip Matrices

ACTUAL WORK
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.000	.008	0.000	0.000	.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.024
3	0.000	.015	0.000	0.000	.031	.015	0.000	0.000	0.000	.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.015
4	0.000	.040	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.020	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.060
5	0.000	.042	0.000	0.000	.042	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.042
6	0.000	.074	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	.008	0.000	0.000	0.000	.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.008	0.000	0.000	.024
8	0.000	0.000	0.000	0.000	0.000	0.000	.013	.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.027
9	0.000	.040	0.000	0.000	0.000	.020	0.000	0.000	0.000	.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.010	0.000
10	0.000	.124	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	.031	0.000	.010	0.000	0.000	0.000	0.000	0.000	0.000	.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.010
12	0.000	.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.007	0.000	0.000	.007	0.000	0.000	0.000	.007
13	0.000	.003	0.000	0.000	.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.003
14	0.000	.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	.021	0.000	.005	0.000	.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.005	0.000	0.000	0.000	0.000	.016
16	0.000	.012	0.000	0.000	0.000	.012	0.000	.012	.012	.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.006	0.000	0.000
17	0.000	.042	0.000	0.000	0.000	.007	0.000	.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.011
TOTAL	0.000	.426	0.000	.024	.082	.059	.013	.042	.012	.051	.010	0.000	.007	0.000	.005	.007	.008	.006	.010	.238

PREDICTED WORK
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.000	.010	0.000	.000	.002	.003	.000	.001	.000	.000	.000	0.000	.000	0.000	.000	.000	.000	.000	.000	.007
2	0.000	.044	0.000	.002	.005	.002	.000	.001	.000	.001	.000	0.000	.000	0.000	.000	.000	.000	.000	.000	.017
3	0.000	.072	0.000	.004	.006	.002	.000	.001	.000	.002	.001	0.000	.000	0.000	.001	.000	.000	.000	.001	.027
4	0.000	.097	0.000	.006	.012	.003	.000	.001	.000	.002	.001	0.000	.000	0.000	.001	.000	.000	.000	.001	.038
5	0.000	.042	0.000	.002	.015	.006	.001	.002	.001	.002	.000	0.000	.000	0.000	.000	.000	.000	.000	.001	.025
6	0.000	.012	0.000	.001	.005	.007	.001	.003	.000	.002	.000	0.000	.000	0.000	.000	.000	.000	.000	.001	.011
7	0.000	.007	0.000	.000	.002	.004	.001	.003	.000	.001	.000	0.000	.000	0.000	.000	.000	.000	.001	.001	.007
8	0.000	.009	0.000	.000	.002	.004	.002	.012	.000	.001	.000	0.000	.000	0.000	.000	.002	.001	.001	.002	.009
9	0.000	.011	0.000	.001	.005	.005	.002	.003	.003	.005	.001	0.000	.001	0.000	.000	.000	.001	.000	.001	.012
10	0.000	.032	0.000	.002	.010	.009	.002	.003	.002	.015	.002	0.000	.001	0.000	.000	.000	.001	.001	.001	.029
11	0.000	.019	0.000	.002	.004	.002	.000	.001	.000	.004	.002	0.000	.000	0.000	.000	.000	.000	.000	.000	.012
12	0.000	.010	0.000	.001	.003	.002	.001	.001	.001	.004	.001	0.000	.001	0.000	.000	.000	.001	.000	.000	.009
13	0.000	.010	0.000	.001	.003	.003	.000	.001	.001	.005	.001	0.000	.001	0.000	.000	.000	.001	.000	.000	.009
14	0.000	.007	0.000	.000	.003	.003	.001	.002	.002	.004	.000	0.000	.001	0.000	.000	.000	.001	.000	.000	.008
15	0.000	.040	0.000	.002	.003	.001	.000	.000	.000	.001	.000	0.000	.000	0.000	.002	.000	.000	.000	.001	.014
16	0.000	.003	0.000	.000	.001	.001	.001	.006	.000	.000	.000	0.000	.000	0.000	.000	.003	.001	.000	.001	.003
17	0.000	.001	0.000	.000	.000	.000	.000	.001	.000	.000	.000	0.000	.000	0.000	.000	.000	.002	.000	.000	.001
TOTAL	0.000	.426	0.000	.024	.082	.059	.013	.042	.012	.051	.010	0.000	.007	0.000	.005	.007	.008	.006	.010	.238

Table 5.10 Aggregate Work Trip Matrices

ACTUAL WORK
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	.001	.003	0.000	0.000	.000	.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.007
2	0.000	.005	.001	.001	.001	0.000	.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.001	0.000	.005
3	0.000	.028	.012	.002	.007	.007	.002	0.000	0.000	.005	0.000	0.000	0.000	0.000	.002	0.000	0.000	0.000	.005	.023
4	.003	.055	.006	.009	.006	.003	0.000	0.000	0.000	.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.003	0.000	.051
5	0.000	.016	0.000	.003	.010	.015	.003	.010	.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.003	0.000	.048
6	.002	.011	0.000	.002	.002	.003	.004	.006	.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.002	0.000	0.000	.024
7	.001	.006	0.000	.001	.002	.007	.004	.005	.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.011	.001	0.000	.016
8	0.000	.001	.001	.001	.001	.007	.005	.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.001	.001	.001	.014
9	.002	.036	0.000	.002	.004	.005	.002	.004	.006	.009	0.000	0.000	0.000	0.000	.002	0.000	0.000	0.000	.002	.016
10	0.000	.050	0.000	.003	.003	.003	.009	.003	0.000	.006	.003	0.000	0.000	0.000	.003	0.000	0.000	0.000	0.000	.025
11	0.000	.019	.002	.005	.005	.011	.002	0.000	0.000	.003	.002	0.000	0.000	0.000	.002	0.000	0.000	0.000	0.000	.014
12	.001	.006	0.000	0.000	.001	.004	.001	.001	0.000	0.000	0.000	0.000	.001	0.000	.001	.001	.011	.001	0.000	.004
13	0.000	.001	0.000	0.000	.000	.000	.000	0.000	.000	.003	.000	0.000	.000	0.000	0.000	0.000	.010	.000	.000	.002
14	0.000	.002	.000	.000	.001	.000	.001	.000	.000	.000	0.000	0.000	0.000	.000	0.000	0.000	0.000	.000	.000	.002
15	0.000	.019	0.000	.003	.004	.003	.001	.001	0.000	0.003	0.000	0.000	0.000	0.000	.009	0.000	0.000	.001	.001	.013
16	0.000	.014	0.000	.002	0.000	.011	.002	.009	.003	.002	0.000	0.000	0.000	0.000	0.000	.009	0.000	.002	0.000	.015
17	0.000	.006	0.000	0.000	0.000	.003	0.000	.004	.001	.001	0.000	0.000	0.000	.001	0.000	0.000	.012	0.000	0.000	.011
TOTAL	.010	.279	.021	.035	.048	.104	.037	.059	.016	.034	.005	0.000	.001	.001	.020	.010	.007	.043	.010	.291

PREDICTED WORK
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	.001	.007	.000	.000	.001	.005	.001	.002	.000	.000	.000	0.000	.000	.000	.000	.000	.000	.001	.000	.010
2	.001	.033	.001	.002	.003	.004	.001	.001	.000	.000	.000	0.000	.000	.000	.001	.000	.000	.001	.000	.021
3	.000	.044	.007	.006	.003	.003	.001	.001	.000	.001	.000	0.000	.000	.000	.005	.000	.000	.001	.001	.026
4	.001	.071	.004	.011	.007	.005	.001	.001	.000	.001	.001	0.000	.000	.000	.003	.000	.000	.001	.001	.044
5	.001	.031	.001	.003	.010	.012	.002	.002	.001	.002	.000	0.000	.000	.000	.001	.000	.000	.001	.001	.034
6	.001	.008	.000	.001	.003	.015	.005	.004	.001	.002	.000	0.000	.000	.000	.000	.000	.000	.001	.001	.018
7	.001	.004	.000	.000	.001	.003	.004	.005	.001	.001	.000	0.000	.000	.000	.000	.001	.000	.001	.001	.011
8	.001	.005	.000	.000	.001	.007	.005	.013	.000	.001	.000	0.000	.000	.000	.000	.002	.001	.002	.002	.013
9	.001	.006	.000	.001	.003	.003	.004	.004	.004	.003	.000	0.000	.000	.000	.000	.000	.001	.001	.001	.016
10	.001	.020	.001	.003	.006	.017	.005	.005	.003	.011	.001	0.000	.000	.000	.001	.000	.001	.001	.001	.039
11	.000	.012	.001	.002	.002	.003	.001	.001	.001	.003	.001	0.000	.000	.000	.000	.000	.000	.000	.000	.014
12	.000	.005	.000	.001	.002	.004	.001	.001	.001	.003	.000	0.000	.000	.000	.000	.000	.001	.000	.000	.010
13	.000	.005	.000	.001	.002	.004	.001	.001	.001	.003	.000	0.000	.000	.000	.000	.000	.000	.000	.000	.010
14	.000	.003	.000	.000	.002	.004	.002	.002	.002	.002	.000	0.000	.000	.000	.000	.000	.001	.000	.000	.009
15	.000	.022	.005	.003	.001	.001	.000	.000	.000	.000	.000	0.000	.000	.000	.008	.000	.000	.001	.001	.012
16	.000	.001	.000	.000	.000	.002	.001	.007	.000	.000	.000	0.000	.000	.000	.000	.004	.000	.001	.000	.003
17	.000	.000	.000	.000	.000	.000	.000	.001	.000	.000	.000	0.000	.000	.000	.000	.000	.001	.000	.000	.001
TOTAL	.010	.279	.021	.035	.048	.104	.037	.059	.016	.034	.005	0.000	.001	.001	.020	.010	.007	.013	.010	.291

Table 5.11 Sum of Income Groups 1 to 4 Work Trip Matrices

ACTUAL WORK
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	.001	.005	0.000	0.000	.001	.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.006
2	0.000	.008	.001	.002	.001	0.000	.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.001	0.000	.010
3	0.000	.028	.012	.002	.007	.007	.002	0.000	0.000	.005	0.000	0.000	0.000	0.000	.002	0.000	0.000	0.000	.005	.023
4	.003	.047	.004	.006	.004	.007	0.000	0.000	0.000	.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.002	0.000	.038
5	0.000	.022	0.000	.004	.015	.021	.004	.007	.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.005	0.000	.047
6	.001	.022	0.000	.006	.002	.012	.003	.004	.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.012	0.000	0.000	.021
7	.002	.008	0.000	.001	.002	.006	.003	.006	.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.001	.001	0.000	.015
8	0.000	.001	.002	.002	.001	.007	.007	.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.001	.002	.001	.015
9	.001	.023	0.000	.002	.004	.004	.001	.003	.014	.006	0.000	0.000	0.000	0.000	.001	0.000	0.010	0.000	.001	.012
10	0.000	.056	0.000	.002	.003	.008	.008	.003	0.000	.005	.016	0.000	0.000	0.000	.003	0.000	0.000	0.000	0.000	.020
11	0.000	.018	.001	.004	.005	.012	.002	0.000	0.000	.003	.001	0.000	0.000	0.000	.002	0.000	0.000	0.000	0.000	.014
12	.001	.006	0.000	0.000	.001	.004	.001	.001	0.000	0.000	0.000	0.000	.001	0.000	.001	.001	.001	.001	0.000	.004
13	0.000	.001	0.000	0.000	.000	.000	.000	0.000	.000	.000	.000	0.000	.000	0.000	0.000	0.000	.000	.000	.000	.003
14	0.000	.002	.000	.000	.000	.000	.000	.000	.000	.000	0.000	0.000	0.000	.000	0.000	0.000	0.000	.000	.000	.002
15	0.000	.016	0.000	.002	.004	.002	.001	.001	0.000	0.000	0.000	0.000	0.000	0.000	.008	0.000	0.000	.001	.001	.016
16	0.000	.014	0.000	.001	0.000	.006	.001	.005	.002	.001	0.000	0.000	0.000	0.000	0.000	.009	0.000	.001	0.000	.020
17	0.000	.011	0.000	0.000	0.000	.003	0.000	.002	.001	.001	0.000	0.000	0.000	.001	0.000	0.000	.002	0.000	0.000	.009
TOTAL	.009	.295	.020	.035	.050	.102	.037	.046	.021	.028	.018	0.000	.001	.001	.017	.010	.007	.014	.008	.278

PREDICTED WORK
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	.002	.007	.000	.000	.001	.006	.001	.001	.000	.000	.000	0.000	.000	.000	.000	.000	.000	.001	.000	.008
2	.001	.033	.001	.002	.003	.004	.001	.001	.000	.000	.000	0.000	.000	.000	.001	.000	.000	.001	.000	.015
3	.000	.047	.008	.006	.004	.003	.001	.001	.000	.001	.000	0.000	.000	.000	.005	.000	.000	.001	.001	.022
4	.001	.070	.004	.010	.007	.006	.001	.001	.000	.001	.000	0.000	.000	.000	.002	.000	.000	.001	.001	.034
5	.001	.031	.001	.003	.011	.010	.002	.002	.000	.001	.000	0.000	.000	.000	.001	.000	.000	.001	.000	.029
6	.001	.009	.000	.001	.003	.018	.005	.002	.000	.001	.000	0.000	.000	.000	.000	.000	.000	.001	.001	.036
7	.001	.005	.000	.000	.001	.007	.006	.003	.000	.001	.000	0.000	.000	.000	.000	.001	.000	.001	.001	.014
8	.001	.006	.000	.000	.001	.006	.004	.020	.000	.001	.000	0.000	.000	.000	.000	.002	.001	.002	.001	.009
9	.001	.008	.000	.001	.003	.008	.004	.003	.014	.003	.000	0.000	.000	.000	.000	.001	.001	.001	.000	.014
10	.001	.023	.001	.003	.006	.015	.005	.003	.002	.009	.001	0.000	.000	.000	.001	.001	.011	.001	.001	.049
11	.000	.013	.001	.002	.002	.003	.001	.001	.000	.002	.015	0.000	.000	.000	.000	.000	.000	.000	.000	.013
12	.000	.006	.000	.001	.002	.004	.001	.001	.001	.002	.000	0.000	.000	.000	.000	.000	.001	.000	.000	.008
13	.000	.007	.000	.001	.002	.004	.001	.001	.001	.001	.000	0.000	.000	.000	.000	.000	.060	.000	.000	.008
14	.000	.005	.000	.001	.002	.005	.002	.001	.001	.002	.000	0.000	.000	.001	.000	.000	.001	.000	.000	.007
15	.000	.025	.005	.003	.002	.001	.000	.000	.000	.000	.000	0.000	.000	.000	.007	.000	.000	.001	.000	.011
16	.000	.002	.000	.000	.000	.002	.001	.005	.000	.000	.000	0.000	.000	.000	.000	.004	.000	.001	.000	.003
17	.000	.001	.000	.000	.000	.001	.000	.001	.000	.000	.000	0.000	.000	.000	.000	.000	.001	.000	.000	.001
TOTAL	.003	.296	.020	.035	.051	.102	.037	.046	.021	.023	.018	0.000	.001	.001	.017	.010	.007	.014	.008	.278

5.4 Final Results of the Analysis

A preliminary analysis has examined the significance of the differences in mean cost estimates. The results confirm the premise that significant differences in work trip characteristics exist between income groups. Both model versions, CALBTN2.1 and CALBTN2.2, have then been calibrated. Population, work trip and shop trip distribution matrices have been synthesized for four income categroies and have been compared to the corresponding observed matrices using the different measures of comparison described in the previous section. The results are given in Table 5.3 to 5.5. These results show that the various goodness of fit measures are somewhat crude in that they fail to reach agreement on the relative degree of accuracy of model predictions in every case. Although the goodness of fit tests have revealed a lack of discriminatory ability, they have served to distinguish the accuracy of model predictions to some extent.

In general the results of income group disaggregation are less than encouraging even if the possible errors due to small sample sizes and other factors are taken into account. The lowest and highest income groups appear to be most susceptible to sample size errors as evidenced by the variation in mean cost values and the poorness of fit between synthesized and observed matrices. Due to the inherent middle income group bias of the sample data (73% of sample households fall into the two middle income groups) this result is to be expected. In terms of goodness of fit the results for income groups 2 and 3 have demonstrated the best performance as compared with the results for categories 1 and 4. For population and shop trip predictions, the sum of income groups 1 to 4 has shown closer fit to the observed data than the aggregate model predictions (treating the entire sample as a homogeneous group). For work trip predictions, the aggregate model has produced marginally better results than those obtained by summing the predictions for income groups 1 to 4. A byproduct of income disaggregation is the prediction of specific income group behaviour. Since actual data were not available for this purpose, no analysis of specific income groups has been performed.

The conclusive test of the disaggregate model was to examine the fit of the sum of the predicted income-specific matrices to the aggregate observed probability matrices. Each set of separately synthesized category matrices was summed and the resultant matrices were compared to the aggregate observed matrices. The results of these goodness of fit comparisons are given in the last columns of Tables 5.3, 5.4 and 5.5. These results indicate that there is little or no advantage to be gained from summing the separately calibrated income-specific matrices in order to describe and reproduce the overall locational and trip distribution patterns of the entire survey. For this purpose the aggregate model version, CALBTN2.1, produces the most satisfactory results especially when considering the trade-off between the cost of using the disaggregated model (data estimation requirements) and the gain in predictive accuracy.

It should be mentioned here that the results obtained from the aggregate model version, CALBTN2.1, represent improvements over the results of previous aggregate model versions. The most recent predecessor to CALBTN2.1 produced the following results.

	Population Distribution	Shop Trip Distribution	Work Trip Distribution
Coefficient of Determination (R^2)	.654	.138	.649

A comparison of the goodness of fit performance of CALBTN2.1 with the above figures shows that the current aggregate model has produced a 6% improvement in population predictions, a 112% improvement in shop trip predictions, and a 7% improvement in work trip predictions. Evidently the model has accrued benefits from the incorporation of the improvements outlined in Chapter Three; most notably, the inclusion of more shopping facilities and a greater number of household and trip type categories.

Although the aggregate shop trip predictions show the highest percentage improvement over previous aggregate model versions they nonetheless remain the weakest aspect of model predictions for both CALBTN2.1 and CALBTN2.2. This problem may stem from several sources. Consideration must be given to the difficulties in assigning twenty shopping facilities in 'chunks' when the observed patronage is more continuous. Inevitably residuals will be present. The assignment of one shopping unit for every $\frac{1}{20}$ observed demand is unrealistic since the demand for shop facilities can not be so neatly broken down into discrete units. However, the provision of twenty centres is certainly more realistic than the allocation of two or five facilities, as was the case in previous model versions.

Another possible source of error relates to the definition of shop trips, as given in the previous chapter. Experience with the survey data indicates that there is considerably more complex, multi-purpose, multi-stop trip making behaviour for shopping trips than for relatively single purpose, one-stop work trips. The definition of a shop trip is restricted to a sequence of stops of which the first stop must be for shopping purposes. It is apparent that the model may fail to capture a significant portion of the shop trip making behaviour of a sample household simply because the purpose of the first stop of a trip was other than shopping. In contrast, the results of model testing would indicate that on most work trips the first stop is the only stop and is for working purposes.

The results of the disaggregate model are immediately apparent in the different goodness of fit results produced for different income groups. These findings reveal that the income group disaggregation renders CALBTN2.2 less robust to the effects of sampling errors and small sample sizes. As Openshaw (1976, 35) has shown, the effect of disaggregation is similar to that of reducing the sample size and thus increasing the relative importance of sampling errors. This problem may be magnified in the disaggregate model version tested here by its use of population weighting factors. In an empirical study conducted by Hall (1975, 96) it was found that the R^2 measure was adversely affected by the use of population weighted scaling factors to the extent that goodness of fit values dropped 10% on average. This may have some bearing on the low values of R^2 encountered in this study. The problem of small sample sizes and exaggerated sampling errors may also be compounded by what some authors consider to be a statistical averaging effect which is inherent to the information-minimising model structure (Senior, 1976).

In summary, the results of this analysis indicate that, overall, there is little predictive advantage to be gained by income group disaggregation of a comprehensive spatial interaction model unless the specific focus of the study is to discern the differences that exist in locational and trip making behaviour between income categories. If such is the case then insight can be gained into subgroup differences especially via preliminary analysis, as has been done here. It is concluded that income disaggregation should be employed only when it serves to fulfill the primary objectives of the study.

CHAPTER SIX

CONCLUSIONS

This study has focussed upon the development and testing of an operational information-minimising model which has taken into account the differences that exist in locational, shop and work trip behaviour between different income categories of the urban population. In addition to the investigation into income group disaggregation this thesis has examined the effects of other improvements made to the model; i.e. increased number of shopping facilities, increased number of household categories, improved parameter search, and the use of better quality data. The modifications incorporated into the two current model versions, CALBTN2.1 and CALBTN2.2, reflect theoretical considerations and insights gained in the development of predecessors to this model. Of particular operational relevance are the data processing procedures required to process the new survey data and the improved calibration methods devised in an effort to alleviate the inherent combinatoric difficulties of the model (which have become even more pronounced due to income disaggregation).

The results of the empirical investigation into disaggregation, which follows directly from the analysis presented in Chapter Five, are summarized as follows. The findings of the preliminary analysis are the most encouraging with respect to income group disaggregation. The fact that significant differences exist in the mean cost estimates and the spatial distribution of work trip ends between categories lends some support to the use of disaggregation as a means of more accurately representing a system of interest. The discovery that mean travel times increase with household income reinforces an intuitive notion as to expected differences in travel behaviour between various categories of the population. In this respect income group disaggregation has served a useful purpose by capturing behaviour patterns which are concealed within a more aggregative model. Disaggregation has also proven to be an effective mechanism for discerning characteristics of sample data which may be overlooked when treating the data as a homogeneous unit. One such insight gained in this study has been the pronounced middle income bias of the survey data and its subsequent effects upon model predictions for both aggregate and disaggregate model versions.

Calibration and testing of the disaggregated model yields further insights into model structure and sheds some interesting light on the ability of the model to describe and reproduce a survey trip distribution pattern. In terms of goodness of fit correlations the results have proven to be less than encouraging. In the majority of observed/predicted comparisons for separate income groups the disaggregated model has failed to describe the observed spatial structure in an adequate manner. In large part the reason for this failure may be the effect of sampling errors and small sample sizes, induced by disaggregation. An examination of the fit of the sum of the income group predicted matrices to the sum of the income group observed matrices has revealed a performance which is comparable to the aggregate model version.

One redeeming feature of the modelling exercise is that the aggregate model version (treating the sample as a homogeneous group) has shown definite signs of improvement over previous aggregate models. These improvements may be attributed to the inclusion of more realistic representations of shopping facility distribution and the increased number of household categories, which undoubtedly captures more of the variation in underlying sample behaviour than previous models.

One other outcome of this study, which is of considerable importance, is that by exploring disaggregation possibilities some of the more pronounced problem areas of the model have been revealed. A clearer understanding of these problem areas suggests avenues for futher research. The poorness of the disaggregate model predictions may be a result of some degree of insensitivity inherent to the model. This may be attributable to the coarseness of the classification schemes used or to other aspects of the model which tend to average out variations in locational, work and shop trip making behaviour.

It is possible that some of the model's hypotheses and assumptions do not capture in sufficient detail the preferences and constraints which differentially affect the various categories of household. Several sources of complexity inherent to households and trip making behaviour have been ignored by the assumptions made in the model in order to render it more readily operational. These problem areas are now briefly summarized.

The residential location component of the model assumes that only heads of household choose residence - workplace combinations which affect the residential choice of the entire household. This ignores the effect that dependent workers within the household have upon the residential location decision. Although this hypothesis may be considered unrealistic, recent operational work by Senior (1977, 306) shows that there are considerable technical difficulties in overcoming this simplification and empirically little to be gained. Another assumption inherent to the residential location component of CALBTN2.1 and CALBTN2.2 concerns the supply of housing which is assumed to meet demand. This simplification may adversely affect the accuracy of model predictions, particularly for the income disaggregated model version. A more realistic location mechanism would limit the population in zones and would give lower income households the last choice of available housing.

The definition of trip types according to purpose of first stop may be resulting in inadequate representations of observed travel characteristics since the purpose of the first stop may only be incidental to the main purpose of the trip (e.g. stopping to purchase cigarettes on the journey to work). The problem is most pronounced for shopping trips, which in many cases are multi-purpose, multi-stop excursions. Although the bulk of trips for working purposes are probably single-purpose, single-stop journeys, the work trip component of the model does present some difficulties. One is faced with the problem of how to adequately represent trip costs for unemployed or retired households. The strategy employed in this model is to compute work trip costs for the no work category in any one zone as the average of work trip costs to all other zones. A more realistic method would be to utilize the social and recreational cost data available from the travel diaries. A closely related issue concerns the use of work trip travel cost matrices computed using off-peak hour travel times. This is questionable since the perceived cost of travel on congested routes during rush hour is undoubtedly greater. The incorporation of a capacity constrained algorithm for computing the work trip cost matrix, c_{ik}, would more realistically reflect the effects of congestion on work trip travel times.

With respect to income group disaggregation, a number of problems arise. The most urgent difficulty concerns data requirements. The results of testing CALBTN2.2 have indicated that additional income-specific data are required, most notably: (i) more complete samples of lower and higher income groups; (ii) observed zonal populations for each income group, to compute the population weighted mean cost estimates, and; (iii) data to construct prior probabilities which more accurately reflect the propensity of origin zones to generate work and shop trips. According to Alonso,

> "it is perfectly conceivable that we can devise predictive models which are beyond the capacity of the data, in the sense that, although they are more 'accurate' in their specification, the quality of the data results in a deterioration of prediction."
>
> (Alonso, 1968, 251)

The results of the analysis conducted in this thesis have indicated that model predictions for income group 1 are very different from the average. It is quite possible that behaviour patterns of low income households are considerably different than those of other income groups (owing largely to the unemployed or retired status of residents). Wilson (1970, 70) has suggested that behavioural differences in low income groups are sufficiently pronounced to violate the fundamental assumption of residential location models, i.e. that the distribution of residences is determined by worker's accessibilities to workplace. This might certainly be the case for low income families living in public housing. For this segment of the population Wilson proposes that it is more realistic to assume production-constrained residential location behaviour since workers are allocated to workplaces around fixed residences. Of course, the decision to build a residential location model specifically for low income households would be dictated by the research objectives.

In conclusion, it can be seen that a number of departure points exist for improving the operational model developed in this thesis. Unfortunately, from a practical-procedural perspective there are disincentives to experiment with potential improvements to the model, owing largely to the elaborate data estimation procedures. These procedures, in turn, are necessitated by the comprehensive nature of the model which creates problems of translating into concrete (operational) terms the various hypotheses embodied in the underlying theory. The experience gained in this study indicates that the most effective way to incorporate further improvements to the model involves considerable change in research strategy. It is therefore suggested that some effort be directed towards more clearly distinguishing between the residential location, work trip and shop trip subcomponents. The most direct way to improve each subcomponent would be to develop separate shopping, work trip and residential location models within a systems framework. The development of each model would involve empirical testing using the common study area and would utilize the data to their best advantage. Insights gained from the testing of each separate model would then guide the construction of an improved comprehensive model, taking into account income disaggregation where appropriate. The development of an improved comprehensive model, properly integrated and consistent, will be a difficult task demanding a balance between theory, objectivity and intuition.

APPENDIX 1

Appendix 1: The Newton-Raphson Method

The Newton-Raphson technique is one of the most widely used methods in numerical analysis. The following is a brief explanation of the method for the univariate case. Complete details may be found in many texts including those of Pennington (1970, 286), Gerald (1973, 7) and Acton (1970, 41).

If we let x_0 represent an approximate value for the root and h denote the correction that must be applied to give the true value of the root (denoted by x), then

$$x = x_0 + h$$

The equation for the real roots of $f(x) = 0$ is then,

$$f(x_0 + h) = 0.$$

Using a Taylor's Series expansion we obtain,

$$f(x_0+h) = f(x_0) + hf'(x_0) + (h^2/2)f''(x_0+\theta h), \text{ where}$$

$$0 < \theta < 1.$$

Therefore,

$$f(x_0) + hf'(x_0) + (h^2/2)f''(x_0+\theta h) = 0$$

If x_0 is a good approximation, then h will be small enough that we may ignore the term containing h^2, which leaves,

$$f(x_o) + hf'(x_0) \simeq 0.$$

We denote this approximate value of h by h_1 so that

$$h_1 = -\frac{f(x_o)}{f'(x_o)}$$

and x_1 is given by

$$x_1 = x_o + h = x_o - \frac{f(x_o)}{f'(x_o)}.$$

This should be a better approximation of the root than was x_o. By continuing the process we achieve better estimates. Using generalized notation, the formulation can be expressed as,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad \text{for } n = 1, 2, 3...$$

This iterative algorithm is continued until the successive x-values (iterates) converge upon the true root to the desired accuracy. A graphical description of the method, in the neighborhood of the root, is shown in Figure 1. The solution method proceeds as follows: at x_o draw a vertical line intersecting the curve at p_o. At p_o draw the tangent line. This intersects the x-axis at x_1. Continue by drawing a vertical line at x_1 which intersects the curve at p_1. Draw the tangent line at p_1. This intersects the axis at x_2, and so on. The ordinate of p_o is $f(x_o)$, therefore the slope of the tangent line is $f'(x_o)$. From the right triangle $x_o\ p_o\ x_1$ we see that,

$$\tan\theta = f'(x_o) = -\frac{f(x_o)}{x_1 - x_o}$$

or

$$x_1 = x_o - \frac{f(x_o)}{f'(x_o)} .$$

Hence, the formula can be derived from the figure. At each step of the process the tangent line is followed to the x-axis rather than the curve itself. Convergence is expected to be good as long as the derivative does not change too rapidly or the curve does not become nearly horizontal at the crossing point. Problems will arise if f' is zero or if it is small. If the function contains maxima and minima which are quite flat, then the chance that an arbitrary starting value for x will land on a dangerous flat part of the curve is large. Figure 2 is an example of this situation. In this case, the Newton-Raphson method will not converge. Starting with x_1 one never reaches the root r. The starting value makes a considerable difference in the Newton-Raphson method.

The principal advantage of the Newton-Raphson method is that it is rapidly (quadratically) convergent in the vicinity of the root. The net result of this "second-order" convergence is that the number of decimal places of accuracy nearly doubles at each iteration. The reason for this rapid convergence is now given.

The algorithm

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 1, 2, 3\ldots$$

is of the form

$$x_{n+1} = g(x_n)$$

for which successive iterations converge if

$$|g'(x)| < 1.$$

Since,

$$g(x) = x - \frac{f(x)}{f'(x)},$$

Figure 1   Graphical representation of the Newton-Raphson Method

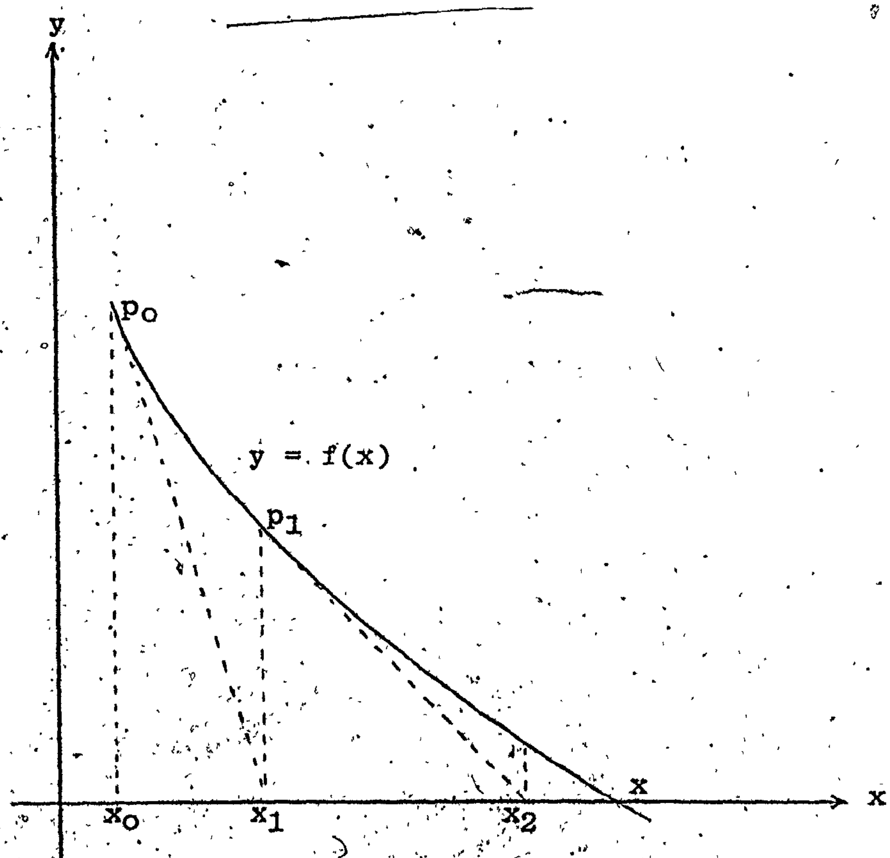
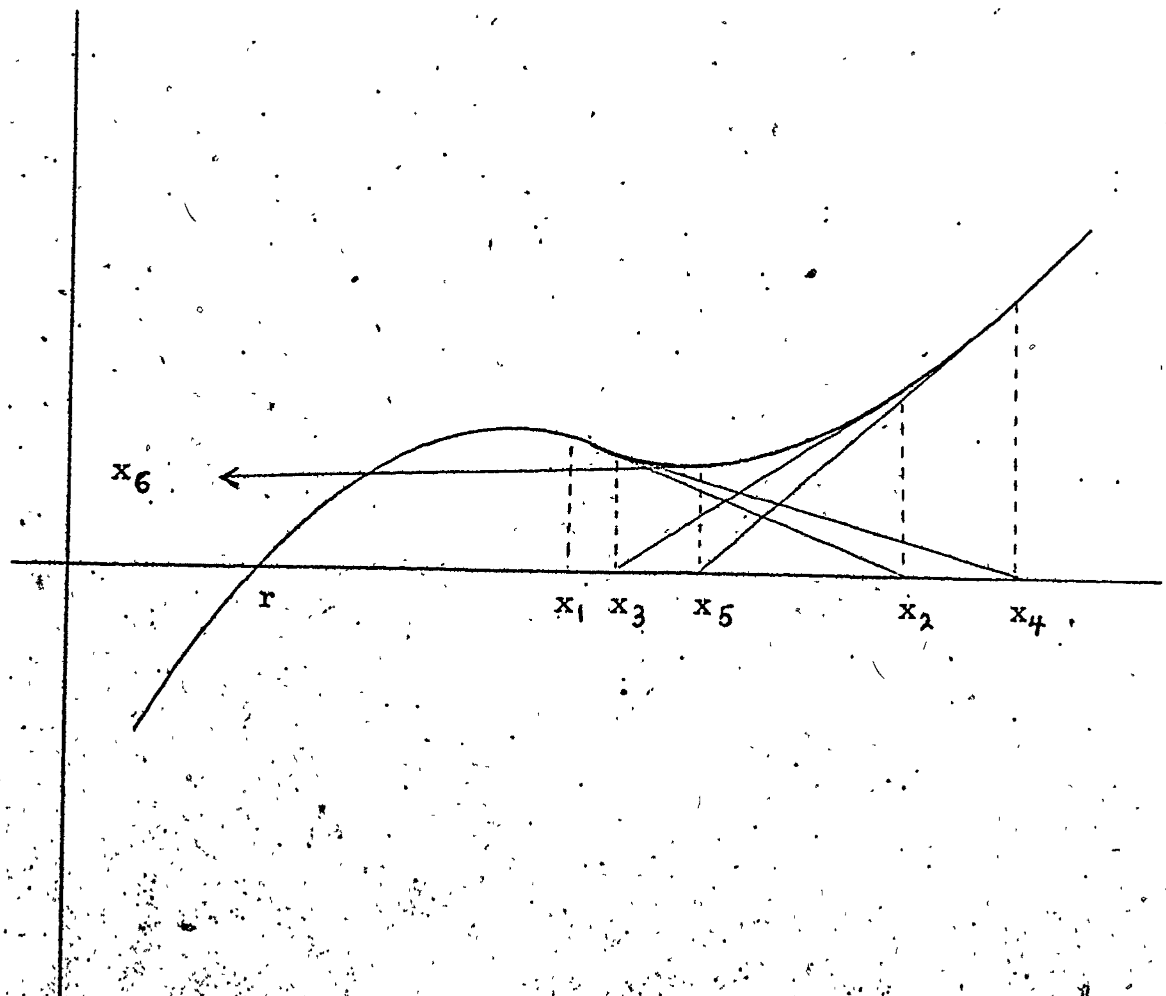


Figure 2 Graphical Representation of a Situation in which the Newton-Raphson Method does not converge



then

$$g'(x) = 1 - \frac{f'(x)f'(x) - f(x)f''(x)}{(f'(x))^2}$$

$$= \frac{f(x)\ f''(x)}{(f'(x))^2}$$

Hence, if

$$\left| \frac{f(x)\cdot f''(x)}{(f'(x))^2} \right| < 1 \qquad (1)$$

on an interval about the root r, then the method will converge for any initial value x_1 in this interval. This formulation (1) is called the convergence criterion. The usual requirement of continuity and existence of $f(x)$ and its derivatives must be met. It is also obvious that $f'(x)$ must not be zero.

If the convergence criterion is satisfied, then the speed of convergence can be shown to increase rapidly as the error becomes smaller. This is shown as follows.
The basic convergence scheme is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad (2)$$

If we denote the error at step n as e_n, then

$$e_n = x - x_n$$

and

$$e_{n+1} = x - x_{n+1}$$

By subtracting both sides of (2) from x we obtain

$$x - x_{n+1} = x - x_n + \frac{f(x_n)}{f'(x_n)}$$

Expressing this in terms of the errors,

$$e_{n+1} = e_n + \frac{f(x-e_n)}{f'(x-e_n)}$$

Now using a Taylor's series expansion,

$$e_{n+1} = e_n + \frac{f(x) - e_n\, f'(x) + (e_n^2/2)\, f''(x) + \ldots}{f'(x) - e_n\, f''(x) + (e_n^2/2)\, f'''(x) + \ldots}$$

We know that $f(x) = 0$ at the root. Using this information and dividing the denominator into the numerator, we obtain

$$e_{n+1} = e_n - e_n - \frac{e_n^2}{2}\,\frac{f''(x)}{f'(x)} + \text{higher-order terms}$$

or by omitting the higher-order terms, the resulting expression is

$$e_{n+1} \approx -\frac{e_n^2}{2}\,\frac{f''(x)}{f'(x)}\,.$$

This tells us that the error at step n+1 is proportional to the square of the error at step n. In other words, if the answer is correct to 1 decimal place at one step, at the next step it would be accurate to 2 decimal places, then 4 at the next step and so on. This rapid convergence is called "second-order" or quadratic convergence.

APPENDIX 2

Appendix 2.1 Diary Sample Page

(THE DIARY CONTAINED 14 PAGES LIKE THIS, ONE FOR EACH DAY)

HOUSEHOLD
Address:

Date: ___________

How many trips did the people in this household make today? ____

(1) START OF TRIP	(2)		(3)	(4)	(5)	(6)	(7) IF SHOPPING OR RECREATIONAL	(8)	(9) IF SOCIAL
Time of start	Place of start		Person(s) starting trip (all)	Means of travel	Reason for stop	Address of stop	Goods bought/ kind of recreation	Amount spent	State purpose
		FIRST STOP							
		SECOND STOP							
		THIRD STOP							
		FOURTH STOP							
		RETURN HOME			Time of arrival home: ______				
		FIRST STOP							
		SECOND STOP							
		THIRD STOP							
		FOURTH STOP							
		RETURN HOME			Time of arrival home: ______				
		FIRST STOP							
		SECOND STOP							
		THIRD STOP							
		FOURTH STOP							
		RETURN HOME			Time of arrival home: ______				

Appendix 2.2

CODING, HOUSEHOLD CARD

(One card per household)

COLUMN	DATA
1-2	Zone number
3-5	Neighbourhood number of present address (299 = missing, no information) (zones outside the region are also coded)
6-15	Surname
16-17	Initials
18-37	Apt., Street Number, Street Name
38-39	Day number that diary starts
40	Month (5 = May, 6 = June)
41-42	Day number of interview
43	Month
44-45	Household number (sequentially, by zone)
46-47	Interviewer who delivered diary
48-49	Interviewere who collected diary
50-52	Total number of trips made by household
53-55	Length of residence at this address (in tenths of a year)
56-58	Neighbourhood of previous address (000 = new household)
59-60	Number of permanent residents in household
61	Number of non-permanent residents
62-63	Number of permanent residents at time of move to this address
64	Number of children now living at home (i.e. persons less than or equal to 16 years of age or still attending high school)
65	Number of children at home at time of move to this address
66	Number of cards for person characteristics
67	Own (1) or rent (0) residence
68	Rent or mortgage payments: 1 = $ 0 - 100 per mo. 2 = $101 - 200 per mo. 3 = $201 - 300 per mo. 4 = $301 - 400 per mo. 5 = $401 - 500 per mo. 6 = greater than $500 per mo.
69	Gross household income per annum: 1 = $ 0 - 10,000 2 = $10,001 - 20,000 3 = $20,001 - 30,000 4 = $30,001 - 40,000 5 = $40,001 - 50,000 6 = exceeds $50,000

Appendix 2.3

CODING, PERSON CARD

(1 card for each person now in household <u>or</u> in household at time of move)

COLUMN	DATA
1-2	Zone Number
3-4	Household Number
5-6	Person Number
7	Sex and Kinship 1 = Male head of household, e.g. husband 2 = Female head of household, e.g. wife 3 = Child (see col 65, card 1 for definition) 4 = Other relative resident in household, e.g. adult son and daughter, mother of wife, aunt, uncle 5 = Non-relative 6 = Relative non-resident 9 = Missing, no information
8	Permanent Resident (1), Other (0)

If permanent resident now, fill in cols 9-23 otherwise go to col 24.

COLUMN	DATA
9-10	Year of last job e.g. 76 if last job in 1976 78 if lost job in 1978 79 if presently employed 00 never employed outside home 99 missing
11-12	Occupation = type of present job (if employed outside home) = type of last job (if not employed outside home)
13-14	Industry = type of employer
15-17	Neighbourhood of work or school Home = 300 Travelling/varied work location = 301
18	License 1 = yes Missing = 9
19	Car owner 1 = yes Missing = 9
20	Cycle owner 1 = yes Missing = 9
21	Motor cycle owner 1 = yes Missing = 9

Appendix 2.3 (cont'd)

COLUMN	DATA
22-23	Year of Birth (last two digits of year) 00 = 1900 or before 99 = missing/no answer
	If a resident when household moved to this address, complete cols 24-30 otherwise skip to col 31.
24-25	Occupation at Time of Move
26-27	Industry at Time of Move
28-30	Neighbourhood of Work or School at Time of Move
31-36	Cost of Car (ignore)
	If this person is not a child now resident in household.
37-40	For each neighbourhood visited for an activity
41-44	Code the neighbourhood number (1st 3 columns) and number of
45-48	times that neighbourhood is visited for that activity (last
49-52	columns) up to 11 activities may be coded.
53-56	
57-60	
61-64	
65-68	
69-72	
73-76	
77-78	

Appendix 2.4

CODING, TRIP CHARACTERISTICS

(One card for each trip)

COLUMN	DATA
1-2	Zone Number
3-4	Household Number
5-6	Trip Number (in sequence)
7	0
8-9	Day Number of Trip
10	Month (5 = May, 6 = June)
11-14	Time Trip Starts 24 hour clock 9999 = no information
15	1 if trip starts at residence blank otherwise
16-18	Neighbourhood of Start
19	Number of Stops on Trip Counting Return Home as One Stop
20-23	Time Trip Ends 24 hour clock
24	1 if trip ends at residence blank otherwise
25-26	Day Number of Trip End
27	Month of Trip End

Appendix 2.5

CODING, STOP INFORMATION

(One card for each stop on a trip)

COLUMN	DATA
1-2	Zone Number
3-4	Household Number
5-6	Trip Number
7	Stop Number (in sequence for this trip)
8	No. of Resident Adults Going to Stop
9	No. of Resident Children Going to Stop
10	No. of Non-residents Going to Stop
11	Means of Travel 1 = auto drive 2 = auto passenger 3 = bus 4 = walk 5 = taxi 5 = bike 7 = other 8 = motor cycle 9 = no info/missing
12	Purpose 1 = shopping 2 = work 3 = recreation 4 = social 5 = picking up/ dropping off 7 = other 8 = return home 9 = missing info
13	Second Purpose (if stated) as above
14	Third Purpose (if stated)
15-17	Neighbourhood of Stop

<u>If not a shopping stop, skip to col 31.</u>

18-19	Goods bought: see list
20-21	(up to 5 categories of goods may be specified
22-23	for each stop. 99: missing info.)
24-25	
26-27	
28-30	Total Amount Spend (whole dollars) 998 = ≥ $998 999 = missing info.

<u>If not a recreation stop, skip to col 36.</u>

31	Kind of Recreation: see list 9 = missing info.

Appendix 2.5 (cont'd)

COLUMN	DATA
32	Kind of Recreation: see list 9 = only 1 kind of recreation
33-35	Total Amount Spent

<u>If not a social stop, skip to col 41.</u>

36	Social Purpose: see list
37	Social Purpose 9 = only 1 social purpose
38-40	Total Amount Spend

Now list the number of every person (adult/child permanent member of household) who went to that stop.

41-42	1st person on trip
43-44	2nd
45-46	
.	
.	
.	

and keep going until all permanent members of household going to this stop are listed.

Appendix 2.6

LIST OF INDUSTRIES

00	Student, Housewifery, No Industry
10	Agriculture, Forestry, Fishing
20	Mines, Quarries
41	Food-Beverage Industries
42	Tobacco Products Industries
43	Rubber and Plastics Products Industries
44	Leather Industries
45	Textile Industries and Knitting Mills
47	Clothing Industries
48	Wood Industries
49	Furniture and Fixture Industries
50	Paper and Allied Industries
51	Printing, Publishing and Allied Industries
52	Primary Metal Industries
53	Metal Fabricating Industries
54	Machinery Industries
55	Transportation Equipment Industries
56	Electrical Products Industries
57	Non-metallic Mineral Products Industries
58	Petroleum and Coal Products Industries
59	Chemical and Chemical Products Industries
60	Miscellaneous Mfg. Industries
61	Construction
71	Transportation, Storage
73	Communication
74	Electric Power, Gas and Water Utilities
81	Wholesale Trade
82	Retail Trade
90	Finance and Real Estate
91	Education
92	Health and Welfare
93	Amusement and Recreation
94	Services to Business Management
95	Accomodation and Food Services
96	Personal and Miscellaneous Services
97	Public Administration and Defence
98	Unspecified
99	No information

<u>Source</u>: *Standard Industrial Classification Manual*, Dominion Bureau of Statistics (Ottawa, 1970). This source provides a more detailed breakdown of each class.

Appendix 2.6 (cont'd)

LIST OF OCCUPATIONS

11	Managerial, Administrative and Related Occupns.
21	Occupns. in Natural Sciences, Engineering, Mathematics
23	Occupns. in Social Sciences
25	Occupns. in Religion
27	Teaching and Related Occupns.
31	Occupns. in Medicine and Health
33	Artistic, Literary and Performing Arts
37	Occupns. in Sport and Recreation
41	Clerical and Related Occupns.
51	Sales Occupns.
61	Service Occupns.
71	Farming, Horticultural and Husbandry Occupns.
73	Fishing, Hunting and Trapping Occupns.
75	Forestry and Logging Occupns.
77	Mining and Quarrying Occupns.
81	Processing Occupns.
83	Machinery Occupns.
85	Product Fabricating, Assembling and Repairing
87	Construction Trades
91	Transport Equipment Operating
93	Materials- Handling Occupns.
95	Other Crafts
99	Occupns. not elsewhere classified
98	Missing Information
97	Housewife
96	Student
94	Retired
92	Unemployed
89	Too young to go to school

Appendix 2.7

Recreation Purposes

0	Visit library
1	Visit gallery
2	Concerts and movies
3	Sporting events
4	Bars and restaurants
5	Parks
6	Participating in sports
7	Walking, cycling
8	Night classes and other lessons
9	Missing information

Social Purposes

1	Visit friends
2	Visit relatives
3	Clubs
4	Organizations
5	Place of worship
6	Volunteer work
7	Other
8	Dance
9	No information

List of Shopping Goods

01	Groceries
02	Clothes (includes shoes, accessories etc.)
03	Appliances
04	Furniture
05	Garden supplies
06	Toys, sporting goods and hobby equipment
07	Beer and liquor
08	'Take out' food (includes candy, hamburgers, ice-cream etc.)
09	Jewellery
10	Tobacco, newspapers and magazines
11	Hardward and houseware
12	Drugs and cosmetics
13	Books, stationery, cards and records
14	Gas, oil, tires and car parts
15	Car repair or car wash
16	Car purchase
17	Appliance repair
18	Dry cleaning and shoe repairs
19	Barber and beauty salon
20	Banking and other financial services (we do not want to know how much you withdrew)
21	Dentist, physician, lawyer or veterinarian
22	Other/window shopping
23	Purchases in a post office
24	Return goods
99	No information

Appendix 2.8 The Data Contained in IDEN

IDEN is a file of neighbourhood characteristics, with data as follows:

IDO, NAM, ID1, ID2, POP77, BASA, BASU, SERA, SERU, ID3, SUPER, DEPT1, DEPT2, MAJ, MAJST, MIN, MINST, STRIP
with format

I3, 1X, A10, 1X, I4, I3, F6.0, F6.1, F4.0, F6.1, 1X, F4.0, 1X, I3, 1X, I2, 1X, I2, 1X, I2, 1X, I2, 2X, I2, 1X, I2, 2X, I2, 1X, I2.
The variables are:

IDO: position of neighbourhood on travel time arrays. IDO = 1 implies that this neighbourhood occupies the first row of the travel time array.

NAM: name of neighbourhood.

ID1: Regional Municipality of Hamilton Wentworth city planning division neighbourhood identifies.

ID2: zone to which neighbourhood belongs.

POP77: 1977 neighbourhood population estimate.

BASA: 1977 acreage in industrial land uses.

BASU: 1977 number of industrial assessment units.

SERA: 1977 acreage in service land uses.

SERU: 1977 number of service assessment units.

ID3: neighbourhood reference number, used in coding diaries.

SUPER: number of supermarkets, June 1979.

DEPT1: number of major department stores, June 1979.

DEPT2: number of minor department stores, June 1979.

MAJ: number of major shopping centres, June 1979.

MAJST: number of stores in major shopping centres, June 1979.

MIN: number of minor shopping centres, June 1979.

MINST: number of stores in minor shopping centres, June 1979.

STRIP: number of shopping strips.

APPENDIX 3

Appendix 3:

The Significance of Differences in Mean Cost Values

Suppose that two separate samples of trip making behaviour are made. One possibility would be that two surveys are carried out by sampling the same categories of trip maker at different times of the year. Important questions which would arise when the results of the two surveys are compared relate to the significance of differences between the two surveys, for example, could the differences in mean cost values found for the same category at different times of the year be explained by sampling error? The purpose of this appendix is to examine the significance of differences in average cost values for the different income categories, as presented in Table 5.1 of Chapter 5.

The mean cost values presented in that table were based upon an areally stratified random sample of households in 1978, and are liable to sample error. The t test described below is used to test for significance in differences in mean cost values between income categories. The existence of four noninteracting (income) groups of trip maker is postulated. The problem is to decide whether observed differences among the sample means can be attributed to chance or whether there are real differences among the means of the populations sampled.

The appropriate null hypothesis and alternate hypothesis are therefore:

$$H_o\ (\mu_1 - \mu_2 = 0)$$

and

$$H_a\ (\mu_1 - \mu_2 \neq 0).$$

We need a measure of the discrepancies among the means and with it a rule which tells us when the discrepancies are so large that the null hypothesis should be rejected. The measure used is the variance of the means; the rule is the t statistic for small sample tests concerning the difference between two means (Freund, 1979, 278.) The t test involves the assumption that the two sample distributions have the same variance and that the two independent random samples are drawn from populations which can be approximated closely by the normal distribution. The t statistic is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left[\frac{(n_1-1)\cdot S_1^2 + (n_2-1)\cdot S_2^2}{n_1 + n_2 - 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}},$$

with $n_1 + n_2 - 2$ degrees of freedom (Freund, 1979, 278). The t test is done on a pairwise basis for each mean cost value of the four income groups. Therefore, for each of the three mean costs (mean shop trip cost, mean work trip cost and mean number of shop trips per week) there are $\binom{4}{2}$, or 6 t tests.

Prior to conducting the t tests it was decided to more carefully examine the assumption regarding equal variance for the two sample distributions. A variance ratio (F) statistic was used in a test concerning the equality of the standard deviations (Freund, 1979, 292). The null and alternate hypotheses were:

$$H_o(\sigma_1 = \sigma_2) \text{ and } H_a(\sigma_1 \neq \sigma_2).$$

The following statistic determines if the above assumption regarding equal variance is actually tenable:

$$F = \frac{S_1^2}{S_2^2} \text{ or } \frac{S_1^2}{S_1^2},$$

(whichever is larger), with n_1-1 and n_2-1 degrees of freedom. Again, the test must be done for each pairwise combination of standard deviations of the four income groups. The results given in Table 1 indicate that in the majority of cases the assumption regarding equal variance is warranted, thereby supporting the use of the t test.

The t test results are presented in Table 2. It is found that for mean work trip costs the only sufficient evidence to reject the null hypothesis (at the 97.5% level of confidence) is between group 1 and all other groups. The differences in mean shop trip cost between income groups are significant except between groups 1 and 2. However with respect to the mean number of shop trips per week, there is not enough evidence to reject the null hypothesis at the 97.5% level of confidence. In most cases, the differences between mean shop trip and mean work trip costs of different income groups are significant and the observed differences between the mean number of shop trips per week of different income groups are not significant.

Table 1 Test Concerning Equal Variance

	$\bar{c}_s^v$	$\bar{c}_w^v$	$\bar{x}^v$
F(1,2)	4.12*	1.69	4.06*
F(1,3)	2.49*	1.43	9.02*
F(1,4)	2.51*	1.64	1.00
F(2,3)	1.65	1.18	2.22
F(2,4)	1.63	1.03	4.04*
F(3,4)	1.00	1.14	8.96*

Note: Critical value is $F_{.05} = 2.35$

F(1,2) refers to F ratio comparing income groups 1 and 2

*denotes rejection of $H_o(\sigma_1 = \sigma_2)$

Table 2 t Test Concerning the Difference Between Mean Values

	$\bar{c}_s^v$	$\bar{c}_w^v$	$\bar{x}^v$
t(1,2)	0.202*	22.04	0.01*
t(1,3)	6.99	20.75	0.15*
t(1,4)	9.23	20.24	0.70*
t(2,3)	10.79	0.63*	0.24*
t(2,4)	14.17	1.95*	0.91*
t(3,4)	3.00	1.25*	0.83*

Note: Critical value is $t_{.025} = 1.96$

t(1,2) refers to t test comparing income groups 1 and 2

*denotes acceptance of $H_o(\mu_1 - \mu_2) = 0$

APPENDIX 4

APPENDIX 4.1 POPULATION DISTRIBUTIONS

EXPECTED POPULATION PROPORTIONS

Zone	Actual	1	2	3	4	SUM	AGGREGATE
1	.0166	.0295	.0302	.0290	.0253	.0291	.0316
2	.0239	.0199	.0708	.0632	.0737	.0624	.0689
3	.0924	.0244	.1075	.1065	.1175	.0981	.0996
4	.1195	.0175	.1670	.1385	.1632	.1386	.1538
5	.1259	.0527	.1041	.0916	.0973	.0927	.1021
6	.0740	.2410	.0593	.0552	.0438	.0421	.0616
7	.0471	.0803	.0377	.0383	.0291	.0421	.0409
8	.0536	.0845	.0483	.0594	.0444	.0559	.0590
9	.0797	.1115	.0507	.0580	.0517	.0610	.0547
10	.1237	.1785	.1153	.1116	.1108	.1217	.1163
11	.0626	.1370	.0445	.0418	.0467	.0559	.0425
12	.0264	.0021	.0284	.0343	.0348	.0278	.0297
13	.0065	.0014	.0289	.0348	.0367	.0284	.0301
14	.0066	.0089	.0267	.0356	.0336	.0282	.0294
15	.0529	.0107	.0560	.0676	.0648	.0551	.0540
16	.0602	.0001	.0179	.0261	.0195	.0184	.0208
17	.0283	.0000	.0046	.0084	.0070	.0055	.0058

Appendix 4.2

Shop Trip Distribution Matrices

Income Group 1

ACTUAL SHOP TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	.003	0.000	0.000	.002	.001	.006	0.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	.006	.004	.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	.040	.017	.003	.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.014	0.000	0.000
4	0.000	0.000	.021	.082	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.009	0.000	0.000
5	0.000	0.000	0.000	.008	.059	.028	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	.002	.001	.036	.005	0.000	0.000	.001	.002	0.000	0.000	0.000	0.000	.003	.002
7	0.000	0.000	0.000	0.000	0.000	.005	.021	.001	0.000	0.000	.001	0.000	0.000	0.000	0.000	.007	0.000
8	0.000	0.000	.001	0.000	0.000	.002	0.000	.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.003	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.062	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	.019	0.000	0.000	0.000	0.000	.097	.174	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.020	.065	0.000	.007	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.025	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.065	0.000
17	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.003	0.000	0.000	.001	.004	0.000	0.000	0.000	.014
TOTAL	.003	0.000	.071	.136	.068	.088	.026	.016	.162	.200	.068	.001	.011	0.000	.049	.086	.016

PREDICTED SHOP TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0.000	0.000	.002	.004	.004	.007	.002	.002	.001	.003	.002	0.000	0.000	0.000	.001	.002	0.000
2	0.000	0.000	.002	.004	.003	.003	.001	.001	.001	.002	.003	0.000	0.000	0.000	.001	.001	0.000
3	0.000	0.000	.004	.005	.002	.003	.001	.001	.001	.003	.003	0.000	0.000	0.000	.002	.001	0.000
4	0.000	0.000	.002	.004	.002	.002	.001	.001	.001	.002	.002	0.000	0.000	0.000	.001	.001	0.000
5	0.000	0.000	.004	.009	.008	.010	.003	.002	.003	.005	.005	0.000	0.000	0.000	.002	.003	0.000
6	0.000	0.000	.015	.031	.028	.054	.017	.013	.014	.027	.022	0.000	0.000	0.000	.006	.018	0.000
7	0.000	0.000	.005	.009	.008	.017	.006	.006	.005	.004	.007	0.000	0.000	0.000	.002	.008	0.000
8	0.000	0.000	.005	.009	.008	.015	.006	.008	.004	.003	.006	0.000	0.000	0.000	.002	.012	0.000
9	0.000	0.000	.007	.013	.012	.019	.007	.006	.009	.014	.011	0.000	0.000	0.000	.003	.008	0.000
10	0.000	0.000	.013	.025	.018	.023	.009	.007	.011	.027	.025	0.000	0.000	0.000	.005	.010	0.000
11	0.000	0.000	.012	.023	.013	.014	.006	.005	.007	.014	.023	0.000	0.000	0.000	.005	.006	0.000
12	0.000	0.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	0.000	0.000	0.000	.000	.000	0.000
13	0.000	0.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	0.000	0.000	0.000	.000	.000	0.000
14	0.000	0.000	.001	.001	.001	.001	.001	.000	.001	.001	.001	0.000	0.000	0.000	.000	.001	0.000
15	0.000	0.000	.002	.002	.001	.001	.000	.000	.000	.001	.001	0.000	0.000	0.000	.001	.000	0.000
16	0.000	0.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	0.000	0.000	0.000	.000	.000	0.000
17	0.000	0.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	0.000	0.000	0.000	.000	.000	0.000
TOTAL	0.000	0.000	.075	.140	.107	.179	.060	.050	.058	.114	.110	0.000	0.000	0.000	.031	.072	0.000

Income Group 2

ACTUAL SHOP
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	.005	0.000	0.000	.000	.001	.004	0.000	0.000	0.000	.001	.001	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	.006	.005	.010	.003	.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.001	0.000	0.000
3	0.000	.001	.075	.010	0.000	.001	0.000	0.000	0.000	.001	.006	0.000	0.000	0.000	.018	0.000	0.000
4	0.000	.001	.005	.102	.012	.008	0.000	0.000	0.000	0.000	.002	0.000	0.000	0.000	.002	0.000	0.000
5	0.000	.002	0.000	.018	.079	.013	0.000	.003	0.000	.003	.003	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	.002	.047	.003	.002	.001	.002	0.000	0.000	0.000	0.000	0.000	.002	0.000
7	.001	0.000	.001	.003	.005	.012	.027	.033	.001	.001	0.000	0.000	0.000	0.000	0.000	.004	0.000
8	0.000	0.000	0.000	0.000	0.000	.006	.003	.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.013	0.000
9	0.000	0.000	.001	.001	0.000	.010	.001	0.000	.048	.023	.003	0.000	0.000	0.000	.001	.002	.011
10	0.000	0.000	0.000	.001	.001	.068	0.000	0.000	.003	.055	.026	0.000	0.000	0.000	0.000	.001	0.000
11	0.000	0.000	0.000	.002	.002	.002	0.000	0.000	0.000	.003	.044	.001	.002	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	.000	.000	.001	0.000	.000	.001	.003	.014	.002	.001	.000	0.000	0.000	0.000
13	0.000	0.000	.000	.000	.000	.000	0.000	0.000	.001	.001	.001	.000	.002	0.000	0.000	0.000	.000
14	0.000	0.000	0.000	.000	0.000	.001	.000	.000	.003	.001	.000	0.000	.001	.002	0.000	.000	0.000
15	.001	0.000	.012	.002	0.000	.002	.001	0.000	0.000	.011	0.000	0.000	0.000	0.000	.040	0.000	0.000
16	0.000	.001	0.000	0.000	.001	.001	0.000	.001	0.000	.002	.003	0.000	0.000	0.000	0.000	.052	0.010
17	0.000	0.000	0.000	0.000	.001	.002	.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.001	.021
TOTAL	.006	.010	.030	.149	.107	.120	.036	.037	.058	.107	.102	.003	.006	.002	.061	.076	.032

PREDICTED SHOP
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0.000	0.000	.002	.004	.004	.007	.002	.002	.001	.003	.002	0.000	0.000	0.000	.001	.002	0.000
2	0.000	0.000	.008	.015	.009	.011	.003	.003	.003	.006	.006	0.000	0.000	0.000	.003	.004	0.000
3	0.000	0.000	.017	.023	.011	.012	.004	.003	.004	.003	.011	0.000	0.000	0.000	.008	.004	0.000
4	0.000	0.000	.022	.039	.019	.032	.006	.005	.007	.015	.019	0.000	0.000	0.000	.009	.007	0.000
5	0.000	0.000	.009	.018	.016	.019	.005	.004	.005	.011	.010	0.000	0.000	0.000	.004	.006	0.000
6	0.000	0.000	.004	.008	.007	.013	.004	.003	.003	.007	.005	0.000	0.000	0.000	.002	.004	0.000
7	0.000	0.000	.003	.004	.004	.008	.003	.003	.002	.004	.003	0.000	0.000	0.000	.001	.004	0.000
8	0.000	0.000	.003	.005	.004	.003	.004	.004	.002	.004	.004	0.000	0.000	0.000	.001	.002	0.000
9	0.000	0.000	.003	.006	.005	.003	.003	.003	.004	.007	.005	0.000	0.000	0.000	.001	.004	0.000
10	0.000	0.000	.003	.016	.012	.019	.006	.005	.007	.018	.016	0.000	0.000	0.000	.003	.006	0.000
11	0.000	0.000	.004	.008	.004	.006	.002	.001	.002	.006	.008	0.000	0.000	0.000	.002	.002	0.000
12	0.000	0.000	.002	.004	.003	.004	.001	.001	.002	.004	.004	0.000	0.000	0.000	.001	.002	0.000
13	0.000	0.000	.003	.004	.003	.004	.001	.001	.002	.004	.004	0.000	0.000	0.000	.001	.001	0.000
14	0.000	0.000	.002	.003	.003	.014	.002	.001	.002	.004	.003	0.000	0.000	0.000	.001	.002	0.000
15	0.000	0.000	.010	.011	.005	.006	.002	.002	.002	.004	.005	0.000	0.000	0.000	.006	.002	0.000
16	0.000	0.000	.001	.002	.001	.003	.001	.001	.001	.001	.001	0.000	0.000	0.000	.000	.004	0.000
17	0.000	0.000	.000	.000	.000	.001	.000	.000	.000	.000	.000	0.000	0.000	0.000	.000	.001	0.000
TOTAL	0.000	0.000	.033	.170	.110	.158	.050	.043	.051	.107	.104	0.000	0.000	0.000	.043	.062	0.000

Income Group 3

ACTUAL SHOP TRIP INTERCHANG MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	.002	0.000	0.000	.001	0.000	.011	0.000	.001	0.000	.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	.004	.001	.004	.004	.002	0.000	0.000	0.000	0.000	0.000	.001	0.000	0.000	.000	0.000	0.000
3	0.000	0.000	.072	.008	0.000	0.000	0.000	0.000	.001	.001	.001	0.000	0.000	0.000	.003	0.000	0.000
4	0.000	0.000	.025	.086	0.000	.005	0.000	.003	0.000	0.000	0.000	0.000	0.000	0.000	.007	0.000	0.000
5	0.000	0.000	.003	.050	.050	.013	0.000	0.000	0.000	.005	0.000	0.000	.003	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	.003	.043	.003	0.000	0.000	.015	.003	0.000	0.000	0.000	.003	.003	0.000
7	0.000	0.000	0.000	.008	.004	.013	.020	.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.003	0.000
8	0.000	0.000	.001	.001	.002	.008	.001	.026	0.000	0.000	0.000	0.000	0.000	0.000	.001	.011	.010
9	0.000	0.000	0.000	0.000	0.000	.004	0.000	0.000	.051	.013	.002	0.000	0.000	0.000	0.000	.001	.001
10	0.000	0.000	.001	.001	.003	.022	0.000	0.000	.028	.066	.025	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	.003	.002	.002	.004	0.000	0.000	.003	.003	.042	.002	.001	0.000	.001	0.000	0.000
12	0.000	0.000	0.000	.000	0.000	.001	0.000	.001	.000	.007	.015	.001	.001	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	.001	0.000	0.000	.002	.002	.000	.000	.002	0.000	0.000	0.000	0.000
14	0.000	0.000	.000	0.000	0.000	.000	.000	0.000	.003	.001	.000	.000	.000	.002	0.000	0.000	0.000
15	0.000	0.000	.007	.001	0.000	.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.029	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	.007	0.000	.002	0.000	0.000	.001	0.000	0.000	0.000	0.000	.056	.001
17	0.000	0.000	0.000	0.000	0.000	.001	0.000	0.000	.003	.001	.000	0.000	0.000	0.000	0.000	.003	.023
TOTAL	.002	.004	.112	.162	.068	.159	.024	.035	.091	.137	.090	.004	.006	.002	.045	.078	.034

PREDICTED SHOP TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0.000	0.000	.003	.004	.003	.005	.002	.002	.001	.003	.003	0.000	0.000	0.000	.001	.003	0.000
2	0.000	0.000	.007	.011	.007	.010	.003	.003	.003	.006	.006	0.000	0.000	0.000	.003	.005	0.000
3	0.000	0.000	.013	.019	.011	.015	.005	.004	.005	.010	.011	0.000	0.000	0.000	.006	.007	0.000
4	0.000	0.000	.016	.025	.015	.020	.006	.006	.006	.014	.015	0.000	0.000	0.000	.007	.010	0.000
5	0.000	0.000	.009	.015	.011	.015	.005	.004	.005	.009	.009	0.000	0.000	0.000	.004	.007	0.000
6	0.000	0.000	.005	.008	.006	.010	.003	.003	.003	.006	.005	0.000	0.000	0.000	.002	.005	0.000
7	0.000	0.000	.003	.005	.004	.007	.002	.002	.002	.004	.004	0.000	0.000	0.000	.001	.004	0.000
8	0.000	0.000	.005	.008	.006	.010	.004	.004	.003	.006	.005	0.000	0.000	0.000	.002	.007	0.000
9	0.000	0.000	.005	.008	.006	.009	.003	.003	.004	.007	.006	0.000	0.000	0.000	.002	.005	0.000
10	0.000	0.000	.010	.016	.011	.018	.006	.005	.006	.013	.013	0.000	0.000	0.000	.005	.009	0.000
11	0.000	0.000	.004	.007	.004	.006	.002	.002	.002	.005	.005	0.000	0.000	0.000	.002	.003	0.000
12	0.000	0.000	.003	.005	.003	.005	.002	.002	.002	.004	.004	0.000	0.000	0.000	.001	.003	0.000
13	0.000	0.000	.003	.005	.003	.005	.002	.002	.002	.004	.004	0.000	0.000	0.000	.001	.003	0.000
14	0.000	0.000	.003	.005	.004	.006	.002	.002	.002	.004	.004	0.000	0.000	0.000	.001	.003	0.000
15	0.000	0.000	.009	.011	.007	.009	.003	.003	.003	.006	.007	0.000	0.000	0.000	.004	.005	0.000
16	0.000	0.000	.002	.003	.002	.004	.001	.002	.001	.002	.002	0.000	0.000	0.000	.001	.004	0.000
17	0.000	0.000	.001	.001	.001	.001	.000	.000	.000	.001	.001	0.000	0.000	0.000	.000	.001	0.000
TOTAL	0.000	0.000	.099	.157	.104	.154	.051	.047	.051	.104	.104	0.000	0.000	0.000	.046	.043	0.000

Income Group 1

ACTUAL SHOP TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	.002	0.000	0.000	0.000	.002	.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	.017	.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	.055	.015	.002	.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.006	0.000	0.000
4	0.000	0.000	0.000	.077	.006	0.000	0.000	0.000	0.000	0.000	.003	0.000	0.000	0.000	.006	0.000	0.000
5	0.000	0.000	0.000	.029	.081	0.000	0.000	0.000	0.000	.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	.020	.020	0.000	.031	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	.088	.031	.009	0.000	.002	0.000	0.000	0.000	0.000	0.000	.004	0.000
8	0.000	0.000	0.000	0.000	.004	.002	.007	.033	0.000	.032	0.000	0.000	0.000	0.000	0.000	.009	0.000
9	0.008	0.000	0.000	0.000	0.000	.010	0.000	0.000	.040	.030	.001	0.000	0.000	.003	0.000	0.000	.031
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.017	.184	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	.001	0.000	.003	0.000	0.000	0.000	.001	.032	.003	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.003	.009	.002	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	.000	.000	0.000	0.000	.000	.001	.002	.000	.000	.002	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.001	.002	.000	0.000	.000	.002	0.000	0.000	0.000
15	0.000	0.000	.010	.001	0.000	.003	0.000	0.000	0.000	.001	0.000	0.000	0.000	0.000	.033	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	.002	0.000	0.000	0.000	.001	0.000	0.000	0.000	0.000	0.000	.055	.001
17	0.000	0.000	.000	.000	0.000	.000	0.000	.000	.002	.000	0.000	0.000	0.000	0.000	0.000	.003	.015
TOTAL	.002	.017	.073	.124	.115	.062	.033	.073	.061	.245	.045	.005	.002	.004	.045	.072	.011

PREDICTED SHOP TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0.000	0.000	.002	.004	.003	.004	.001	.001	.001	.002	.002	0.000	0.000	0.000	.001	.002	0.000
2	0.000	0.000	.008	.012	.008	.011	.004	.003	.003	.007	.007	0.000	0.000	0.000	.004	.006	0.000
3	0.000	0.000	.014	.020	.012	.016	.005	.005	.005	.011	.012	0.000	0.000	0.000	.007	.009	0.000
4	0.000	0.000	.018	.029	.017	.024	.008	.007	.008	.016	.017	0.000	0.000	0.000	.008	.012	0.000
5	0.000	0.000	.003	.016	.011	.016	.005	.005	.005	.010	.010	0.000	0.000	0.000	.004	.008	0.000
6	0.000	0.000	.004	.006	.005	.008	.002	.002	.002	.005	.004	0.000	0.000	0.000	.002	.004	0.000
7	0.000	0.000	.002	.004	.003	.005	.002	.002	.002	.003	.003	0.000	0.000	0.000	.001	.003	0.010
8	0.000	0.000	.004	.006	.004	.007	.003	.003	.002	.004	.004	0.000	0.000	0.000	.002	.005	0.000
9	0.000	0.000	.004	.007	.005	.008	.003	.003	.003	.004	.005	0.000	0.000	0.000	.002	.005	0.000
10	0.000	0.000	.010	.016	.011	.017	.006	.005	.006	.013	.013	0.000	0.000	0.000	.005	.009	0.010
11	0.000	0.000	.005	.007	.005	.007	.002	.002	.002	.005	.006	0.000	0.000	0.000	.002	.004	0.000
12	0.000	0.000	.003	.005	.003	.005	.002	.002	.002	.004	.004	0.000	0.000	0.000	.001	.003	0.000
13	0.000	0.000	.003	.005	.004	.006	.002	.002	.002	.004	.004	0.000	0.000	0.000	.002	.003	0.000
14	0.000	0.000	.003	.005	.003	.005	.002	.002	.002	.004	.003	0.000	0.000	0.000	.001	.003	0.000
15	0.000	0.000	.008	.011	.006	.009	.003	.003	.003	.006	.006	0.000	0.000	0.000	.004	.005	0.000
16	0.000	0.000	.002	.003	.002	.003	.001	.001	.001	.002	.002	0.000	0.000	0.000	.001	.003	0.000
17	0.000	0.000	.001	.001	.001	.001	.000	.000	.000	.001	.001	0.000	0.000	0.000	.000	.001	0.000
TOTAL	0.000	0.000	.100	.157	.104	.153	.050	.047	.051	.103	.104	0.000	0.000	0.000	.047	.084	0.000

Aggregate

ACTUAL SHOP TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	.003	0.000	0.000	.001	.001	.005	0.000	.000	0.000	.003	.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	.004	.002	.005	.002	.001	0.000	0.000	0.000	0.000	0.000	.000	0.000	0.000	.000	0.000	0.000
3	0.000	.000	.065	.011	.001	.002	0.000	0.000	.000	.001	.003	0.000	0.000	0.000	.011	0.000	0.000
4	0.000	.000	.015	.114	.008	.005	0.000	.001	0.000	0.000	.002	0.000	0.000	0.000	.006	0.000	0.000
5	0.000	.001	.001	.023	.065	.015	0.000	.001	0.000	.004	.001	0.000	.001	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	.001	.002	.045	.004	.002	.000	.004	.001	0.000	0.000	0.000	.000	.003	.001
7	.000	0.000	.000	.002	.002	.009	.023	.063	.000	.001	.000	0.000	0.000	0.000	0.000	.004	0.000
8	0.000	0.000	.001	.000	.001	.005	.002	.028	0.000	.000	0.000	0.000	0.000	0.000	.000	.013	0.000
9	0.000	0.000	.000	.001	0.000	.069	.001	0.000	.052	.022	.003	0.000	0.000	.001	.000	.001	.001
10	0.000	0.000	.001	.002	.002	.012	0.000	0.000	.015	.063	.022	0.000	0.000	0.000	0.000	.001	0.000
11	0.000	0.000	.001	.002	.001	.003	0.000	0.000	.001	.008	.045	.001	.002	0.000	.000	0.000	0.000
12	0.000	0.000	0.000	.000	.000	.001	0.000	.000	.000	.005	.011	.001	.001	.000	0.000	0.000	0.000
13	0.000	0.000	.000	.000	.000	.003	0.000	.000	.001	.002	.001	.000	.002	0.000	0.000	0.000	.000
14	0.000	0.000	.000	.000	0.000	.003	.000	.000	.002	.001	.000	.000	.000	.002	0.000	.000	0.000
15	.000	0.000	.011	.001	0.000	.002	.000	0.000	0.000	.001	0.000	0.000	0.000	0.000	.040	0.000	0.000
16	0.000	.000	0.000	0.000	.000	.004	0.000	.001	0.000	.001	.001	0.000	0.000	0.000	0.000	.059	.000
17	0.000	0.000	.000	.000	.000	.001	.000	.000	.002	.000	.000	.000	.001	0.000	0.000	.003	.021
TOTAL	.004	.005	.097	.162	.087	.123	.030	.038	.075	.113	.089	.003	.005	.002	.059	.083	.023

PREDICTED SHOP TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0.000	0.000	.002	.005	.004	.007	.002	.002	.002	.003	.003	0.000	0.000	0.000	.001	.003	0.000
2	0.000	0.000	.008	.013	.009	.011	.003	.003	.003	.006	.006	0.000	0.000	0.000	.003	.004	0.000
3	0.000	0.000	.015	.020	.010	.012	.004	.003	.004	.003	.011	0.000	0.000	0.000	.007	.005	0.000
4	0.000	0.000	.019	.033	.017	.024	.005	.005	.006	.015	.017	0.000	0.000	0.000	.008	.009	0.000
5	0.000	0.000	.009	.017	.014	.015	.005	.004	.005	.010	.010	0.000	0.000	0.000	.004	.005	0.000
6	0.000	0.000	.004	.008	.007	.013	.004	.003	.004	.007	.006	0.000	0.000	0.000	.002	.005	0.000
7	0.000	0.000	.003	.005	.004	.003	.003	.003	.002	.004	.004	0.000	0.000	0.000	.001	.004	0.000
8	0.000	0.000	.004	.007	.006	.010	.004	.005	.003	.005	.005	0.000	0.000	0.000	.002	.008	0.000
9	0.000	0.000	.004	.007	.006	.003	.003	.003	.004	.007	.005	0.000	0.000	0.000	.002	.004	0.000
10	0.000	0.000	.009	.017	.012	.019	.006	.005	.007	.016	.015	0.000	0.000	0.000	.004	.007	0.000
11	0.000	0.000	.004	.007	.004	.005	.002	.002	.002	.005	.007	0.000	0.000	0.000	.002	.002	0.000
12	0.000	0.000	.002	.004	.003	.005	.002	.001	.002	.004	.004	0.000	0.000	0.000	.001	.002	0.000
13	0.000	0.000	.002	.004	.003	.005	.001	.001	.002	.004	.004	0.000	0.000	0.000	.001	.002	0.000
14	0.000	0.000	.002	.004	.003	.005	.002	.001	.002	.004	.003	0.000	0.000	0.000	.001	.002	0.000
15	0.000	0.000	.008	.010	.005	.005	.002	.002	.002	.004	.005	0.000	0.000	0.000	.005	.003	0.000
16	0.000	0.000	.001	.002	.002	.003	.001	.002	.001	.002	.002	0.000	0.000	0.000	.001	.004	0.000
17	0.000	0.000	.000	.001	.001	.001	.000	.000	.000	.001	.001	0.000	0.000	0.000	.000	.001	0.000
TOTAL	0.000	0.000	.097	.163	.108	.155	.051	.045	.051	.105	.106	0.000	0.000	0.000	.043	.070	0.000

Sum of Income Groups 1 to 4

ACTUAL SHOP
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	.001	0.000	0.000	.001	.001	.007	0.000	.000	0.000	.001	.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	.006	.003	.006	.003	.001	0.000	0.000	0.000	0.000	0.000	.000	0.000	0.000	.000	0.000	0.000
3	0.000	.000	.067	.011	.001	.002	0.000	0.000	.000	.001	.003	0.000	0.000	0.000	.011	0.000	0.000
4	0.000	.000	.013	.091	.006	.005	0.000	.001	0.000	0.000	.001	0.000	0.000	0.000	.005	0.000	0.000
5	0.000	.001	.001	.028	.068	.013	0.000	.001	0.000	.005	.001	0.000	.001	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	.000	.005	.041	.003	.005	.000	.006	.001	0.000	0.000	0.000	.001	.002	.000
7	.000	0.000	.000	.004	.003	.011	.025	.003	.000	.001	.000	0.000	0.000	0.000	0.000	.004	0.000
8	0.000	0.000	.000	.000	.001	.006	.002	.026	0.000	.000	0.000	0.000	0.000	0.000	.000	.012	0.000
9	0.000	0.000	.000	.001	0.000	.007	.001	0.000	.049	.017	.002	0.000	0.000	.000	.000	.001	.001
10	0.000	0.000	.000	.003	.001	.010	0.000	0.000	.025	.093	.019	0.000	0.000	0.000	0.000	.000	0.000
11	0.000	0.000	.001	.002	.001	.003	0.000	0.000	.001	.003	.044	.001	.002	0.000	.000	0.000	0.000
12	0.000	0.000	0.000	.000	.000	.001	0.000	.000	.000	.006	.012	.001	.001	.000	0.000	0.000	0.000
13	0.000	0.000	.000	.000	.000	.000	0.000	.000	.001	.002	.001	.000	.002	0.000	0.000	0.000	.000
14	0.000	0.000	.000	.000	0.000	.000	.000	.000	.002	.001	.000	.000	.000	.002	0.000	.000	0.000
15	.000	0.000	.010	.001	0.000	.002	.000	0.000	0.000	.000	0.000	0.000	0.000	0.000	.034	0.000	0.000
16	0.000	.000	0.000	0.000	.000	.003	0.000	.001	0.000	.001	.001	0.000	0.000	0.000	0.000	.055	.000
17	0.000	0.000	.000	.000	.000	.001	.000	.000	.001	.000	.000	.000	.001	0.000	0.000	.002	.070
TOTAL	.004	.008	.095	.148	.091	.113	.031	.059	.082	.142	.080	.003	.006	.002	.052	.077	.071

PREDICTED SHOP
TRIP INTERCHANGE MATRIX

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0.000	0.000	.002	.004	.003	.006	.002	.002	.001	.003	.002	0.000	0.000	0.000	.001	.002	0.000
2	0.000	0.000	.007	.012	.008	.010	.003	.003	.003	.006	.006	0.000	0.000	0.000	.003	.004	0.000
3	0.000	0.000	.014	.019	.010	.012	.004	.003	.004	.009	.013	0.000	0.000	0.000	.006	.006	0.000
4	0.000	0.000	.017	.029	.015	.019	.006	.005	.006	.013	.015	0.000	0.000	0.000	.007	.008	0.000
5	0.000	0.000	.008	.016	.012	.016	.005	.004	.005	.009	.009	0.000	0.000	0.000	.004	.006	0.000
6	0.000	0.000	.006	.011	.009	.017	.005	.004	.004	.009	.007	0.000	0.000	0.000	.002	.006	0.000
7	0.000	0.000	.003	.005	.004	.008	.003	.003	.002	.004	.004	0.000	0.000	0.000	.001	.004	0.000
8	0.000	0.000	.004	.007	.005	.010	.004	.004	.003	.005	.005	0.000	0.000	0.000	.002	.007	0.000
9	0.000	0.000	.004	.006	.006	.018	.004	.003	.004	.007	.006	0.000	0.000	0.000	.002	.005	0.000
10	0.000	0.000	.010	.017	.012	.015	.006	.005	.007	.017	.016	0.000	0.000	0.000	.004	.008	0.000
11	0.000	0.000	.005	.009	.005	.008	.002	.002	.003	.007	.009	0.000	0.000	0.000	.002	.003	0.000
12	0.000	0.000	.002	.004	.003	.004	.001	.001	.002	.013	.003	0.000	0.000	0.000	.001	.002	0.000
13	0.000	0.000	.002	.004	.003	.004	.001	.001	.002	.004	.004	0.000	0.000	0.000	.001	.002	0.000
14	0.000	0.000	.002	.004	.003	.004	.002	.001	.002	.003	.003	0.000	0.000	0.000	.001	.002	0.000
15	0.000	0.000	.008	.010	.004	.007	.002	.002	.002	.005	.005	0.000	0.000	0.000	.004	.003	0.000
16	0.000	0.000	.001	.002	.002	.003	.001	.001	.001	.002	.001	0.000	0.000	0.000	.001	.003	0.000
17	0.000	0.000	.000	.001	.001	.001	.000	.000	.000	.001	.001	0.000	0.000	0.000	.000	.001	0.000
TOTAL	0.000	0.000	.096	.160	.107	.153	.051	.046	.052	.107	.106	0.000	0.000	0.000	.043	.073	0.000

BIBLIOGRAPHY

Acton, F.S. (1970): Numerical Methods That Work, Harper and Row, New York.

Alonso, W. (1960): "A theory of the urban land market", Papers of the Science Association, 6, 149-157.

Alonso, W. (1964): Location and Land Use, Harvard University Press.

Angel, S. and Hyman, G.M. (1976): Urban Fields A Geometry of Movement for Regional Science, Pion Ltd.

Batty, M. (1971): "Design and construction of a sub-regional land use model", Socio-Economic Planning Sciences, 5, 97-124.

Batty, M. (1976): Urban Modelling, Cambridge U. Press, Cambridge.

Batty, M. (1978): "Urban Models in the Planning Process" in D.T. Herbert and R.J. Johnston (eds.) Geography and the Urban Environment, Vol. 1, Wiley, Chichester, pp. 63-134.

Batty, M. and Mackie, S. (1972): "The calibration of gravity, entropy and related models of spatial interaction", Environment and Planning, A, 205-233.

Batty, M., Foot, D.H.S., Alonso, L., Bray, G., Breheny, M., Constable, D., Dugmore, K., Ellender, R., Shepherd, J. and Williams, J. (1973): "Spatial Systems Design and Fast Calibration of Activity Interaction-Allocation Models", Regional Studies, 7, 351-366.

Batty, M. and March, L. (1976): "The method of residues in urban modelling", Environment and Planning, 8, 189-214.

Beaumont, J. (1978): "Spatial Interaction Models and the Location-Allocation Problem", Working Paper 240, School of Geography, University of Leeds, Leeds. Urban Doc. Centre #91.

Berry, B.J.L. (1967): Geography of Market Centres and Retail Distribution, Prentice-Hall, Englewood Cliffs, New Jersey.

Brillouin, L. (1962): Science and Information Theory, Academic Press, New York.

Broadbent, T.A. (1970): "Notes on the design of operational models", Environment and Planning, 2, 469-476.

Broyden, C.G. (1965): "A class of methods for solving non-linear simultaneous equations", Math. Comp., 19, 577-593.

Carrothers, G.A.P. (1956): "An historical review of the gravity and potential concepts of human interaction", JAIP, 22, 94-102.

Cesario, F. (1973): "A note on the entropy model of trip distribution", Transportation Research, 7, 331-333.

Cesario, F. (1975): "Linear and non-linear models of spatial interaction", Economic Geography, 51, 69-77.

Cesario, F. (1975): "A primer on entropy modelling", JAIP, 41, 40-48.

Cesario, F.J. (1977): "A New Interpretation of the 'Normalizing' or 'Balancing' Factors of Gravity-Type Spatial Models", Socio-Econ. Plan. Sci., 11, 131-136.

Chapin, F.S. and Weiss, S.F. (1962): Factors Influencing Land Development, Institute for Research in Social Science, University of North Carolina, Chapel Hill, North Carolina.

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

Christaller, W. (1933, translated 1966): Central Places in Southern Germany, Prentice-Hall, Englewood Cliffs, New Jersey.

Coelho, J.D. and Wilson, A.G. (1975): Some Equivalence Theorems to Integrate Entropy Maximizing Submodels within Overall Mathematical Programming Frameworks, Working Paper III, Department of Geography, University of Leeds.

Cordey-Hayes, M. and Wilson, A.G. (1971): "Spatial Interaction", Socio-Economic Planning Sciences, 6, 365-385.

Cox, R.T. (1961): The Algebra of Probable Inference, John Hopkins, Baltimore.

Cripps, E.L. and Foot, D.H.S. (1970): "The urbanization effects of a third London airport", Environment and Planning, 2, 153-192.

Echenique, M., Crowther and Lindsay, W. (1969): "A spatial model of urban stock and activity", Regional Studies, 3, 281-312.

Evans, R.A. (1969): "The principle of minimum information", I.E.E.E. Transactions on Reliability, R18, 3, 87-90.

Evans, S.P. (1973): "A Relationship Between the Gravity Model for Trip Distribution and the Transportation Problem in Linear Programming", Transp. Res., 7, 39-61.

Freund, J.E. (1979): Modern Elementary Statistics, Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Garin, R.A. (1966): "A matrix formulation of the Lowry model for intrametropolitan activity location", JAIP, 32, 361-364.

Georgescu-Roegen (1971): The Entropy Law and the Economic Process, Harvard University Press, Cambridge.

Gerald, C.F. (1973): Applied Numerical Analysis, Addison-Wesley, Massachusetts.

Goldner, W. (1971): "The Lowry model heritage", JAIP, 37, 100-110.

Good, I.J. (1956): "Some terminology and notation in information theory", Proceedings IEE, Part c, 103, 200-204.

Gould, P. (1972): "Pedagogic Review", AAAG, 62, 689-700.

Greenhut, M. (1963): Microeconomics and the Space Economy, Scott, Foresman, Chicago.

Haggett, P. (1965): Locational Analysis in Human Geography, MacMillan, Toronto.

Hall, P.D. (1975): "An Assessment of the Calibration of Spatial Interaction Models", Unpublished M.A. Thesis, Department of Geography, McMaster University, Hamilton, Ontario.

Hamburg, J.R. and Creighton, R.I. (1959): "Predicting Chicago's Land Use Pattern", JAIP, 25(2), 67-72.

Hansen, W.G. (1959): "How Accessibility Shapes Land Use", JAIP, 25, 73-76.

Harris, B. (1961): "Some Problems in the Theory of Intra-Urban Location", Operations Research, 9, 695-721.

Harris, B. (1962): "Linear programming and the projection of land uses", Paper 20, Penn-Jersey Transportation Study, Philadelphia.

Harris, B. (1966): "The Uses of Theory in the Simulation of Urban Phenomena", HRR, 126, 1-16.

Harris, B. (1968): "Quantitative models of urban development: their role in Metropolitan Policy making", in Issues in Urban Economics, H.S. Perloff and L. Wingo, Jr. (Eds.), John Hopkins, Baltimore, 363-412.

Hathaway, P.J. (1975): "Trip distribution and disaggregation", Environment and Planning A, 7, 71-97.

Herbert, J. and Stevens, B.H. (1960): "A model for the distribution of residential activity in urban areas", Journal of Regional Science, 2, 21-36.

Hobson, A. (1969): "A new theorem of information theory", Journal of Stat. Physics, 1, 383-391.

Hobson, A. and Cheng, B.K. (1973): "A comparison of the Kullback and Shannon information measures", Journal of Stat. Physics, 7, 301-310.

Hoover, E.M. (1948): The Location of Economic Activity, McGraw-Hill, New York.

Huff, D.L. (1963): "A probabilistic analysis of shopping center trade areas", Land Economics, 39, 81-90.

HWPDD (1975): "Transportation: A Substudy of the Regional Official Plan", Report Number 2, Hamilton, October, 1975.

Isard, W. (1956): Location and the Space Economy, John Wiley, New York.

Jarratt, P. (1970): "A Review of Methods for solving non-linear algebraic equations in one variable", in P. Rabinowitz (ed.) Numerical Methods for Non-Linear Algebraic Equations, Gordon and Breach, London, 1-26.

Jaynes, E.T. (1957a): "Information theory and statistical mechanics I", Phys. Review, 106, 620-630 and (1957b): "Information theory and statistical mechanics II", 108, 171-190.

Kain, J.F. (1962): "The journey-to-work as a determinant of residential locations", Papers of R.S.A., 9, 137-160.

Khinchin, A.J. (1957): Mathematical Foundations of Information Theory, Dover, New York.

Kullback, S. (1959): Information Theory and Statistics, Wiley, New York.

Lakshmanan, T.R. and Hansen, W.G. (1965): "A retail market potential model", J.A.I.P., 31, 134-143.

Lee, C. (1973): Models in Planning, Pergamon Press, Oxford.

Lee, D.B. (1973): "Requiem for Large Scale Models", J.A.I.P., 39, 163-178.

Levine, R.D. and Tribus, M. (eds.)(1979): The Maximum Entropy Formalism, M.I.T. Press, Cambridge, Mass.

Losch, A. (1944, translated 1954): The Economics of Location, Yale University Press, New Haven, Connecticut.

Lowry, I.S. (1964): A Model of Metropolis, Rand Corporation, Santa Monica.

Lowry, I.S. (1966): Migration and Metropolitan Growth: Two Analytical Models, Chandler, San Fransisco.

Lowry, I.S. (1968): "Seven models of urban development: a structural comparison", Special Report 97, Highway Research Board, Washington, D.C., 121-145.

Macgill, S.M. (1975): Balancing Factor Methods in Urban and Regional Analysis, Working Paper 124, Department of Geography, University of Leeds.

Macgill, S.M. (1975): "A convergence theorem for the multipliers in biproportional matrices", working paper 113, Department of Geography, University of Leeds.

Macgill, S.M. (1977): "Theoretical properties of biproportional matrix adjustments", Environment and Planning A, 9, 687-701.

Macgill, S.M. and Wilson, A.G. (1978): "Equivalences and Similarities Between Some Alternative Urban and Regional Models", Working Paper 239, School of Geography, University of Leeds, Leeds. Urban Doc. Centre #107.

March, L. (1969): "Urban Systems: A Generalised Distribution Function", Working Paper 24, Land Use and Built Form Studies, University of Cambridge, Cambridge.

Mathai, A.M. and Rathie, P.N. (1975): Basic Concepts in Information Theory and Statistics, Wiley (New Delhi).

McLoughlin, B. (1969): Urban and Regional Planning: a systems approach, Faber, London.

Mitchell, R. and Rapkin, C. (1954): Urban Traffic: A Function of Land Use, Columbia U.P., New York.

Mogridge, M.J.H. (1972): "The use and misuse of entropy in urban and regional modelling of economic and spatial systems", working paper 80, Centre of Environmental Studies, London.

Morphet, R. (1975): "A note on the calculation and calibration of doubly constrained trip distribution models", Transportation, 4, 43-53.

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

Olsson, G. (1965): Distance and Human Interaction: A Review and Bibliography, R.S.R.I., Philadelphia.

O'Kelly, M.E. (1978): "Empirical tests of an entropy maximizing model of retail location and consumer behaviour", unpublished research paper, Department of Geography, McMaster University, Hamilton.

O'Kelly, M.E. (1980): "A Model of the Demand for Retail Facilities Incorporating Multi-stop, Multi-purpose trips", unpublished manuscript, Department of Geography, McMaster University, Hamilton.

Openshaw, S. (1976): "An empirical study of some spatial interaction models", Environment and Planning A, 8, 23-41.

Papageorgiou, G.J. (Ed.)(1976): Mathematical Land Use Theory, D.C. Heath, Lexington, Massachusetts.

Pitfield, D.E. (1978): "Algorithm 6: The χ^2 test for predicted trip matrices, Environment and Planning A, 10, 1201-1206.

Powell, M.J.D. (1970): "A hybrid method for non-linear equations" in P. Rabinowitz (ed.), Numerical Methods for Non-Linear Algebraic Equations, Gordon and Breach, London, 87-150.

Putman, S. (1975): "Urban Land Use and Transportation Models: A State-of-the-Art Summary", Transportation Research, 9, 187-202.

Quandt, R.E. (1964): "Statistical Discrimination Among Alternative Hypotheses and Some Economic Regularities", Journal of Regional Science, V.5, 2, 1-23.

Ravenstein, E.G. (1885): "The Laws of Migration", Journal of the Royal Statistical Society, 52, 241-305.

Senior, M.L. (1973): "Approaches to residential location modelling 1: urban ecological and spatial interaction models", Environment and Planning, 5, 165-197.

Senior, M.L. (1974): "Approaches to residential location modelling 2: urban economic models and some recent developments", Environment and Planning A, 6, 369-409.

Senior, M.L. (1976): "The use of the entropy-maximizing methodology for model building in geography", Working Paper 144, Department of Geography, University of Leeds.

Senior, M.L. (1977): Chapter 6: Residential location, in A.G. Wilson, P.H. Rees, and C.M. Leigh (eds.)(1977): Models of Cities and Regions, Theoretical and Empirical Developments, John Wiley and Sons, Chichester.

Shannon, C.E. (1948): "A mathematical theory of communication", Bell System Technical Journal, 27, 379-423 and 623-656.

Snickars, F. and Bull, W.E. (1977): "A minimum information principle, theory and practice", Reg. Sci. and Urb. Econ., 7, 137-168.

Stillwell, J.C.H. (1975): "Models of Interregional Migration: A Review", Working Paper No. 100, Department of Geography, University of Leeds.

Stillwell, J.C.H. (1977): "Some historical tests of spatial interaction models for inter-area migration: Part II: Inter-region migration", Working Paper 199, Department of Geography, University of Leeds.

Stillwell, J.C.H. (1978): "Interzonal Migration: Some Historical tests of Spatial-Interaction Models", Environment and Planning A, 10, 1187-1200.

Stouffer, S.A. (1940): "Intervening Opportunities: A theory relating mobility and distance", American Sociological Review, 5, 845-867.

Tribus, M. (1969): Rational Descriptions, Decisions and Designs, Pergamon, Oxford.

Tribus, M. and Rossi, R. (1973): "On the Kullback information measure as a basis for information theory: Comments on a proposal by Hobson and Chang", Journal of Statistical Physics, 9(4), 331-338.

Voorhees, A.M. (1955): "A General Theory of Traffic Movement", Proceedings of the Institute of Traffic Engineering, 1, 46-56.

Walsh, J.A. and Webber, M.J. (1977): "Information theory: some concepts and measures", Environment and Planning A, 9, 395-417.

Webber, M.J. (1975): "Entropy maximizing location models for non-independent events", Environment and Planning A, 7, 99-108.

Webber, M.J. (1976a): "The Meaning of Entropy Maximizing Models", in G.J. Papageorgiou (ed.): Mathematical Land Use Theory, Lexington (Heath), 277-292.

Webber, M.J. (1976b): "Elementary Entropy Maximizing Probability Distributions: Analysis and Interpretation", Economic Geography, Vol. 52, 218-227.

Webber, M.J. (1977): "Pedagogy Again: What is Entropy?", Annals, A.A.G., 67, 254-266.

Webber, M.J., O'Kelly, M.E., and Hall, P.D. (1978): "Empirical Tests on an information-minimizing model of consumer characteristics and facility location", a paper presented at the 74th Annual Meeting of the A.A.G., New Orleans, April, 1978.

Webber, M.J., O'Kelly, M.E., Hall, P.D., Munro, D., and Beeson, D. (1979a): "A realistic information-minimizing model", Modelling and Simulation, 10, Part 4, Socio-Economic Systems, Proceedings of the Tenth Annual Pittsburgh Conference.

Webber, M.J., O'Kelly, M.E., and Hall, P.D. (1979b): "Empirical tests on an information-minimizing model of consumer characteristics and facility location", Ontario Geography, 13, 61-80.

Webber, M.J. (1979c): Data Description, unpublished manuscript, Department of Geography, McMaster University, Hamilton.

Weber, A. (1909, translated 1929): On the Location of Industries, University of Chicago Press, Chicago.

Williams, H.C.W.L. and Wilson, A.G. (1979): Some Comments on the Theoretical and Analytic Structure of Urban and Regional Models, Working Paper 258, School of Geography, University of Leeds.

Wilson, A.G. (1967): "A Statistical Theory of Spatial Distribution Models", Transp. Res. 1, 253-269.

Wilson, A.G. (1969): "The use of entropy maximizing models in the theory of trip distribution, mode split and route split", J. Transp. Econ. Policy, 3, 108-126.

Wilson, A.G. (1970a): Entropy in Urban and Regional Modelling, Pion, London.

Wilson, A.G. (1970b): "Inter-regional commodity flows: entropy maximizing approaches", Geog. Anal., 2, 255-282.

Wilson, A.G. (1970c): "Advances and Problems in Distribution Modelling", Transp. Res. 4, 1-18.

Wilson, A.G. (1971): "A family of spatial interaction models, and associated developments", Env. and Plan., 3, 1-32.

Wilson, A.G. (1973): "Further developments of entropy maximizing transport models", Trans. Plan. and Tech., 1, 183-193.

Wilson, A.G. (1974a): Urban and Regional Models in Geography and Planning, Wiley, London.

Wilson, A.G. (1974b): Some New Forms of Spatial Interaction Model: A Review, Working Paper 55, Department of Geography, University of Leeds.

Wilson, A.G. (1979): Criticality and Urban Retail Structure: Aspects of Catastrophe Theory and Bifurcation Working Paper 241, School of Geography, University of Leeds.

Wilson, A.G. (1979): Theory in Human Geography: A Review Essay, Working Paper 253, School of Geography, University of Leeds.

Wilson, A.G., Rees, P.H., and Leigh, C.M. (eds.)(1977): Models of Cities and Regions, Theoretical and Empirical Developments, John Wiley and Sons, Chichester.