AN ASSESSMENT OF THE CALIBRATION OF SPATIAL INTERACTION MODELS

AN ASSESSMENT OF THE CALIBRATION OF SPATIAL INTERACTION MODELS

by

PETER DOUGLAS HALL

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Arts

McMaster University

December 1975

MASTER OF ARTS (1975)
(Geography)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: An Assessment of the Calibration of Spatial Interaction Models

AUTHOR: Peter Douglas Hall, B.A. (McMaster)

SUPERVISOR: Dr. F. L. Hall

NUMBER OF PAGES: viii, 124

SCOPE AND CONTENTS:

This paper is concerned with assessing the procedures used in calibrating spatial interaction models. It critically reviews calibration methodologies which have been proposed in the literature and determines that the statistical estimation techniques of maximum likelihood and least-squares are particularly suited to this estimation problem.

The calibration statistics from the maximum likelihood and least-squares estimators are developed from first principles and special note is made of the behavioral assumptions implicit in each.

Two issues are then reviewed: the reliability of the random sample in representing the mean distribution of trips, and the definition of variables in calibration statistics. A hypothetical framework is proposed, within which an examination of these issues is made.

The study results indicate that the sample reasonably represents the mean distribution and also that the incorporation of implicit behavioral assumptions does not necessarily result in better model predictions.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Fred Hall for his encouragement and continued guidance throughout the course of this research. His suggestions and comments have improved the content of this paper. The assistance of Dr. Michael Webber and Dr. Lee Liaw is also gratefully acknowledged, as is that of Ruth Fincher for suggesting many improvements to the manuscript. Initial discussions with Michael Batty stimulated my research in this area, and his assistance in providing the calibration programs is deeply appreciated.

TABLE OF CONTENTS

		Page
	SCOPE AND CONTENTS	ii
	ACKNOWLEDGEMENTS	iii
	TABLE OF CONTENTS	iv
	LIST OF FIGURES	vi
	LIST OF TABLES	vii
CHAPTER		
1	AN ASSESSMENT OF CALIBRATION METHODOLOGIES	1
	INTRODUCTION	1
	SPATIAL INTERACTION MODELS	3
	APPROACHES TO MODEL CALIBRATION	6
	GRAPHICAL CURVE FITTING	6
	TABULATION	8
	SYSTEMATIC SEARCH	11
	DIRECT EVALUATION	12
	INDIRECT EVALUATION	16
	STATISTICAL ESTIMATORS	18
	SUMMARY	21
2	THE MAXIMUM LIKELIHOOD METHOD OF PARAMETER ESTIMATION	24
	INTRODUCTION	24
	HYPOTHESIS EVALUATION AS A METHOD OF PARAMETER CALIBRATION: AN APPRAISAL OF HYMAN'S APPROACH	25
	THE MAXIMUM LIKELIHOOD ESTIMATOR	28

CHAPTER		Page
2	THE DEVELOPMENT OF CALIBRATION STATISTICS FROM THE METHOD OF MAXIMUM LIKELIHOOD	30
	EVANS APPROACH TO CALIBRATION	30
	KIRBY'S APPROACH TO CALIBRATION	35
	ASSUMPTIONS OF THE TWO DERIVATIONS	39
	GENERAL CONDITIONS REQUIRED BY THE MAXIMUM LIKELIHOOD ESTIMATOR	43
	SUMMARY	47
3	THE LEAST SQUARES METHOD OF PARAMETER ESTIMATION	48
	INTRODUCTION	48
	THE LEAST-SQUARES ESTIMATOR	49
	THE DEVELOPMENT OF CALIBRATION STATISTICS FROM THE METHOD OF LEAST-SQUARES: CESARIO'S APPROACH	51
	ASSUMPTIONS OF THE LEAST-SQUARES APPROACH AND A COMPARISON TO MAXIMUM LIKELIHOOD	54
	STATISTICAL CONDITIONS REQUIRED BY THE LEAST-SQUARES ESTIMATOR	61
	SUMMARY	62
4	SOME RELATED CALIBRATION PROBLEMS IN URBAN SYSTEMS MODELLING	64
	INTRODUCTION	64
	THE "PROBLEM" OF MODEL CALIBRATION	65
	THE TRAVEL SURVEY AND ITS RELATIONSHIP TO THE POPULATION IT REPRESENTS	67
	VARIABLE DEFINITION IN CALIBRATION STATISTICS	69
	A REVIEW OF PROBLEMS IN MODEL CALIBRATION	74
	A RESEARCH DESIGN TO EXAMINE THE ISSUES IN CALIBRATION	76

CHAPTER		Page
5	AN ANALYSIS OF THE RELATIONSHIP OF THE TRAVEL-SURVEY TO THE MEAN TRAVEL DISTRIBUTION, AND OF THE EFFECT OF BEHAVIORAL CONSIDERATIONS ON MODEL PERFORMANCE	79
	INTRODUCTION	79
	THE RELATIONSHIP BETWEEN OPTIMUM FIT AND OPTIMUM PARAMETER VALUES	82
	THE PHYSICAL SETTING FOR THE ANALYSIS	85
	A DESCRIPTION OF THE ALGORITHMS IN THE ANALYSIS	89
	SYSTEM SIMULATION	89
	CORRESPONDENCE MEASURES	92
	THE RELIABILITY OF A SINGLE TRIP SURVEY IN MODEL CALIBRATION	94
	A COMPARISON OF MODEL PREDICTIONS GENERATED BY THE MAXIMUM LIKELIHOOD AND LEAST-SQUARES ESTIMATORS	102
	FACTORS IN MODEL APPLICATION WHICH REDUCE THEORETICAL CALIBRATION REQUIREMENTS	108
	SUMMARY	111
6	SUMMARY AND CONCLUSIONS	113
	BIBLIOGRAPHY	121

LIST OF FIGURES

FIGURE		Page
1	Grid Search in a Two Parameter Space	10
2	The Response Surface of the Coefficient of Determination (R^2) Against the Parameter β	13
3	Sample Trip Matrix	19
4	Configuration and Zoning of the Idealized Urban Areas	86
5	Sequence of Operations in the Analysis	90
6	Scaled-Up Sample Predictions and the Mean Distribution of Trips Generated from Zone Twenty-Five	99
7	The Deviation of Scaled-Up Poisson and Normal Samples from the Mean Trip Distribution Generated from Zone Twenty-Five	101
8	Comparison of Maximum Likelihood and Least-Squares Predictions with Poisson Sample of Trips Generated from Zone Twenty-Five	104
9	Comparison of Maximum Likelihood and Least-Squares Predictions with Normal Sample for Trips Generated from Zone Twenty-Five	105
10	Scaled-Up Model Predictions as Calibrated by Maximum Likelihood and the Mean Trip Distribution Generated from Zone Twenty-Five	107

LIST OF TABLES

TABLE		Page
1	TRIP ORIGINS AND DESTINATIONS ALLOCATED TO EACH ZONE	88
2	THE MEAN, POISSON AND NORMAL TRAVEL PATTERN GENERATED FROM ZONE TWENTY-FIVE	97
3	R^2 BETWEEN MODELS ESTIMATED BY MAXIMUM LIKELIHOOD OR LEAST-SQUARES, AND THE SAMPLE DATA USED TO CALIBRATE THE MODEL	102
4	MEASURES OF CORRESPONDENCE AND PARAMETER VALUES OF MODELS ESTIMATED BY MAXIMUM LIKELIHOOD AND LEAST-SQUARES	110

CHAPTER 1

AN ASSESSMENT OF CALIBRATION METHODOLOGIES

INTRODUCTION

The calibration of mathematical models involves finding the best (in some well defined sense) values of their parameters. Calibration transforms the general model structure into a set of exact empirically tested relationships by giving precise empirical definitions to the variables and numerical values to the parameters. A model is calibrated to improve its predictive or descriptive capability. The theoretical principles used to develop the model are seldom sufficient to indicate more than the appropriate sign and probable order of magnitude of the model parameters. Since the parameters are measures of the relationships between numerical variables, the precise empirical definition of these variables affects the parameter values (Lowry, 1965, p. 163).

Mackie (1972, p. 39) identifies three components of the calibration process:

(1) Specification of the type of model to be calibrated;

(2) Selection of a suitable statistic to optimize, which yields the "best" parameter estimates, and

(3) Selection of an accurate and efficient technique to solve the equations derived by the statistic.

This paper is concerned with the calibration of a particular type of geographic model: *the spatial interaction model.* The model has been used as a trip distribution sub-model in travel forcasting studies, and, operating within the Lowry framework, as part of the large-scale modelling efforts in Britain (Batty, 1970c, p. 95; Batty, 1972, p. 152).

The paper will investigate the calibration of the doubly-constrained spatial interaction model, which will be outlined later. Therefore, Mackie's first component of the calibration process is defined. Batty (1970c, p. 114) emphasizes the need for better calibration statistics to measure the model's goodness-of-fit, so that a unique set of parameter values can be derived (Mackie's second component). Also important is the development of more efficient and faster numerical methods of solution without a loss of accuracy, which is Mackie's third component. Although several solution techniques will be reviewed in this paper, its primary task will be to investigate Mackie's second component of the calibration process: *the selection of suitable statistics*, the optimization of which will yield the "best" parameter estimates.

In this chapter, a brief description of the spatial interaction model will be followed by a review and assessment of the various approaches to calibration which have been undertaken, particularly in the British context. From the assessment, two significant calibration approaches will be identified.

Chapters two and three will explore these two calibration approaches and examine the assumptions implicit in each.

Chapter four will critically assess the general problem of calibration as it is applied to urban systems modelling, and stress the areas of weakness. From this review, a research design will be proposed and tested in chapter five.

Finally, chapter six will summarize the research findings of this paper and evaluate their significance.

SPATIAL INTERACTION MODELS

Spatial interaction models are a family of models which describe the interaction between sets of activities in terms of flows of people or commodities. The equation which describes this, the gravity law, was originally applied to the geographic field by analogy to Newtonian mechanics. It states that the intensity of interaction between two zones i and j, is a function of the population masses at i and j, and of the impedance to interaction, measured by an inverse function of distance. Wilson (1970) develops a general theoretical derivation of the gravity model from the fields of statistical mechanics and information theory. The gravity model is derived by analogy to principles in statistical mechanics by finding the most probable distribution of trips, subject to a set of constraints placed upon the system, which restrict the number of assignments giving rise to a distribution. The same model can be derived from information theory by defining the entropy of a system to be a measure of its uncertainty (Shannon and Weaver, 1949). The probability distribution which results from maximizing the entropy, subject to whatever information is known about the system (the constraints), is minimally biased yet maximally non-committal with regard to missing information (Webber, 1975, p. 14). Wilson (1970, p. 8) shows that the distribution derived from the information theoretic approach is equivalent to the most probable distribution derived from principles of statistical mechanics.

From the general formulation of the spatial interaction model

$$\hat{t}_{ij} = a_i b_j \, f(\beta, c_{ij}) \tag{1.1}$$

where $\hat{t}_{ij}$ = the predicted number of trips from i to j,

a_i = a factor related to the ability of zone i to generate trips,

b_j = a factor related to the ability of zone j to attract trips,

$f(\beta, c_{ij})$ = the impedence to interaction,

four variations may be derived, depending on the constraints imposed upon the distribution: (1) *unconstrained* flows; (2) *production constrained* flows; (3) *attraction constrained* flows; and (4) *production-attraction constrained* flows.

The unconstrained model is simply equation (1.1). There are no constraints on the distribution, and the model estimates the number of interchanges between each zone pair, the $\hat{t}_{ij}$, the number of origins $\sum_j t_{ij}$, and destinations, $\sum_i \hat{t}_{ij}$, within the framework. Both the production constrained and attraction constrained models are examples of *singly* constrained spatial interaction models. In this case, the $\hat{t}_{ij}$ are subject to the constraint

$$\sum_j \hat{t}_{ij} = O_i \quad \text{where} \quad O_i = a_i/A_i$$

for the production constrained model, or

$$\sum_i \hat{t}_{ij} = D_j \quad \text{where} \quad D_j = b_j/B_j$$

for the attraction constrained model. The models estimate the $\hat{t}_{ij}$ and the $\sum_i \hat{t}_{ij}$ (destinations), or the $\sum_j \hat{t}_{ij}$ (origins), depending on whether the distribution is production or attraction constrained. The production-attraction constrained model is a *doubly* constrained interaction model, as the $\hat{t}_{ij}$ are subject to both of the above constraints. Since both the $\sum_j \hat{t}_{ij}$ and $\sum_i \hat{t}_{ij}$ are estimated externally, only the $\hat{t}_{ij}$ are estimated by the model.

The doubly constrained model is of interest for two reasons. First, this variation of the spatial interaction model is generally used for predicting the distribution of trips in transportation studies (Mackie, 1972, p. 27). For this, the calibration procedure is of some practical significance. Secondly, the inclusion of constraints makes it more difficult to calibrate the model (Mackie, 1972, p. 24).

The unconstrained gravity model can be calibrated by transforming the equation into logarithmic form and estimating the parameters by regression techniques (Olsson, 1965, p. 37), although Siedmann (1969) stresses the problems of this approach. The singly and doubly constrained models on the other hand, because of their intrinsically non-linear character, cannot be linearized by a simple transformation (Draper and Smith, 1966, p. 264), and thus require more sophisticated calibration techniques. Any attempt at linearization, i.e., truncating a Taylor's series expansion at the first order, may lead to biased parameter estimates (Batty and Mackie, 1972, p. 209). Presumably, a calibration procedure developed for a doubly constrained model should be applicable to both singly constrained and unconstrained models.

APPROACHES TO MODEL CALIBRATION

The most important task in application is to calibrate the model so that the most realistic distribution is generated, or so the model "best fits" the survey data collected. Batty (1972, p. 156) notes some related calibration problems that have arisen in model application in Britain and emphasizes the importance of this aspect of design.

Because of the singly and doubly constrained models' inherent nonlinear character, the model parameters have been estimated by several different methods. Specifically, four different approaches to the calibration problem can be identified. Early attempts include *graphical curve fitting* and *tabulation* methods, while more recent work has employed *systematic search* algorithms and *statistical estimators.* An outline of each of these approaches follows.

GRAPHICAL CURVE FITTING

Initial attempts at calibration can be seen in the work of Lowry (1963). Although the allocation sub-models used are not the Wilson-type spatial interaction models, the potential models used by Lowry can be related to the gravity model (Isard, 1960), and thus the calibration problem is much the same. Lowry estimates the model parameters outside the framework of the model by approximating frequency functions to empirical data manually (Lowry, 1963; Reif, 1973, p. 181). He takes data on the relative frequency of work-trips by distance, disaggregated to different socio-economic classes, and finds the distributions to closely approximate a negative power function, i.e.,

$$f(r) = ar^{-x} \tag{1.2}$$

where r = distance from the origin zone

$f(r)$ = the relative trip frequency

a, x = parameters to be estimated.

The parameter values derived by Lowry are given by Reif (1973, p. 181). Lowry's calibration technique, then, is simply a graphical curve-fitting procedure, in which the parameter values are derived so the hypothesized function best fits the given data. The trip distribution index is obtained from the point density function:

$$G = \frac{f(r)}{2\pi r} = \frac{1}{t_{ij}} \tag{1.3}$$

Thus,

$$t_{ij} = \frac{2\pi r}{ar^{-x}} = \frac{2\pi}{a}\, r^{(x+1)} \tag{1.4}$$

The trip distribution elements, $\{t_{ij}\}$, are then substituted into the potential sub-model of locational choice. The entire process is repeated to derive service locations using another frequency function (Reif, 1973, pp. 180-185).

The major weakness of Lowry's approach to calibration is that because the parameters are estimated outside the model framework, the interdependencies between the parameter values and the model are ignored (Batty, 1972, p. 164). Lowry (1965, p. 163) acknowledges the dependence of the parameters on the model variables (see page 1) but does not incorporate this into his calibration technique.

TABULATION

Subsequent British work (Batty, 1970a, 1970b; Cripps and Foot, 1969a, 1969b; Turner and Williams, 1970; Masser, 1970) utilizes the tabulation method to calibrate the parameter values. This approach involves the testing of different combinations of parameter values, which are fixed within some predetermined range. If one assumes that a unique optimum exists, then the search is for that combination of parameter values which yields a best fit to the data. The correspondence between the predicted and observed sets of trips can be measured by various statistics. Usually a *statistic of correlation*, such as the *coefficient of determination* is used (Batty and Mackie, 1973; Batty, Foot, *et al.*, 1973, p. 356), and the technique tests to find the maximum correlation between the predicted and observed values. For $\hat{t}_{ij}$ predicted and t_{ij} observed, the algorithm measures:

$$R^2 = 1 - \frac{\sum_i \sum_j (\hat{t}_{ij} - t_{ij})^2}{\sum_i \sum_j (t_{ij} - \frac{1}{N} \sum_i \sum_j t_{ij})^2} \tag{1.5}$$

where N equals the total number of variables. Wilson (1974, p. 317) states that the statistic varies between one, for an exact correspondence between the model predictions and the observations, and zero, for no correspondence. However, the range of the statistic is actually between one and $-\infty$. The denominator of (1.5) is simply the sum of squared deviations of the observation from its mean. If the predicted values, $\hat{t}_{ij}$, significantly differ from the observed values, t_{ij}, and the deviation of the observed values from the mean is small, the right hand term in (1.5) will be greater than one and negative values for R^2 will result.

An alternative statistic of correspondence which is commonly used is the *chi-squared statistic*, defined as:

$$\chi^2 = \sum_i \sum_j \frac{(t_{ij} - \hat{t}_{ij})^2}{t_{ij}} \tag{1.6}$$

The method searches for the value or combination of parameter values for which the statistic is at a minimum. Evans (1971, p. 25) notes that this statistic is a reasonable measure of the model's goodness-of-fit with the data. He postulates that the test statistic approximates a chi-squared distribution if the data does arise in the way postulated by the model and all the t_{ij}'s are reasonably large.

Other correlation statistics that have been tested in model calibration include the *root mean square error* statistic, defined by Hill, *et al.* (1965), and the *standard deviation* (Batty, 1970c, p. 104).

The tabulation method sets up a grid of combinations of parameter values; for example, if a two parameter interaction model is to be calibrated, a two dimensional grid of pairs of parameter values is constructed (Figure 1).

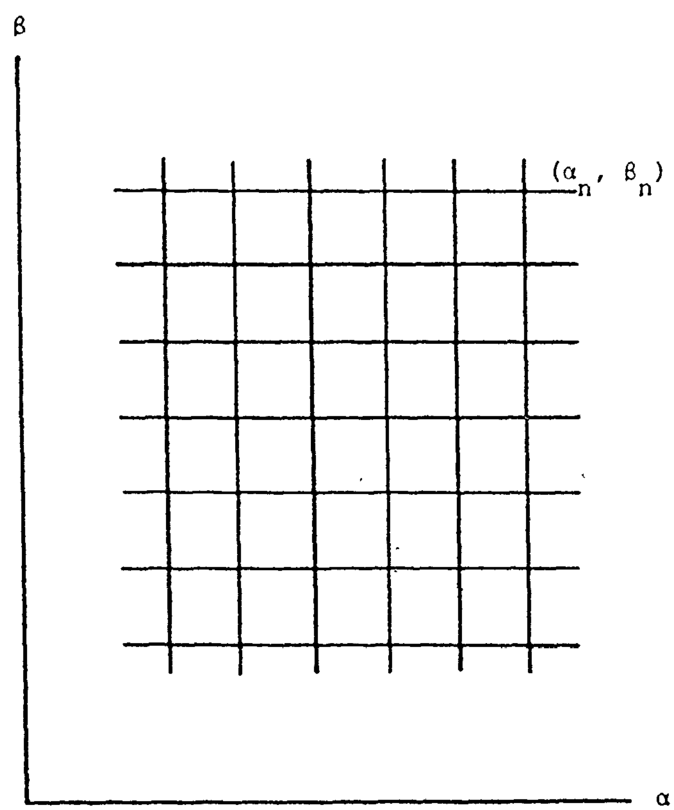

FIGURE 1: Grid Search in a Two Parameter Space.

Each node of the grid represents a particular parameter combination. The model is then run at each node within a predetermined range of parameter values. Alternatively, the nodes upon which the model is tested in the parameter space can be selected at random. More commonly, however, the nodes are chosen by trial and error, i.e., successive choice is made of the nodes in the parameter space which appear to be approaching the optimum (Mackie, 1972, p. 38). The distribution generated by that combination of parameter values which optimizes the test statistic yields the best fit to the survey data.

The principal draw-backs of this approach are that the method is slow, inefficient, and inaccurate (Mackie, 1972, p. 38). Since the correlation statistics require model output as variables, the model must be run for each combination of parameter values. To improve accuracy requires a vast number of tabulations to be performed. Referring to this, Batty (1971, p. 425) notes that computer time increases directly with the number of functional evaluations, which, for a model of x parameters, can be approximated by n^x, where n is the number of evaluations to be made in the specified parameter range.

SYSTEMATIC SEARCH

If, however, the test statistics are plotted over a range of parameter values, a *response surface* can be generated, measuring the model prediction's correlation to the data. Batty (1970c,p. 111) shows the regularity of these surfaces, and several authors (Batty and Mackie, 1972; Batty and Mackie, 1973; Batty, Foot, *et al.*, 1973; Batty *et al.*, 1974; Wilson, 1974) suggest how numerical methods could be applied, using the properties of the surface to quickly find that set of parameter values which gives the model predictions a best fit to the data. This technique, defined as systematic search (Batty and Mackie, 1972; Batty and Mackie, 1973) is simply a search algorithm, using standard mathematical optimization principles designed to calibrate the interaction models. The model is calibrated by optimizing a given test statistic, such as the coefficient of determination or chi-squared. Two different optimizing approaches have been employed. One is *direct evaluation* of the statistics' response surface and the other optimizes by *indirect evaluation.*

DIRECT EVALUATION

Direct evaluation methods use a set of direction vectors throughout the search, and explorations are made along these directions on the response surface. Subsequent action in directing the search is determined by the results obtained on the previous iteration. Direct evaluation can be based upon *linear methods*, such as the Newton-Raphson technique, Fibonacci sequences, and search by Golden Section, in all of which the direction vectors are *univariate*, or it can be based upon *quadratic methods*, which specify the optimum point by approximating the objective function by a *quadratic.* A numerical method of this type which has been used in model calibration is quadratic search by conjugate directions.

The mechanics of all these search procedures are illustrated by the method of Fibonacci sequences (Wilson, 1974, pp. 321-322). Consider a spatial interaction model of the form:

$$\hat{t}_{ij} = a_i b_j f(\beta, c_{ij}) \tag{1.7}$$

The model predictions, $\hat{t}_{ij}$, are functions of the parameter β. The response surface, which is simply a function of the t_{ij} observed and the $\hat{t}_{ij}$ predicted, is also a function of β. The test statistic, such as the coefficient of determination, will generally vary with β in the following manner (Figure 2).

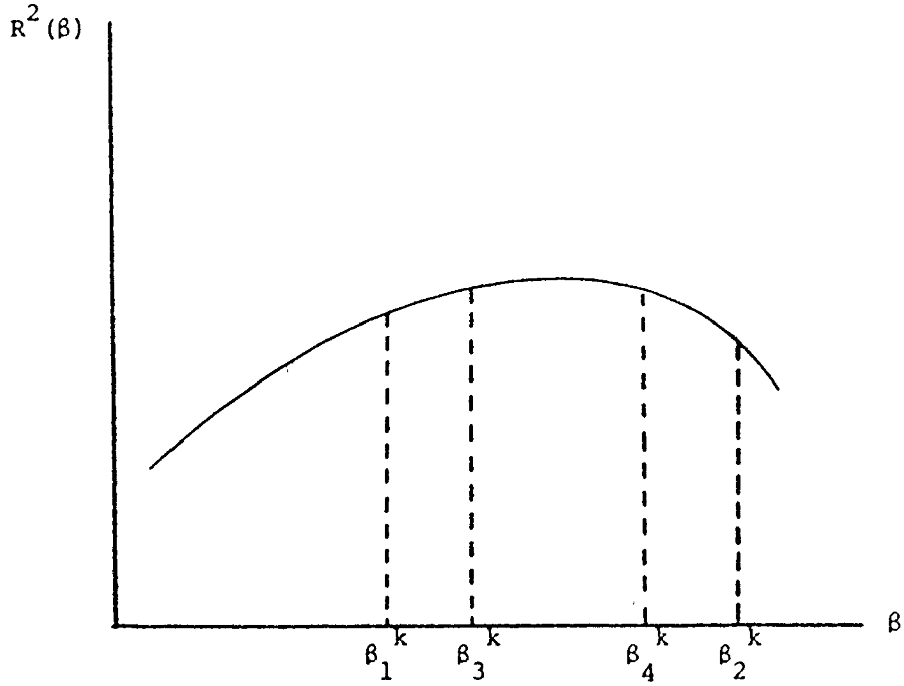

FIGURE 2: The Response Surface of the Coefficient of Determination (R^2) Against the Parameter β

(Source: Wilson, 1974, p. 321)

Therefore, the calibration task is the unconstrained optimization of the function, $R^2(\beta)$. If we assume that at the k^{th} step, the method has established that β lies within the values $\beta_1{}^k$ and $\beta_2{}^k$, the method then finds values for $\beta_3{}^k$ and $\beta_4{}^k$ such that,

$$\beta_1{}^k < \beta_3{}^k < \beta_4{}^k < \beta_2{}^k \tag{1.8}$$

by the equations,

$$\beta_3{}^k = \frac{F_{N-1-k}}{F_{N+1-k}} (\beta_2{}^k - \beta_1{}^k) + \beta_1{}^k \tag{1.9}$$

$$\beta_4{}^k = \frac{F_{N-k}}{F_{N+1-k}} (\beta_2{}^k - \beta_1{}^k) + \beta_1{}^k \tag{1.10}$$

where N is the total number of evaluations, and F_y are Fibonacci numbers defined by:

$$\left.\begin{array}{l} F_0 = F_1 = 1 \\ F_n = F_{n-1} + F_{n-2} \end{array}\right\} \quad n \geq 2 \tag{1.11}$$

The procedure then determines which interval to evaluate in the $(k + 1)^{st}$ step by evaluating the surface at the four points, β_1, β_2, β_3, and β_4 (Wilson, 1974, p. 322). The total number of functional evaluations, N, is determined from the desired interval of search after the N iterations

(Wilson, 1974, p. 322).

Other numerical methods in this class, although different in structure, basically operate according to the same principle, in which the response surface is incrementally searched for an optimum point. Further information on the techniques within the modelling context is available in the literature: Fibonacci sequences (Batty and Mackie, 1973; Wilson, 1974), Newton-Raphson (Batty and Mackie, 1972, 1973; Batty, Foot, *et al.*, 1973; Batty, *et al.*, 1974; Wilson, 1974), and quadratic search by conjugate directions (Batty and Mackie, 1972). Also, Mackie (1972) gives an excellent account of several calibration algorithms, including those discussed in this chapter.

The basic problem with this class of methods is that the search vectors may diverge from the global optimum to local optima on the response surface if poor initial parameter values are chosen. This may be corrected by damping the procedure, i.e., by transforming the slope of the response surface to a more regular shape (Batty and Mackie, 1973), or by choosing the initial parameter values close to the optimum point so the solution does not degenerate (Batty, Foot, *et al.*, 1973, pp. 359-362; Batty, *et al.*, 1974, p. 471). Hyman (1969, p. 110) suggests that since, in many cases, the value $\beta\bar{C}$, where $\bar{C}$ is the mean trip cost, lies between one and two, a reasonable starting value is given by the equation:

$$\beta = \frac{3}{2\bar{C}} \qquad (1.12)$$

INDIRECT EVALUATION

The method of indirect evaluation does not require explicit appraisal of the slope of the response surface (Batty and Mackie, 1973), but performs functional evaluations at the vertices of some geometric configuration generated in the parameter space (Mackie, 1972, p. 39). The only method of indirect evaluation to be applied to the calibration problem in spatial interaction modelling appears to be the Simplex method of sequential search (Mackie, 1972, pp. 53-56; Batty and Mackie, 1972, pp. 222-224; Batty and Mackie, 1973).

For the calibration of an n-parameter model, the simplex is generated by evaluating the objective function at n + 1 vertices in an n-parameter space. The vertex having the worst performance with respect to the optimization of the test statistic, i.e., maximize or minimize, is identified, and the simplex is reflected away from this vertex. If this operation improves its performance, the simplex is expanded; if not, it is contracted. An illustration of these basic operations is given in Mackie (1972, p. 54). The method iterates by reflecting across the response surface and adjusting its shape until the optimum is reached.

Calibration by the simplex method is more reliable than direct search techniques because it overcomes the problem of convergence to local optima on the response surface. The method, however, takes somewhat longer to compute.

Despite the advances made in solving the statistics, there are problems concerning the statistics themselves which cannot be overcome.

Correlation statistics, such as the coefficient of determination, are not as sensitive to changes in parameter values as simpler *performance measures*, such as mean trip length (Batty 1970c, pp. 108-109; Batty, 1971, p. 416). Furthermore, in the calibration of multi-parameter interaction models, no single goodness-of-fit statistic can determine the parameter values simultaneously (Wilson, 1974, p. 323). A unique set of optimum parameter values can only be derived if each parameter is related to a particular calibration statistic (Batty, Foot, *et al.*, 1973, p. 358). In other words, there must be as many calibration statistics as there are parameters (Batty and Mackie, 1973). Also, correlation statistics, in particular cases (Wilson, 1974, p. 343), may lead to *bogus calibration*, which occurs when the response surface is peaked towards the maximum at the axis of one of the parameters. Wilson (1974, p. 342) states that the bogus calibration problem can render certain correlation statistics virtually useless in parameter estimation.

Finally, one cannot, with any confidence, draw statistical inferences from these correlation statistics because their distributions are unknown. One is restricted to getting a feel for the goodness-of-fit of the model to the observed data, when interpreting the results. Although, this restriction can be somewhat overcome by choosing more robust correlation statistics, such as chi-squared, which place less stringent assumptions on the data, it is more meaningful to derive new calibration statistics based upon statistical assumptions which consider all the information available concerning the problem (Batty and Mackie, 1973). The statistic measuring the goodness-of-fit must take the sample data into account in order to derive the "best" parameter values.

STATISTICAL ESTIMATORS

Batty and Mackie (1973) suggest *maximum likelihood* techniques as a meaningful approach to deriving calibration statistics. Based on the work of Hyman (1969) and Evans (1971), for a trip distribution model of the form

$$\hat{p}_{ij} = a_i b_j \; f(\beta, \; c_{ij}) \tag{1.12}$$

where $\hat{p}_{ij}$ = the probability of a trip maker living in i and having his destination in zone j,

the maximum likelihood estimator derives a set of 2n + 1 conditions for the 2n + 1 unknowns. Given the sample data from a trip survey (Figure 3), the balancing factors, a_i and b_j, are chosen such that the proportion of trips generated from and distributed to each zone by the model equals the proportion of trips leaving and the proportion arriving at each zone as observed in the sample, i.e., in the row and column totals of the sample matrix. The parameter β in the impedence function is calibrated against the mean trip cost, and is at its optimum value when the mean trip cost predicted by the model equals the mean trip cost calculated from the sample.

		Destination Zone 1	2	. . .	n	Trip Origins
	1	t_{11}	t_{12}	. . .	t_{1n}	$\sum_j t_{1j} = O_1$
	2	t_{21}	t_{22}	. . .	t_{2n}	$\sum_j t_{2j} = O_2$
Origin Zone	.					
	.					
	.					
	n	t_{n1}	t_{n2}	. . .	t_{nn}	$\sum_j t_{nj} = O_n$
Trip Destinations		$\sum_i t_{i1} = D_1$	$\sum_i t_{i2} = D_2$	. . .	$\sum_i t_{in} = D_n$	$\sum_i \sum_j t_{ij} = \sum_i O_i = \sum_j D_j = T$

FIGURE 3: Sample Trip Matrix

Cesario (1975) proposes the alternative principle of *least-squares* to derive calibration statistics for spatial interaction models of the same form. The conditions derived state that the parameters of the model must be such that the sum of squared residuals, defined as the sum of the squares of the difference between the number of trips from zone i to j observed in the sample and the number of trips between i and j predicted by the model, equals zero. Consistency in the balancing factors is achieved on the sum of squares of row and column elements, not simply on the sums of these elements as in maximum likelihood (Cesario, 1975, p. 15).

The statistics generated by the maximum likelihood and least-squares estimators possess characteristics which give them several advantages over correlation statistics. The calibration statistics are simpler and more sensitive to changes in parameter values. The maximum likelihood estimator, under certain assumptions on the nature of travel cost (Hyman, 1969, pp. 108-109), derives mean trip cost as the statistic against which to calibrate β; Batty (1970c,pp. 108-109) shows the sensitivity of this statistic. The statistics themselves are functions of *interaction variables*, the t_{ij}. Batty (1971, p. 416) finds that statistics using these variables are far more sensitive to variations in parameter values than statistics which measure *distributions of activity*, such as population or employment.

Secondly, the statistical estimators derive as many calibration statistics as there are parameters, so that a unique set of "best" parameter values can be determined. The statistics are generated by optimizing the estimator with respect to each unknown parameter. Batty and Mackie (1972, p. 214) develop calibration statistics for a two parameter shopping model and suggest several numerical methods for their solution. Subsequent British work on multi-parameter models bases its calibration strategies on these statistics (Batty, Foot, *et al.*, 1973, p. 359; Batty, *et al.*, 1974, p. 466).

Thirdly, the statistics derived from the maximum likelihood and least-squares methods do not lead to bogus calibration problems, as do correlation statistics (Wilson, 1974, p. 343).

But perhaps the most significant advantage of the maximum likelihood and least-squares approaches is that one can look into the construction of the estimator to see exactly what assumptions are being made about the sample data. If the data reasonably satisfy the assumptions, then one should be able to make inferences about the model's goodness-of-fit to the sample observations.

SUMMARY

This chapter has considered four approaches to the calibration problem which have been proposed in the literature: *graphical curve fitting*, *tabulation*, *systematic search*, and *statistical estimators*.

In the evaluation of these approaches, the method of graphical curve fitting is rejected. It fails to consider the interdependencies between the model and parameters, by estimating the parameter values outside the model framework. Tabulation methods are rejected too. Although they use a statistical measure of the model's goodness-of-fit to the survey data, they are too slow and inaccurate to be useful in the calibration of interaction models.

Systematic search techniques, which use numerical methods to optimize the model's goodness-of-fit to the sample data, based on a given correlation statistic, are shown to be a better calibration method than tabulation approaches. However, correlation statistics have several properties which make them undesirable measures of the model's goodness-of-fit. The statistics are relatively insensitive to changes in parameter values. Some statistics tend to optimize to a bogus solution. The statistics also fail to yield unique parameter values for multi-parameter spatial interaction models. But most important, because the assumptions placed upon the sample data by correlation statistics are unknown, it is impossible to make statistical inferences on the parameter values and the model's goodness-of-fit.

Statistics derived from statistical estimators, such as maximum likelihood and least-squares estimators are preferred, since they do not possess the undesirable properties of correlation statistics outlined above. Furthermore, because these statistics are derived from theoretical principles of statistical estimation, one should be able to deduce the assumptions made by the statistics about the data, and thus be able to make inferences about the model's goodness-of-fit. Therefore, this approach to calibration is selected as the most appropriate for estimating "best" parameter values.

It now remains to compare the statistics derived from the maximum likelihood and least-squares estimators through an examination of the estimators themselves. This is of interest because each estimator derives statistics against which to calibrate the model parameters from different basic assumptions about the data. Because the statistics differ, so do the subsequent parameter estimates. Mackie (1972, p. 36) asserts that a particular set of statistical conditions is based upon specific decision functions "embedded" in the statistics. The decision function can be related to trip purpose through the interzonal probability density function assumed by the statistical estimator (Kirby, 1974, p. 101). Therefore, the different parameter estimates from alternative statistical hypotheses relate to the behavioral characteristics of the trip-maker, of which trip purpose is a major factor. Since we are attempting to find the "best" parameter estimates in the calibration process, a particular statistical estimator may be more appropriate in deriving calibration statistics, depending on the type of interaction being modelled.

The next two chapters will examine in detail the maximum likelihood and least-squares estimators. The appropriate calibration statistics will then be derived and the assumptions that the statistics make upon the data will be defined.

CHAPTER 2

THE MAXIMUM LIKELIHOOD METHOD OF PARAMETER ESTIMATION

INTRODUCTION

This chapter will examine the mathematical approaches which have been used to derive the maximum likelihood conditions for optimum parameter estimates. In doing this, the intention is to define the behavioral assumptions which each approach implies.

In the literature, two distinct methodologies are applied. Hyman (1969) defines the calibration problem to be one of *hypothesis evaluation*. Evans (1971) and Kirby (1974) define the problem to be one of *point estimation*. Hypothesis evaluation is based on the assumption that competing hypotheses can be evaluated in terms of the survey data and that inferences can be made about which hypothesis best represents the observed distribution. Point estimation, on the other hand, involves the estimation of unknowns of a given hypothesis from a single function of the sample data (Freeman, 1963, p. 229).

Hyman's framework is included in this chapter on maximum likelihood estimators for two reasons. First, it will be shown, through an outline of his approach, that hypothesis evaluation is not a suitable framework upon which to calibrate model parameters. Second, the simplifying assumptions which Hyman uses in his framework eventually reduce the problem to that of maximizing the likelihood function in Bayes' equations, which is equivalent to Evans' (1971) and Kirby's (1974) frameworks when they approach the problem as that of point estimation.

Following the assessment of Hyman's (1969) work, the paper will briefly outline the general principles of the maximum likelihood estimator. This will be followed by the application of the maximum likelihood estimator to parameter calibration problems in spatial interaction models, through the work of Evans (1971) and Kirby (1974). The paper will emphasize the point that Evans (1971) and Kirby (1974) derive the same key calibration statistics from different mathematical approaches, and an attempt will be made to reconcile the two approaches on the basis of their implicit behavioral assumptions. Finally, the chapter will establish the relationship between the derived statistics and the behavioral conditions in the survey and will summarize the findings of the previous sections.

HYPOTHESIS EVALUATION AS A METHOD OF PARAMETER CALIBRATION: AN APPRAISAL OF HYMAN'S APPROACH

Hyman (1969) attempts to calibrate a trip distribution model using the concept of evidence in Bayes' equation. By taking the log-odds form of Bayes' equation (Tribus, 1969, p. 83), he develops the evidence for hypothesis H_1 over H_2, with respective distributions $\{\hat{p}_{ij}\}$ and $\{\hat{p}_{ij}^*\}$, where $\hat{p}_{ij}$ represents the proportion of trips between zones i and j predicted by H_1.

$$\frac{1}{T}ev(H_1|DX) = \sum_i \sum_j p_{ij} \log \hat{p}_{ij} - \sum_i \sum_j p_{ij} \log \hat{p}_{ij}^* \qquad (2.1)$$

where p_{ij} = the proportion of observed trips between i and j

D = the sample data

X = the conditions for the survey

Hyman states that the choice of H_1 which maximizes (2.1) yields a distribution [meaning hypothesis] giving the best possible fit to the survey data. He then defines

$$E(H_1|DX) = \sum_i \sum_j p_{ij} \log \hat{p}_{ij} \qquad (2.2)$$

and states that the choice of parameters which maximizes this expression, subject to the constraint

$$\sum_i \sum_j \hat{p}_{ij} = 1$$

yields a distribution giving a best possible fit to the data.

Several points of criticism are in order. First, Hyman is confusing the issues of parameter estimation and hypothesis evaluation. These are two distinct topics (Mackie, 1972, p. 35). Secondly, because of the framework that he has set up, the approach becomes neither strictly a Bayesian nor a hypothesis testing approach. Hyman's assumption is that there is no prior evidence for one hypothesis over the other

(Hyman, 1969, p. 106). This reduces the problem from a Bayesian one to simply maximizing the likelihood of a hypothesis on the data.

Second, Hyman states that maximizing equation (2.1) yields the hypothesis which best fits the observed distribution and *then*, after dropping a term in the equation, states that maximizing equation (2.2) yields parameter values which give H_1 a best fit. Since there is no prior evidence to support one hypothesis over the other, how without calibrating both hypotheses, can there be evidence for H_1 over H_2? Further, since the likelihood for H_2 is assumed to be zero, Hyman must be assuming that this hypothesis cannot be calibrated, i.e., a uniform distribution. Since the concept of evidence applies to *any* competing distribution, and since such competitors that can be calibrated exist, i.e., the intervening opportunities model (Hutchinson, 1974, pp. 107-113), or the Charnes, Raike and Bettenger (1972) model, Hyman's simplifying assumption is unreasonable.

As a maximum likelihood method of parameter calibration, Hyman's (1969) approach is correct. As a Bayesian approach, as it is credited in the literature (Evans,1971, p. 23; Batty and Mackie, 1972, p. 210; Wilson, 1974, p. 318), it is not. A Bayesian approach requires the hypothesis to be calibrated *a priori* and then to be altered by the data. Since the prior hypothesis in Hyman's framework is discounted, the approach only considers the likelihood of the hypothesis based on the data, and thus is not strictly Bayesian. The reader should refer to Sheppard (1974, pp. 62-63) for a description of a Bayesian framework for parameter calibration.

Hyman's work emphasizes the fact that point estimation is the only valid approach to parameter calibration, since, in effect, he ends up taking this approach in equation (2.2). Other authors have approached the parameter calibration problem using statistical theory on point estimation. Evans (1971) and Kirby (1974) use the method of maximum likelihood as a parameter estimation technique. Following a brief outline of the mechanics of the method, their work will be reviewed.

THE MAXIMUM LIKELIHOOD ESTIMATOR

The method of maximum likelihood in point estimation can be described as follows. Consider a random variable t and a sample of T independent observations, t_{ij} from the same distribution, where t_{ij} represents the number of trip interchanges between zones i and j. The probability of observing t_{ij} is $\phi(t_{ij}|\beta)$, where the form of ϕ is known but the value of β is not. The joint probability of the observations, which is a function of the unknown parameter β, is called the likelihood function.

$$L(t_{ij}|\beta) = \prod_i \prod_j \phi(t_{ij}|\beta) \tag{2.3}$$

According to the maximum likelihood principle, we choose as our estimate of β that value which maximizes the joint probability of the actual observations. The conditions for a maximum are thus:

$$\frac{\partial L}{\partial \beta} = \frac{\partial}{\partial \beta}\left[\prod_i \prod_j \phi(t_{ij}|\beta)\right] = 0 \tag{2.4}$$

and

$$\frac{\partial^2 L}{\partial \beta^2} < 0 \tag{2.5}$$

Since $\log L(t_{ij}|\beta)$ and $L(t_{ij}|\beta)$ reach their maximum at the same values of β (Freeman, 1963, p. 254), then equations (2.4) and (2.5) can be written

$$\frac{\partial}{\partial \beta} \log L(t_{ij}|\beta) = \frac{\partial}{\partial \beta} \sum_i \sum_j \phi(t_{ij}|\beta) = 0 \tag{2.6}$$

and

$$\frac{\partial^2}{\partial \beta^2}\left[\sum_i \sum_j \phi(t_{ij}|\beta)\right] < 0 \tag{2.7}$$

to derive the optimum value of β.

The maximum likelihood estimator has several desirable properties which make it preferable to other point estimation approaches (Larson, 1969, p. 223), and the reader is referred to the literature (Freeman, 1963, pp. 257-262; Larson, 1969, pp. 233-250) for a description of these properties.

Having described the maximum likelihood approach, let us examine its application to parameter calibration in spatial interaction modelling.

THE DEVELOPMENT OF CALIBRATION STATISTICS FROM THE METHOD OF MAXIMUM LIKELIHOOD

This section will review the work of Evans (1971) and Kirby (1974). Basically, it will develop the calibration statistics by the likelihood estimator from the authors' different mathematical approaches, and will define the conditions required for deriving best parameter estimates.

EVANS APPROACH TO CALIBRATION

Evans (1971) derives optimum parameter values for spatial interaction models of the form:

$$\hat{p}_{ij} = a_i b_j \exp(-\beta c_{ij}) \tag{2.8}$$

where $\hat{p}_{ij}$ is defined as the *probability* of a trip originating in zone i and having zone j as its destination. He assumes we are given a sample of trip interchanges, $\{t_{ij}\}$, from which the proportion of trips between each i and j can be calculated:

$$p_{ij} = \frac{t_{ij}}{T} \tag{2.9}$$

where T is the number of observations in the sample. If we interpret the trip proportions to represent probabilities (Lindley, 1965, p. 3; Freund, 1952, p. 112), and their joint distribution to be multinomial (Evans, 1971, p. 23), the likelihood for β on the sample will be (Edwards, 1972, p. 19)

$$P(\hat{p}_{ij} = p_{ij},\ \forall_{i,j}) = \frac{T!}{\prod_i \prod_j t_{ij}!} \prod_i \prod_j (\hat{p}_{ij})^{t_{ij}} \tag{2.10}$$

and

$$\sum_i \sum_j \hat{p}_{ij} = 1 \tag{2.11}$$

or, on converting (2.10) to log-form:

$$\log P(\hat{p}_{ij} = p_{ij},\ \forall_{i,j}) = \log T! - \sum_i \sum_j \log t_{ij}! + \sum_i \sum_j t_{ij} \log \hat{p}_{ij} \tag{2.12}$$

Evans maximizes the likelihood of the sample on the joint distribution of $\hat{p}_{ij}$'s.

$$\max \log P(\hat{p}_{ij} = p_{ij},\ \forall_{i,j}) = \log T! - \sum_i \sum_j \log t_{ij}! + \sum_i \sum_j t_{ij} \log \hat{p}_{ij}$$

$$\text{S.T.} \quad \sum_i \sum_j \hat{p}_{ij} = \sum_i \sum_j a_i b_j \exp(-\beta\, c_{ij}) = 1 \tag{2.13}$$

Forming the Lagrangian,

$$\Lambda = \log T! - \sum_i \sum_j \log t_{ij}! + \sum_i \sum_j t_{ij} [\log a_i + \log b_j - \beta c_{ij}]$$

$$+ \lambda [1 - \sum_i \sum_j a_i b_j \exp(-\beta c_{ij})] \qquad (2.14)$$

and differentiating with respect to the unknowns a_i, b_j and β yields the first order conditions for $\log P(\hat{p}_{ij} = p_{ij}, \forall_{i,j})$ to be a maximum.

$$\frac{\partial \Lambda}{\partial a_i} = \sum_j \frac{t_{ij}}{a_i} - \lambda \sum_j b_j \exp(-\beta c_{ij}) = 0 \qquad \text{for all } i \qquad (2.15)$$

$$\frac{\partial \Lambda}{\partial b_j} = \sum_i \frac{t_{ij}}{b_j} - \lambda \sum_i a_i \exp(-\beta c_{ij}) = 0 \qquad \text{for all } j \qquad (2.16)$$

$$\frac{\partial \Lambda}{\partial \beta} = - \sum_i \sum_j t_{ij} c_{ij} + \lambda \sum_i \sum_j c_{ij} a_i b_j \exp(-\beta c_{ij}) \qquad (2.17)$$

Equation (2.15) is rewritten as

$$\sum_j t_{ij} = \lambda \sum_j a_i b_j \exp(-\beta c_{ij}) \qquad (2.18)$$

and by summing (2.18) over i:

$$\sum_i \sum_j t_{ij} = \lambda \sum_i \sum_j a_i b_j \exp(-\beta c_{ij}) \qquad (2.19)$$

Combining (2.13) and (2.19) gives

$$\lambda = \sum_i \sum_j t_{ij} = T \tag{2.20}$$

Therefore, substituting for λ into (2.15), (2.16) and (2.17), and re-arranging terms results in the first order conditions on (2.12).

$$\sum_j a_i\, b_j \exp\,(-\beta\, c_{ij}) = \frac{1}{T} \sum_j t_{ij} = \sum_j p_{ij} \qquad \text{for all } i \tag{2.21}$$

$$\sum_i a_i\, b_j \exp\,(-\beta\, c_{ij}) = \frac{1}{T} \sum_i t_{ij} = \sum_i p_{ij} \qquad \text{for all } j \tag{2.22}$$

$$\sum_i \sum_j c_{ij}\, a_i\, b_j \exp\,(-\beta\, c_{ij}) = \frac{1}{T} \sum_i \sum_j c_{ij}\, t_{ij} = \sum_i \sum_j c_{ij}\, p_{ij} \tag{2.23}$$

Sufficient second order conditions for the maximum are:

$$\frac{\partial^2 \Lambda}{\partial a_i} < 0,\ \forall_j;\ \frac{\partial^2 \Lambda}{\partial b_j} < 0,\ \forall_i;\ \frac{\partial^2 \Lambda}{\partial \beta} < 0 \tag{2.24}$$

From (2.15), (2.16) and (2.17),

$$\frac{\partial^2 \Lambda}{\partial a_i} = -\frac{1}{a_i^{\,2}} \sum_j t_{ij} < 0 \qquad a_i \in R \tag{2.25}$$

$$\frac{\partial^2 \Lambda}{\partial b_j} = -\frac{1}{b_j^{\,2}} \sum_i t_{ij} < 0 \qquad b_j \in R \tag{2.26}$$

$$\frac{\partial^2 \Lambda}{\partial \beta} = - \lambda \sum_i \sum_j c^2_{ij} \; a_i \; b_j \; \exp\,(-\beta \; c_{ij}) < 0 \qquad (2.27)$$

$$\text{for } a_i,\; b_j > 0$$

Since by definition $a_i > 0$, $b_j > 0$, the second order conditions hold. Therefore, the best parameter values are attained when:

(1) the proportion of trips generated in zone i by the model agrees with the proportion observed in the sample (2.21),

(2) the proportion of trips attracted to each zone j by the model agrees with the proportion observed in the sample (2.22), and

(3) the average generalized cost of travel is the same in both the model and the survey (2.23).

Because the maximum likelihood estimator generates as many equations as unknowns, then theoretically, the system should be solvable for a unique set of values a_i, b_j and β. However, because of the large number of terms usually involved in the system of equations, Evans (1971, p. 30) has proposed an iterative procedure which converges to the optimum solution. Mackie (1972) has shown how the optimization methods discussed in Chapter 1 can be applied to statistics derived by the maximum likelihood estimator, to solve for the parameter values.

KIRBY'S APPROACH TO CALIBRATION

Kirby, on the other hand, assumes that the sample matrix of t_{ij}'s from a traffic survey is only one estimate of the number of journeys from each i to j. He hypothesizes that if the results of several independent surveys were available, a mean number of journeys on each interchange could be established. However, in most circumstances, this additional information is not available.

The model to be calibrated estimates the *mean number* of journeys from i to j,

$$\hat{t}_{ij} = a_i \, b_j \, f(\beta, c_{ij}) \tag{2.28}$$

and the observation, t_{ij}, is regarded as belonging to a probability density function with mean $\hat{t}_{ij}$. The probability $\phi(t_{ij})$ of obtaining an observation t_{ij} is assumed to depend only upon the values t_{ij}, the mean $\hat{t}_{ij}$, and certain properties independent of both i and j (Kirby, 1974, p. 99), such as the sampling process.

Therefore,

$$\phi(t_{ij}) = \phi(t_{ij} | \hat{t}_{ij}) \tag{2.29}$$

Kirby then finds the compound probability of obtaining the sample matrix of trips, $\{t_{ij}\}$, which is, as defined above, the likelihood function.

$$L = \prod_i \prod_j \phi(t_{ij} | \hat{t}_{ij}) \tag{2.30}$$

Since the mean value is unknown from the observation, but is generated in the model

$$L = \prod_{i}\prod_{j} \phi(t_{ij} | a_i\ b_j\ f(\beta,\ c_{ij})) \qquad (2.31)$$

or

$$\log L = \sum_{i}\sum_{j} \log \phi(t_{ij} | a_i\ b_j\ f(\beta,\ c_{ij})) \qquad (2.32)$$

Maximizing the log-likelihood maximizes the compound probability of obtaining the base year matrix of journeys (Kirby, 1974, p. 99). Solving the first order conditions with respect to the unknowns yields parameter values which maximize (2.32). These are

$$\frac{\partial \log L}{\partial a_i} = \frac{\partial \log \phi}{\partial \hat{t}_{ij}} \cdot \frac{\partial \hat{t}_{ij}}{\partial a_i}$$

$$= \sum_{j} \frac{\partial \log \phi}{\partial \hat{t}_{ij}}\, b_j\ f(\beta,\ c_{ij})$$

$$= \frac{1}{a_i} \sum_{j} \frac{\partial \log \phi}{\partial \hat{t}_{ij}} \cdot \hat{t}_{ij} = 0 \qquad \text{for all } i, \qquad (2.33)$$

$$\frac{\partial \log L}{\partial b_j} = \frac{1}{b_j} \sum_{i} \frac{\partial \log \phi}{\partial \hat{t}_{ij}} \cdot \hat{t}_{ij} = 0 \qquad \text{for all } j, \qquad (2.34)$$

$$\frac{\partial \log L}{\partial \beta} = \frac{\partial \log \phi}{\partial \hat{t}_{ij}} \cdot \frac{\partial \hat{t}_{ij}}{\partial \beta}$$

$$= \sum_i \sum_j \frac{\partial \log \phi}{\partial \hat{t}_{ij}} a_i \, b_j \frac{\partial f}{\partial \beta}$$

$$= \sum_i \sum_j \frac{\partial \log \phi}{\partial \hat{t}_{ij}} \hat{t}_{ij} \frac{1}{f} \cdot \frac{\partial f}{\partial \beta}$$

$$= \sum_i \sum_j \frac{\partial \log \phi}{\partial \hat{t}_{ij}} \hat{t}_{ij} \frac{\partial \log f}{\partial f} \cdot \frac{\partial f}{\partial \beta}$$

$$= \sum_i \sum_j \hat{t}_{ij} \frac{\partial \log \phi}{\partial \hat{t}_{ij}} \left(\frac{\partial \log f}{\partial \beta} \right) = 0 \qquad (2.35)$$

If the a_i and b_j are defined to be *strictly* positive, i.e., each zone attracts and generates interzonal trips, the conditions for optimum parameter estimation become:

$$\sum_j \frac{\partial \log \phi}{\partial \hat{t}_{ij}} \cdot \hat{t}_{ij} = 0 \qquad \text{all } i \qquad (2.36)$$

$$\sum_i \frac{\partial \log \phi}{\partial \hat{t}_{ij}} \cdot \hat{t}_{ij} = 0 \qquad \text{all } j \qquad (2.37)$$

$$\sum_i \sum_j \hat{t}_{ij} \frac{\partial \log \phi}{\partial \hat{t}_{ij}} \left(\frac{\partial \log f}{\partial \beta} \right) = 0 \qquad (2.38)$$

Assume the probability density function of the variable t_{ij} to be Poisson. Then,

$$\phi(t_{ij}|\hat{t}_{ij}) = \frac{(\hat{t}_{ij})^{t_{ij}}}{t_{ij}!} \exp(-\hat{t}_{ij}) \tag{2.39}$$

By substituting (2.39) into (2.36), (2.37) and (2.38), first order conditions for the maximum likelihood estimator are derived.

$$\sum_j \frac{\partial}{\partial \hat{t}_{ij}} [(-\hat{t}_{ij}) + t_{ij} \log \hat{t}_{ij} - \log t_{ij}!] \, \hat{t}_{ij} = 0$$

$$\sum_j \left(-1 + \frac{t_{ij}}{\hat{t}_{ij}}\right) \hat{t}_{ij} = 0$$

$$\sum_j (-\hat{t}_{ij} + t_{ij}) = 0 \qquad \text{for all } i \tag{2.40}$$

Similarly,

$$\sum_i (-\hat{t}_{ij} + t_{ij}) = 0 \qquad \text{for all } j \tag{2.41}$$

and

$$\sum_i \sum_j (-\hat{t}_{ij} + t_{ij}) \frac{\partial \log f}{\partial \beta} = 0 \tag{2.42}$$

Providing an appropriate transformation is made on the cost function, log $f(\beta, c_{ij})$, the statistics derived from Kirby's approach are identical to the statistics derived by Evans (1971), although the two are developed on different mathematical frameworks.

ASSUMPTIONS OF THE TWO DERIVATIONS

In order to understand why the two approaches yield the same conditions on the parameters, it is necessary to look into the assumptions, both implicit and explicit, made by both Evans (1971) and Kirby (1974) about the sample data.

Evans (1971) examines the *macro-state* of the distribution. He is interested in finding the *joint probability* of observing a given matrix of trips, and he defines this to be a multinomial density function, assuming that the sample proportions, p_{ij}, differ from the mean $\hat{p}_{ij}$ by reason of chance arising in the sampling of trips (Evans, 1971, p. 23). Thus, for the multinomial, the distribution variable is defined as a probability. Maximization of the density function with respect to the parameters, after applying the required constraint on the variable ($\sum_i \sum_j p_{ij} = 1$), yields conditions for deriving optimum parameter estimates. Evans, therefore, is explicitly assuming the variation of trips between each i-j interchange to be a product of chance in the sampling process.

Kirby (1974), on the other hand, examines the *micro-states* of the distribution. The maximum likelihood estimator requires certain assumptions, whether explicit or implicit, made upon the nature of the distribution of trips between each i-j pair. Kirby (1974, p. 99) calls this the *sampling distribution*. However, it is more appropriate to define the nature of the distribution to be a *probability density function* of the variable t_{ij}, to avoid confusing the terminology with the trip distribution, derived by the model,or with sampling theory.

Therefore, the probability density function must be known in order to use the likelihood function to estimate the parameters. This density function describes the probability that the value of t_{ij} will be observed between a given i-j pair. The likelihood is found by taking the compound probability of the density functions describing the micro-states. The parameters, thus derived, generate a macro-state distribution.

If the Poisson density function describes the variation of t_{ij} in the micro-states, the maximum likelihood estimator derives identical conditions for optimum parameter values to Evans' (1971) method of maximizing the likelihood of the multinomial density function. This may be explained by reexamining the assumptions which Evans (1971) makes about the variation of trips between each i-j interchange.

By assuming the joint probability of the observation to be multinomial, Evans (1971), contrary to his assumption of random error in sampling, is *implicitly* assuming the probability density function for a simple interchange to be *binomial*. Thus, the trip proportions are from a probability density function with mean

$$\mu = T\,\hat{p}_{ij} \tag{2.43}$$

$$= \hat{t}_{ij} \tag{2.44}$$

and variance

$$\sigma^2 = T\,\hat{p}_{ij}\,(1 - \hat{p}_{ij}) \tag{2.45}$$

$$= \hat{t}_{ij}\,(1 - \hat{p}_{ij}) \tag{2.46}$$

Kirby (1974) in assuming the variation of trips between each i-j interchange to be Poisson, is implying a probability density function with mean

$$\mu = \hat{t}_{ij} \tag{2.47}$$

and variance

$$\sigma^2 = \hat{t}_{ij} \tag{2.48}$$

Identical conditions for optimum parameter values have been deduced from both approaches, even though there is a fundamental contradiction in the variance of travel on each i-j pair. However, the difference between the two probability density functions describing the micro-states may not be significant, because the variance of the binomial density function, implied in the multinomial, approaches the mean $\hat{t}_{ij}$ as $\hat{p}_{ij}$ approaches zero. This means that if the number of interchanges in the system is large and the proportion of trips (which equals $\frac{\hat{t}_{ij}}{\sum_i \sum_j \hat{t}_{ij}}$)

on each interchange is small, the density functions are approximately the same. Freeman (1963, p. 105) suggests that practical working values for N and $\hat{p}_{ij}$ are $N > 50$ and $\hat{p}_{ij} \leq 0.10$. A system of n zones generates $N = n^2$ interchanges. Therefore, for a large number of zones, the $\hat{p}_{ij}$ may be small enough for the assumptions made by the two approaches to be equivalent.

The assumption of a Poisson or binomial density function describing the variation of t_{ij} on each interchange gives a statement about the variation of travel between the two zones. Kirby (1974, p. 99) relates the density function to trip purpose, and asserts that one expects to observe a greater variance for certain trip purposes, such as shopping or recreational travel, than more regular travel patterns, such as the journey-to-work.

Statistical conditions derived by the maximum likelihood estimator, with the implicit assumption of a Poisson distribution describing the variation of travel, are usually employed in model calibration of work trips (Batty and Mackie, 1973; Batty, Foot, *et al.*, 1973, p. 359). Although the Poisson is known to describe the variation of traffic on a road reasonably well (Kirby, 1974, p. 103), and interzonal travel might vary in a similar manner, it is possible that the Poisson has too great a variance to accurately portray the variation of journey-to-work trips.

Kirby (1974, p. 101) suggests that other statistics, based on different density functions, which can also be derived by the maximum likelihood estimator, may be more appropriate for modelling journey-to-work travel. One needs to examine the higher moments of t_{ij}, disaggregated by trip purpose in order to derive the appropriate statistics against which to calibrate the model.

GENERAL CONDITIONS REQUIRED BY THE MAXIMUM LIKELIHOOD ESTIMATOR

The *conditions* derived by the maximum likelihood estimator define *statistics* which are calibrated against the model to derive "best" parameter estimates. Because the statistics are derived from statistical estimation theory, they yield the most pertinent information from the sample data to give the model a best fit to the survey. For the single parameter spatial interaction model, the maximum likelihood estimator yields 2n + 1 conditions for the 2n + 1 unknowns. The first 2n conditions (equations (2.36) and (2.37)) require the *balancing factors*, a_i and b_j, to be such that some function of the trip-origins and trip-ends generated by the model agree with the same function of origins and destinations in the sample. The actual form of the function is dependent upon the probability density function which describes the variation of the variable, $\hat{t}_{ij}$. The statistic derived to calibrate the parameter, β, (equation (2.38)) is dependent on the density function of $\hat{t}_{ij}$, and some function of travel cost between zones i and j.

The cost function describes the generalized cost of travel between each i-j pair and is some combination of distance, time and direct money charges to the trip-maker (Evans, 1973, p. 40). It is assumed to have the following properties:

$$\frac{\partial f}{\partial c_{ij}} < 0 \tag{2.49}$$

and

$$\frac{\partial^2 f}{\partial c_{ij}^2} > 0 \tag{2.50}$$

As the cost of travel between two zones increases, the cost function decreases the number of journeys between the zones, *ceteris paribus*, and the function decreases such that the amount of travel between two zones decreases at a decreasing rate.

For a given probability density function, several statistics related to generalized cost can be derived. If the generalized cost function is defined as:

$$f(\beta,\ c_{ij}) = \exp\ (-\beta\ h(c_{ij})) \tag{2.51}$$

where h is some transformation on the generalized cost of travel, then, using Kirby's general conditions and assuming the density function of $\hat{t}_{ij}$ to be Poisson, the calibration statistic is derived as follows. From equation (2.42)

$$\sum_i \sum_j \hat{t}_{ij} \frac{\partial \log f(\beta,\ c_{ij})}{\partial \beta} = \sum_i \sum_j t_{ij} \frac{\partial \log \cdot f(\beta,\ c_{ij})}{\partial \beta} \tag{2.52}$$

$$\sum_i \sum_j \hat{t}_{ij}\ h(c_{ij}) = \sum_i \sum_j t_{ij}\ h(c_{ij}) \tag{2.53}$$

Hyman (1969, p. 109) derives calibration statistics for several cost functions and argues that each function may be appropriate for different trip purposes. He suggests the exponential model

$$f(\beta, c_{ij}) = \exp(-\beta c_{ij}) \tag{2.54}$$

i.e.,

$$h(c_{ij}) = c_{ij} \tag{2.55}$$

as an appropriate measure of cost in the journey-to-work. This transformation is often used for this purpose in modelling applications (Wilson, *et al.*, (1969), p. 339), and it is this cost function which makes Kirby's calibration statistics equivalent to Evans' (1971). The calibration statistic derived under the assumption of a Poisson density function describing the variation of t_{ij} over each interchange yields, from (2.53) and (2.55)

$$\sum_i \sum_j \hat{t}_{ij} c_{ij} = \sum_i \sum_j t_{ij} c_{ij} \tag{2.56}$$

since

$$\sum_i \sum_j \hat{t}_{ij} = \sum_i \sum_j t_{ij} = T \tag{2.57}$$

$$\frac{\sum_i \sum_j \hat{t}_{ij} c_{ij}}{T} = \frac{\sum_i \sum_j t_{ij} c_{ij}}{T} \tag{2.58}$$

Therefore,

$$\sum_i \sum_j \hat{p}_{ij} c_{ij} = \bar{C} \tag{2.59}$$

The parameter β is calibrated against the *mean trip cost*, and is at its optimum value when the mean trip cost predicted by the model equals the mean trip cost observed in the sample.

In general, if the probability density function is assumed to be Poisson, the likelihood estimator calibrates the parameter β against the mean value of the transformation on cost. Optimum conditions state that the mean value of the cost transformation generated in the model must equal the mean value of the transformation observed in the sample. It will be shown in the next chapter, that a different density function, the normal, yields other statistics.

To summarize, the *balancing factors* of the spatial interaction models are determined by statistics describing the variation of the t_{ij}. The model *parameter*, β, is calibrated to a statistic which is a function of the probability density function of the t_{ij}, and the deterence function, which is a measure of the cost of travel between two zones. The work of Hyman (1969) and Kirby (1974) relates the deterence function and density function to trip purpose. Since different combinations of these functions yield different calibration statistics, one particular combination may be the more appropriate against which to calibrate the parameter, depending upon the trip purpose being modelled.

SUMMARY

This chapter has discussed the maximum likelihood approach to parameter estimation. It has argued that parameter calibration is a problem of point estimation, not hypothesis evaluation, and has reviewed the different mathematical approaches of Evans (1971) and Kirby (1974). It has been found that the same key conditions are derived by the two mathematical approaches.

The chapter has looked at the implicit and explicit assumptions that each approach makes on the sample data, and it has been found that in the context of spatial interaction modelling, the assumptions are approximately the same. The chapter concludes by examining the various calibration statistics which can be generated by altering the assumptions made on the variable t_{ij} and the trip cost, and notes that a particular statistic may be better suited for the calibration of a specific trip purpose.

CHAPTER 3

THE LEAST SQUARES METHOD OF PARAMETER ESTIMATION

INTRODUCTION

This chapter will examine another approach to point estimation which has been applied in the context of spatial interaction models. This is the method of least-squares. As in the previous chapter on maximum likelihood, it will briefly outline the general principles of the least-squares estimator. This will be followed by an application of least-squares to the parameter calibration problem through an outline of the work of Cesario (1975). The advantage of this estimator will then be shown by developing the same calibration statistics as least-squares by the maximum likelihood estimator. It will be shown that maximum likelihood makes very stringent assumptions about the data to derive the same calibration statistics as least-squares, and that any less strict assumptions about the data result in different calibration statistics. The chapter will conclude by interpreting the least-squares statistics in the context of spatial interaction modelling and will suggest situations when the least-squares estimator is appropriate in model calibration.

THE LEAST-SQUARES ESTIMATOR

Consider n random variables with known and possibly different means and known and possibly different variances. The variable t_{ij} can be thought of as the outcome of a random sample of size 1 from a population with mean

$$\hat{t}_{ij} = f(\mu_k) \tag{3.1}$$

and variance, σ^2, where μ_k represent the parameters of the model. The observation can thus be represented by

$$t_{ij} = \hat{t}_{ij} + e_{ij} \tag{3.2}$$

where e_{ij} is an error term.

The least-squares principle states that the best linear unbiased estimator of the parameters, μ_k, is the one which minimizes the deviations of the sample variance, defined by

$$e^2_{ij} = (t_{ij} - \hat{t}_{ij})^2 \tag{3.3}$$

The estimates for the parameters are chosen so that when substituted into (3.3), i.e.,

$$S = \sum_i \sum_j e^2_{ij} = \sum_i \sum_j (t_{ij} - f(\mu_k))^2 \tag{3.4}$$

they produce the least possible value for S.

The first order condition for (3.4) being a minimum is:

$$ds = 0 \tag{3.5}$$

Sufficient second order conditions require:

$$d^2S \text{ positive definite} \tag{3.6}$$

Therefore, the optimum values of the parameters, μ_k, are derived when

$$\frac{\partial}{\partial \mu_1} \sum_i \sum_j (t_{ij} - f(\mu_k))^2 = \frac{\partial}{\partial \mu_r} \sum_i \sum_j (t_{ij} - f(\mu_k))^2 = 0 \tag{3.7}$$

and the principal minors, $| H_1 |, \ldots, | H_r |$, of the bordered Hessian,

$$| H | = \begin{vmatrix} \frac{\partial^2 S}{\partial \mu_1^2} & \cdots & \frac{\partial^2 S}{\partial \mu_1 \partial \mu_r} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \frac{\partial^2 S}{\partial \mu_r \partial \mu_1} & \cdots & \frac{\partial^2 S}{\partial \mu_r^2} \end{vmatrix} \tag{3.8}$$

are greater than zero.

The least-squares estimator has several properties which make it the best *linear* unbiased estimator of the parameters, μ_k, and the reader is referred to Freeman (1963, p. 265) and Wonnacott and Wonnacott (1970, pp. 21-30) for a discussion of these characteristics. Specifically, it is called the "best" linear unbiased estimator because the estimates of the μ_k have *minimum variance* (see Freeman, 1965, p. 265) for a note on the correspondence between "best" and minimum variance).

Let us now examine the application of the least-squares approach in the spatial interaction context.

THE DEVELOPMENT OF CALIBRATION STATISTICS FROM THE METHOD OF LEAST-SQUARES: CESARIO'S APPROACH

Cesario (1975) approaches the problem by considering a set of observations $\{t_{ij}\}$ and an estimate for the mean values of t_{ij},

$$\hat{t}_{ij} = a_i b_j f(\beta, c_{ij}) \tag{3.9}$$

where a_i, b_j are balancing factors and β is the parameter of the model which must be estimated.

The least-squares principle requires the minimization of the sum of squared residuals. Therefore, minimize

$$S = \sum_i \sum_j (t_{ij} - \hat{t}_{ij})^2 \tag{3.10}$$

$$= \sum_i \sum_j (t_{ij} - a_i b_j f(\beta, c_{ij}))^2 \tag{3.11}$$

First order conditions for the minimization of S require:

$$\frac{\partial S}{\partial a_i} = \frac{\partial}{\partial a_i}\left[\sum_i \sum_j (t_{ij} - a_i b_j f(\beta, c_{ij}))^2\right] = 0 \tag{3.12}$$

$$2 \sum_j (t_{ij} - a_i b_j f(\beta, c_{ij})) (-b_j f(\beta, c_{ij})) = 0$$

Upon rearranging terms

$$a_i \sum_j b_j^2 f(\beta, c_{ij})^2 - \sum_j t_{ij} b_j f(\beta, c_{ij}) = 0 \tag{3.13}$$

or

$$\sum_j \hat{t}_{ij}^2 - \sum_j \hat{t}_{ij} t_{ij} = 0 \qquad \text{for all } i \tag{3.14}$$

Similarly

$$\frac{\partial S}{\partial b_j} = \frac{\partial}{\partial b_j}\left[\sum_i \sum_j (t_{ij} - a_i b_j f(\beta, c_{ij}))^2\right] = 0 \tag{3.15}$$

$$b_j \sum_i a_i^2 f(\beta, c_{ij})^2 - \sum_i t_{ij} a_i f(\beta, c_{ij}) = 0 \tag{3.16}$$

or

$$\sum_i \hat{t}_{ij}^2 - \sum_i \hat{t}_{ij} t_{ij} = 0 \qquad \text{for all } j \tag{3.17}$$

and

$$\frac{\partial S}{\partial \beta} = \frac{\partial}{\partial \beta}\left[\sum_i \sum_j (t_{ij} - a_i b_j f(\beta, c_{ij}))^2\right] = 0 \tag{3.18}$$

$$2 \sum_i \sum_j (t_{ij} - a_i\, b_j\, f(\beta, c_{ij}))\ (-a_i\, b_j \frac{\partial f}{\partial \beta}) = 0$$

$$\sum_i \sum_j a_i^2\, b_j^2\, f(\beta, c_{ij}) \frac{\partial f}{\partial \beta} - \sum_i \sum_j t_{ij}\, a_i\, b_j \frac{\partial f}{\partial \beta} = 0 \qquad (3.19)$$

or

$$\sum_i \sum_j \hat{t}_{ij}^2 \frac{\partial \log f}{\partial \beta} - \sum_i \sum_j \hat{t}_{ij}\, t_{ij} \frac{\partial \log f}{\partial \beta} = 0 \qquad (3.20)$$

Sufficient conditions for S to be a minimum exist if the second-order conditions (3.8) are satisfied.

$$\frac{\partial^2 S}{\partial a_i^2} = \sum_j b_j^2\, f(\beta, c_{ij})^2 > 0 \qquad (3.21)$$

$$\frac{\partial^2 S}{\partial b_j^2} = \sum_i a_i^2\, f(\beta, c_{ij})^2 > 0 \qquad (3.22)$$

$$\begin{aligned}
\frac{\partial^2 S}{\partial \beta^2} &= \frac{\partial^2}{\partial \beta^2} \sum_i \sum_j (t_{ij} - \hat{t}_{ij})^2 \\
&= \sum_i \sum_j \left[a_i\, b_j \left(\frac{\partial^2 f}{\partial \beta^2} + f \frac{\partial^2 f}{\partial \beta^2} \right) \right] - \sum_i \sum_j t_{ij}\, a_i\, b_j \frac{\partial^2 f}{\partial \beta^2} \\
&= \sum_i \sum_j a_i\, b_j \frac{\partial^2 f}{\partial \beta^2} \left(a_i\, b_j + a_i\, b_j\, f \right) - \sum_i \sum_j a_i\, b_j \frac{\partial^2 f}{\partial \beta^2} t_{ij} \\
&= \sum_i \sum_j a_i\, b_j \frac{\partial^2 f}{\partial \beta^2} \left(a_i\, b_j + \hat{t}_{ij} \right) - \sum_i \sum_j a_i\, b_j \frac{\partial^2 f}{\partial \beta} (t_{ij}) > 0 \qquad (3.23)
\end{aligned}$$

for $a_i, b_j > 0$ all i, j.

The first order conditions yield 2n + 1 equations for the 2n + 1 unknowns so a unique solution exists. However, because of the non-linear character of the 2n + 1 normal equations, Cesario (1975, p. 14) devises an iterative procedure which converges to the optimum solution.

Also, the first order conditions differ from the conditions derived by the maximum likelihood estimator in the previous chapter. Instead of requiring correspondence of trip-end and trip-origin totals, consistency is achieved on the sum of squares of the row and column elements (Cesario, 1975, p. 15). The parameter β is calibrated against a more complex statistic (equation (3.20)), which is a function of the squared trip matrix elements, and of the generalized cost function.

A purpose of this chapter, as previously stated, is to examine the assumptions made by the least-squares estimator on the data $\{t_{ij}\}$, and to compare these assumptions with those of the maximum likelihood estimator. This discussion will lead to an assessment of the behavioral hypotheses implied in maximum likelihood assumptions.

ASSUMPTIONS OF THE LEAST-SQUARES APPROACH AND A COMPARISON TO MAXIMUM LIKELIHOOD

The calibration statistics derived from the method of least-squares (equations (3.14), (3.17), and (3.20)) yield unbiased parameter values with minimum variance. The significance of this property is that it can be proved by the Gauss-Markov theorem on a very weak set of assumptions (Wonnacott and Wonnacott, 1970, pp. 48-51). Specifically, the least-squares estimator requires no assumption about the shape of the density function of the error term (Wonnacott and Wonnacott, 1970, p. 21). This means the estimator requires no information about the density function of the variable t_{ij}.

The importance of this assumption can be shown by developing identical calibration statistics as those of least-squares, by the method of maximum likelihood. Since the maximum likelihood estimator requires the probability density function of t_{ij} to be known, then the assumption on the denstiy function by maximum likelihood can be examined to compare the two methods of point estimation.

Let us consider the distribution of the *observation*, t_{ij}. If the probability density function of this variable is assumed to be normal, with mean $\hat{t}_{ij}$, the probability of obtaining the observation t_{ij} is

$$\phi(t_{ij}|\hat{t}_{ij}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-(t_{ij} - t_{ij})^2/2\sigma^2\right] \qquad (3.24)$$

Using the general maximum likelihood conditions developed by Kirby (1974, p. 100), and upon substituting the probability density function defined above, the following first order conditions for optimum parameter values result.

$$\sum_j \frac{\partial \log \phi}{\partial \hat{t}_{ij}} \hat{t}_{ij} = \sum_j (t_{ij} - \hat{t}_{ij})\, \hat{t}_{ij} = 0 \qquad \text{for all } i \qquad (3.25)$$

$$\sum_i (t_{ij} - \hat{t}_{ij})\, \hat{t}_{ij} = 0 \qquad \text{for all } j \qquad (3.26)$$

$$\sum_i \sum_j (t_{ij} - \hat{t}_{ij})\, \hat{t}_{ij} \frac{\partial \log f}{\partial \beta} = 0 \qquad (3.27)$$

The conditions (3.25), (3.26) and (3.27) are identical to the least-squares conditions (equations (3.14), (3.17) and (3.20)) against which to calibrate parameter values. However, in order to derive these identical conditions, where the least-squares estimator makes no assumptions about the data, the maximum likelihood estimator must assume the t_{ij} to be normally distributed with *common, constant variance*.

The assumption of a normal probability density function for t_{ij} is not at issue here. Several properties of the normal make it an appealing density function to assume in the context of spatial interaction modelling. First, the normal is a reasonable description of the behavior of many observable phenomena, and its application is generally a valid description of observable data (Freeman, 1963, p. 141). Second, the normal probability density function is the limiting form of many other density functions, including the Poisson. The normal approximates other probability density functions when the mean is large (Wetherill, 1967, p. 71). Since the $\hat{t}_{ij}$ on many interchanges are likely to be large values, i.e., $\hat{t}_{ij} \geq 30$, the normal will reasonably approximate the data (Freund, 1952, p. 233). There are some inconsistencies if the t_{ij} are assumed normally distributed, such as the assumption of the variable being continuous when in fact it is discrete, and the allowance of negative values of the variable t_{ij} when the mean is small.

However, the maximum likelihood estimator places a severe restriction on the *variance* of the t_{ij} when the density function is assumed normal. Conditions identical to least-squares can only be derived by maximum likelihood if the variance of each t_{ij} is constant and equal over all interchanges. Since the means ($\hat{t}_{ij}$) can differ by several hundred trips, this assumption about the variance is not reasonable. If we attempt to relax this restriction on the variance by assuming it to be proportional to the mean, i.e.,

$$\sigma^2_{ij} = \alpha^2 \hat{t}_{ij}^{\,2} \tag{3.28}$$

where α is the coefficient of variation, a different set of conditions is derived. Since the probability density function of the variable t_{ij} is now:

$$\phi(t_{ij}|\hat{t}_{ij}) = \frac{1}{\alpha\hat{t}_{ij}\sqrt{2\pi}} \exp\left[-(t_{ij} - \hat{t}_{ij})^2/2\alpha^2\, \hat{t}^2_{ij}\right] \tag{3.29}$$

then

$$\log \phi = -\frac{1}{2\alpha^2}\left[\frac{(t_{ij} - \hat{t}_{ij})^2}{\hat{t}_{ij}^{\,2}}\right] - \log \hat{t}_{ij} - \log \alpha - \log \sqrt{2\pi} \tag{3.30}$$

Using the method of maximum likelihood, the first order conditions for optimum parameter estimates are from equations (2.36), (2.37) and (2.38):

$$\sum_j \frac{\partial \log \phi}{\partial \hat{t}_{ij}} \hat{t}_{ij} = 0$$

$$\sum_j \left[- \frac{1}{2\alpha^2} \left(- \frac{2t^2_{ij}}{\hat{t}^3_{ij}} + \frac{2t_{ij}}{\hat{t}^2_{ij}} \right) - \frac{1}{\hat{t}_{ij}} \right] \hat{t}_{ij} = 0$$

$$\sum_j \frac{1}{\alpha^2} \left(\frac{t^2_{ij}}{\hat{t}^2_{ij}} - \frac{t_{ij}}{\hat{t}_{ij}} \right) - 1 = 0$$

$$\sum_j \frac{t_{ij}}{\hat{t}_{ij}} \left(\frac{t_{ij}}{\hat{t}_{ij}} - 1 \right) = \sum_j \alpha^2 \qquad \text{all } i \qquad (3.31)$$

Similarly

$$\sum_i \frac{t_{ij}}{\hat{t}_{ij}} \left(\frac{t_{ij}}{\hat{t}_{ij}} - 1 \right) = \sum_i \alpha^2 \qquad \text{all } j \qquad (3.32)$$

and

$$\sum_i \sum_j \left[\frac{t_{ij}}{\hat{t}_{ij}} \left(\frac{t_{ij}}{\hat{t}_{ij}} - 1 \right) \right] \frac{\partial \log f}{\partial \beta} = \sum_i \sum_j \alpha^2 \frac{\partial \log f}{\partial \beta} \qquad (3.33)$$

Furthermore, the coefficient of variation must be estimated.

$$\frac{\partial \log f}{\partial \alpha} = \frac{\partial}{\partial \alpha} \sum_i \sum_j \log \phi = 0 \qquad (3.34)$$

$$\frac{\partial}{\partial\alpha} \sum_i \sum_j \left[-\frac{1}{2}\alpha^{-2} (t_{ij}^{\,2}\, \hat{t}_{ij}^{\,-2} - 2t_{ij}\,\hat{t}_{ij}^{\,-1} + 1) - \log \hat{t}_{ij} - \log \alpha - \log \sqrt{2\pi} \right] = 0$$

$$\sum_i \sum_j \frac{1}{\alpha^3} \left(\frac{t_{ij}}{\hat{t}_{ij}} - 1 \right)^2 - \frac{1}{\alpha} = 0$$

$$\sum_i \sum_j \left(\frac{t_{ij}}{\hat{t}_{ij}} - 1 \right)^2 = \sum_i \sum_j \alpha^2 \tag{3.35}$$

$$\alpha^2 = \frac{1}{T} \sum_i \sum_j \left(\frac{t_{ij}}{\hat{t}_{ij}} - 1 \right)^2 \tag{3.36}$$

The n + 2 conditions are derived by the method of maximum likelihood on the assumption of normally distributed variables, t_{ij}, with mean $\hat{t}_{ij}$ and variance

$$\sigma^2_{ij} = \frac{1}{T} \sum_i \sum_j \left(\frac{t_{ij}}{\hat{t}_{ij}} - 1 \right) \hat{t}_{ij}^{\,2} \tag{3.37}$$

It can be seen that the maximum likelihood estimator can only derive the same conditions for optimum parameter calibration as the principle of least-squares, if the variables t_{ij} are assumed to be normally distributed with a common, constant variance. The least-squares estimator derives these same conditions, although it makes *no* assumptions about the data.

Draper and Smith (1966, pp. 60-61) suggest the maximum likelihood estimator as being appropriate if the density function of the variable is known, since the conditions derived for optimum parameter values will

will be different than least-squares conditions, with the exception of a normal probability density function with common, constant variance for all variables. If the density function is not known, however, least-squares is the better estimator to use. This is exactly what Kirby (1974, p. 102) states. The least-squares estimator is appropriate if nothing is known about the sampling distribution [meaning probability density function] of the variable t_{ij}.

However, the comparison of assumptions by the two estimators, to derive identical calibration statistics, suggests that the maximum likelihood assumptions require an overspecification of the data to derive calibration statistics. Since the least-squares estimator makes no assumption about the variable, t_{ij}, there is no behavioral hypothesis relating to trip purpose implied in the calibration statistics. The maximum likelihood estimator, on the other hand, must make a stringent, if not unrealistic, assumption about the nature of the probability density function of t_{ij}. If the assumption is not necessary, then perhaps the behavioral hypotheses "embedded" in calibration statistics are not at issue, and should not be considered when estimating parameters for different trip purposes. If this is the case, then trip purpose should only be a factor in the generalized cost function.

STATISTICAL CONDITIONS REQUIRED BY THE LEAST-SQUARES ESTIMATOR

The conditions derived by least-squares, defining statistics against which the model parameters are calibrated, are more complex than maximum likelihood conditions. Because the conditions are functions of the sums of the *squared* row and column elements, the calibration statistics cannot be directly related to observable phenomena in the sample matrix.

Consider the calibration statistic derived by least-squares for the parameter β

$$\sum_i \sum_j \hat{t}_{ij}^{\,2} \frac{\partial \log f}{\partial \beta} = \sum_i \sum_j \hat{t}_{ij} \, t_{ij} \frac{\partial \log f}{\partial \beta} \qquad (3.38)$$

If we apply the same transformation on the generalized cost function as in the previous chapter,

$$f(\beta, c_{ij}) = \exp(-\beta \, h(c_{ij})) \qquad (3.39)$$

equation (3.38) becomes

$$\sum_i \sum_j \hat{t}_{ij}^{\,2} \, h(c_{ij}) = \sum_i \sum_j \hat{t}_{ij} \, t_{ij} \, h(c_{ij}) \qquad (3.40)$$

Least-squares conditions do not calibrate β against a mean value that can be calculated from the data, as in maximum likelihood conditions, since the model predictions, $\hat{t}_{ij}$, enter both sides of the equation (i.e., (3.41)).

The sensitivity of the calibration statistic has not been discussed in the literature, although the least-squares estimator has been used to calibrate doubly constrained spatial interaction models (Tanner, 1961; Cesario, 1975). However, Turner (1970) reports that parameter values for singly constrained interaction models, calibrated by least-squares, may converge to a local optimum, giving a false solution (Batty and Mackie, 1972, p. 210).

In general, the least-squares estimator calibrates the parameter against a statistic which is a function of the model predictions, $\hat{t}_{ij}$, the observation, t_{ij}, and the transformation on cost, the latter being assumed to be specific to trip purpose (Hyman, 1969, p. 109). The balancing factors, a_i and b_j are determined so that there is consistency for the sum of squares of row and column elements between the prediction and the data.

SUMMARY

This chapter has discussed another approach to point estimation, the method of least-squares. It has shown that the least-squares estimator derives conditions for optimum values without making any assumptions about the probability density function of the variable, t_{ij}. Furthermore, it has shown that the maximum likelihood estimator can derive the same set of conditions only if very restrictive assumptions are made about the nature of the density function, and that any attempt to relax these assumptions results in different conditions on the parameters. From this study, it is suggested that maximum likelihood assumptions may not relate to behavioral hypotheses concerning trip purpose. The chapter concludes by examining the least-squares statistics and interpreting their meaning with respect to the sample.

CHAPTER 4

SOME RELATED CALIBRATION PROBLEMS IN URBAN SYSTEMS MODELLING

INTRODUCTION

This chapter will discuss the assumptions made by the modeller when calibrating a model of spatial interaction. These are distinct from the assumptions on the data implied by the statistical estimators, reviewed in the previous two chapters. Instead, they are the assumptions which must be made about the *observation*, or trip survey, and about the *variables* which must be defined in order to give the parameters numerical values.

The chapter will examine these assumptions by first discussing the information available to the modeller from the trip survey in terms of sample size and method of sampling. It will then discuss the extent to which the observations may deviate from the actual or *mean* values, especially in a traffic survey, where a single random sample is usually taken to calibrate the model. It will also look at how *generalised cost* is usually defined in the calibration statistics derived by maximum likelihood and least-squares, and will outline some weaknesses to the approach.

The chapter will go on to review the basic criticisms which have been raised throughout the paper. It will identify three research areas which should be examined. The paper will then propose a framework for testing the hypotheses.

THE "PROBLEM" OF MODEL CALIBRATION

The urban modeller is faced with the problem of applying a mathematical abstraction of a physical system to a defined set of activities, from which must be generated empirically-relevant output (Lowry, 1965, p. 160). The modeller assumes that a specifically chosen hypothesis sufficiently describes the phenomenon he is studying. The "problem" of model calibration is thus one of *defining* the variables against which to estimate the parameters of the model, and of optimally "fitting" the hypothesis to sample data.

The accuracy of the spatial interaction model's output is closely related to the reliability of the sample data, or trip survey, for it is against these data that the model is calibrated, regardless of which calibration statistics are employed. The modeller takes these data, usually a small percentage of the population (Chaterjee, *et al.*, 1974, p. 3), as a satisfactory representation of the actual distribution, or mean values, he is attempting to mathematically describe. The model is calibrated to these data, or more precisely, the model parameters are estimated from the observation.

Given the parameter values determined from the sample, the model is applied to the population from which the sample is drawn, to generate the existing distribution, or can be applied to projected activity variables to conditionally predict future distributions.

The model predictions are conditional upon three basic assumptions.

1. The hypothesis is a suitable description of the phenomenon under study.

2. The survey, against which the model is calibrated, is a valid representation of the phenomenon under study.

3. The parameter values determined from the sample data, and the functional relationships describing behavior, remain constant over *scale* and over *time* (if used to make conditional predictions in the future).

This paper does not intend to test the validity of the first assumption. It assumes that the spatial interaction model reasonably describes the distribution of travel in an urban area. Nor is it concerned with the effects of scale or time on the generated distribution. Although these two issues must be more thoroughly researched (Shepherd, 1974, pp. 52-68; Wilson, 1974, p. 391), the issue which will be discussed in the chapter is the reliability of the sample data.

This issue will be examined in two parts. One will deal with the trip survey, and the other will examine and assess the variables which give the parameters numerical values.

THE TRAVEL SURVEY AND ITS RELATIONSHIP TO THE POPULATION IT REPRESENTS

Travel data, used to calibrate trip distribution models, are obtained primarily from origin-destination surveys, of which there are two kinds. These are *home-interview* and *roadside-interview* surveys. To determine existing internal travel patterns, home-interview surveys are usually conducted (Chaterjee, *et al.*, 1974, p. 1). The area is divided into a set of zones, and the survey is conducted by interviewing a small percentage of households in each zone randomly. In trip distribution modelling, the survey finds the destination zone of each household for the particular trip purpose being modelled.

After the survey has been taken, the results are aggregated into a *sample trip matrix*, which describes the interaction in the system observed in the survey. From this matrix, calibration statistics, defined by the statistical estimator used to estimate the parameters, are calculated.

The issue to be discussed is the relation the sample bears to the mean distribution of the population. In the context of trip distribution, the mean distribution is the average travel between each origin and destination, for a specified trip purpose and in a defined time period, i.e., journey-to-work trips in a two-hour peak period. The major factor influencing the correspondence between the observed and actual distributions is the *sample size of observations*. The survey data upon which the parameters are estimated are considered to be a random sample of T variables, and of sample size equal to one. The survey, which is but one estimate of the number of journeys between each zone pair for a given time period, may or may not correspond to the actual travel pattern.

The actual number of trips between each i-j pair varies from day-to-day, and can be represented by a probability density function, ϕ, with mean $\hat{t}_{ij}$ and variance σ_{ij}^2. The *sampling distribution* of the mean travel on each interchange, is related to the actual distribution by the following fundamental relationship.

Consider Z random samples taken from a population having mean $\hat{t}_{ij}$ and variance σ_{ij}^2. The mean value, $\bar{x}$, of the random samples, will be distributed in a sampling distribution with mean

$$E(\bar{x}) = \hat{t}_{ij} \tag{4.1}$$

and variance

$$\sigma^2(\bar{x}) = \frac{\sigma_{ij}^2}{Z} \left(\frac{N - Z}{N - 1}\right) \tag{4.2}$$

when taken from a population of finite size, N (Freund, 1952, p. 230). These relationships reveal that, *on the average*, the sample mean equals the population mean, and the variability of the sample mean is equal to or less than the variability of the random variable of the population. The variability of the sample mean decreases as the number of random samples taken increases (Freund, 1952, p. 231).

Since, in actual studies, only one trip survey is usually taken (Kirby, 1974, p. 99), the variance of the sample mean is as large as the variance of the variable t_{ij} in the population. This variance may be large for certain trip purposes (Kirby, 1974, p. 99). Therefore, the data, against which the model is calibrated, may significantly misrepresent the mean travel distribution.

This suggests that better data for calibrating spatial interaction models can be obtained simply by taking more than a single trip survey in the area being modelled. Since trip surveys involve a considerable expense, in terms of time and money, the correspondence between a single random sample and the mean travel distribution should be investigated to determine whether reliable model predictions can be generated on the basis of a single sample.

VARIABLE DEFINITION IN CALIBRATION STATISTICS

The precise empirical definition of variables is important because it affects the values of the model parameters (Lowry, 1965, p. 163). The relationship between variable definition and parameter values can be seen in the calibration statistics derived by the statistical estimators in the previous two chapters.

The maximum likelihood estimator, under the assumption of a Poisson density function, calibrates the parameter β against a transformation of the *generalised cost of travel* between i and j.

$$\sum_i \sum_j \hat{t}_{ij}\, h(c_{ij}) = \sum_i \sum_j t_{ij}\, h(c_{ij}) \qquad (4.3)$$

where

$$h(c_{ij}) = \frac{\partial \log f}{\partial \beta} \tag{4.4}$$

The least-squares estimator calibrates β against the same transformation of cost.

$$\sum_i \sum_j \hat{t}_{ij}\, h(c_{ij}) = \sum_i \sum_j \hat{t}_{ij}\, t_{ij}\, h(c_{ij}) \tag{4.5}$$

In order to derive numerical values for the parameters, it is necessary to define the form of the generalized cost function, and to define generalized cost itself. Hyman (1969, pp. 108-109) suggests several cost functions and relates these to different trip purposes. Wilson (1974, p. 70) argues that if a function of the form

$$f(\beta_1\, c_{ij}) = \exp\,(-\beta c_{ij}) \tag{4.6}$$

is used in the spatial interaction model, the traveller is perceiving cost linearly. Several authors (Batty, 1970c; Batty, 1971; Batty, Foot, *et al.*, 1974) have used this cost function in model application. Under this assumption, the transformation on cost becomes

$$h(c_{ij}) = c_{ij} \tag{4.7}$$

and the calibration statistics, from (4.3) and (4.4), become

$$\sum_i \sum_j \hat{t}_{ij}\, c_{ij} = \sum_i \sum_j t_{ij}\, c_{ij} \tag{4.8}$$

and

$$\sum_i \sum_j \hat{t}_{ij}^{\,2}\, c_{ij} = \sum_i \sum_j \hat{t}_{ij}\, t_{ij}\, c_{ij} \qquad (4.9)$$

Wilson (1974, p. 70) argues that a power function of the form

$$f(\beta,\, c_{ij}) = c_{ij}^{\;-\beta} \qquad (4.10)$$

may be more appropriate for long distance travel, since marginal travel over long distances is not likely to be perceived in the same linear fashion. O'Sullivan (1968) has used this cost function to describe inter-regional freight flows. In this case, the transformation on cost becomes

$$h(c_{ij}) = \log c_{ij} \qquad (4.11)$$

and the calibration statistics are, from (4.3) and (4.5)

$$\sum_i \sum_j \hat{t}_{ij} \log c_{ij} = \sum_i \sum_j t_{ij} \log c_{ij} \qquad (4.12)$$

and

$$\sum_i \sum_j \hat{t}_{ij}^{\,2} \log c_{ij} = \sum_i \sum_j \hat{t}_{ij}\, t_{ij} \log c_{ij} \qquad (4.13)$$

Other authors (Alonso, 1972; Batty and Mackie, 1972; Batty and Mackie, 1973) have used the Tanner model, where the cost function is specified by

$$f(\beta,\ c_{ij}) = c_{ij}^{-\beta} \exp\ (-\mu\ c_{ij}) \tag{4.14}$$

Both the maximum likelihood and least-squares estimators derive statistics for each parameter. The statistics by maximum likelihood are (Batty and Mackie, 1973):

$$\sum_i \sum_j \hat{t}_{ij}\ c_{ij} = \sum_i \sum_j t_{ij}\ c_{ij} \tag{4.15}$$

and

$$\sum_i \sum_j \hat{t}_{ij} \log c_{ij} = \sum_i \sum_j t_{ij} \log c_{ij} \tag{4.16}$$

The statistics from least-squares are:

$$\sum_i \sum_j \hat{t}_{ij}^{\,2}\ c_{ij} = \sum_i \sum_j \hat{t}_{ij}\ t_{ij}\ c_{ij} \tag{4.17}$$

$$\sum_i \sum_j \hat{t}_{ij}^{\,2} \log c_{ij} = \sum_i \sum_j \hat{t}_{ij}\ t_{ij} \log c_{ij} \tag{4.18}$$

In the literature, careful consideration has been given to specifying the form of the cost *function*. However, generalized cost, in most readings, has been defined simply as *trip length*, i.e., trip distance. This is a departure from the definition of "cost" in earlier studies. Wilson, *et al.*, (1969) assume travel cost to be a linear function of several factors: travel time, waiting time (for transit), trip length, parking costs, and a modal "penalty". The precise definition of cost in this study is a result of extensive work in people's valuation of cost for different modes of transportation (Wilson, *et al.*, 1969, pp. 341-342). Although Wilson (1972, p. 14) states that for trip distribution purposes, the cost function can be truncated to the form,

$$c_{ij} = a_1 b_{ij} + a_2 c_{ij} + a_3 d_{ij} \tag{4.19}$$

where b_{ij} = travel time

c_{ij} = excess waiting time

d_{ij} = trip length

a_1, a_2, a_3 = parameters to be estimated (by regression)

There is no evidence in the literature which indicates trip length to be an appropriate surrogate for generalised cost of travel. Trip length would appear to be a poor measure of travel cost under conditions of congestion in the urban area, or when political strategies, such as increasing parking costs in the CBD are involved.

From equation (4.7), it can be seen that if trip length is used as a surrogate for trip cost, calibration statistic may be readily calculated from the sample data, since inter-zonal distances are easily determined.

From (4.8)

$$\frac{1}{T} \sum_i \sum_j \hat{t}_{ij} c_{ij} = \frac{1}{T} \sum_i \sum_j t_{ij} c_{ij} \tag{4.20}$$

$$\sum_i \sum_j \hat{p}_{ij} c_{ij} = \bar{C} \tag{4.21}$$

where $\bar{C}$ = mean trip length (observed).

Although the mean trip length statistic is sensitive to changes in parameter values (Batty, 1970c, p. 109), it is not a valid statistic unless the higher moments of the distribution determine the t_{ij} density function to be Poisson. Other conditions may make the statistic unsuitable as well. For example, shopping models cannot be calibrated to mean trip length because the trip pattern, *a priori*, is not known, (Openshaw, 1973, p. 367).

Variable definition is important in determining parameter values, and parameter estimation is closely related to model performance (Openshaw, 1973, p. 367). Empirical studies, therefore, should be directed towards the precise definition of the generalized cost variable in calibration statistics, so as to improve model performance in application.

A REVIEW OF PROBLEMS IN MODEL CALIBRATION

From the preceding discussion in chapters two, three and four, three distinct problem areas can be identified. Chapters two and three have reviewed two competing statistical methods of parameter estimation. Each method makes different assumptions about the survey data. The maximum likelihood estimator assumes the t_{ij} to vary on each interchange by a *specified* probability density function. The specified density function can be related to trip purpose, since the variance of travel over a given interchange is different for different trip purposes. Because the specified density function yields a unique set of calibration statistics, optimum parameter values, derived from the maximum likelihood estimator, depend upon the travel purpose being modelled.

The least-squares estimator, on the other hand, makes no assumptions about the data. Only one set of calibration statistics are derived, regardless of the data. Therefore, the optimum parameter values derived by least-squares are not dependent upon the probability density function of t_{ij}, and thus do not require any assumptions about trip purpose. Whereas the maximum likelihood estimator makes very restrictive assumptions about the data, the least-squares estimator makes very weak ones.

Does this mean that if the data really do arise in the way postulated by the maximum likelihood estimator, the parameter values derived by this method will give the data a better fit to the data than the least-squares estimator? If so, then Kirby's (1974) assertion that trip purpose, and hence its characteristic probability density function, must be specified, before optimum parameter values can be derived, is correct. The problem is thus to determine whether these assumptions are necessary to give the model a best fit to the data.

Another problem area concerns the relationship the random sample bears to the actual distribution from which it is taken. Assuming there is some variation in travel between each interchange, typically with mean $\hat{t}_{ij}$ and variance σ_{ij}^2, it has been shown that for a *single* trip survey, the variance of the sample mean is as large as the variance in the actual travel on the interchange. Since the variation on a given interchange can be large, a single random sample may not be a reliable representation of the mean travel distribution, which the model is trying to predict. Since the parameters are calibrated against the sample, it is important to determine whether a suitable correspondence between the sample means and actual means exists.

The third problem area concerns the definition of variables in calibration statistics. It has been stated that parameter values depend upon the empirical definition of the model variables (Lowry, 1965, p. 163), and that model performance is dependent upon the parameter values derived by the calibration statistics. If trip length is not a suitable surrogate for the generalized cost variable in the maximum likelihood and least-squares calibration statistics, then calibration against the variable yields a predicted distribution having a sub-optimal fit to the survey data. The problem is therefore to empirically determine what generalized travel cost is for calibration purposes.

The next section will propose a research design to resolve some of these issues.

A RESEARCH DESIGN TO EXAMINE THE ISSUES IN CALIBRATION

This paper now proposes to construct a framework upon which to examine two of the issues discussed in the previous section. The framework is designed to test whether the assumptions about the data, implied by the maximum likelihood estimator, derive parameter values which give the model a better fit to the sample than least-squares, which makes no implicit assumptions. The framework is also designed to assess whether a single random sample is an adequate representation of the actual distribution, upon which to calibrate spatial interaction models. The issue of variable definition must be resolved by empirical examination. In the proposed framework, which is a hypothetical example, this factor cannot be studied, but can be *controlled*, to prevent it from biasing the results of the other two issues.

The paper proposes a hypothetical example, consisting of a "typical" urban area of moderate size, divided into a set of zones of equal area. We are given the number of origins and destinations in each zone, and the generalised cost of travel on each interchange is assumed to be the interzonal distances. Given this information, we intend to use a specified function of cost and parameter value, β, to define a distribution of trips in an urban area.

By constructing the framework in this manner, we possess more data than does the modeller when he applies the model in an empirical study. First, we know the actual distribution of travel in the urban system. Second, we know the cost function that determines the impedence to travel in the system. Finally, we know how cost is defined in the distribution.

If we take a random sample from the generated distribution, we can (1) make a statistical measure of correspondence between the sample and the actual trip distribution, and (2) calibrate a spatial interaction model of the same form with the statistical estimators reviewed in chapters two and three, so as to compare the distributions generated from the estimated parameters with the sample.

We have information about several factors in the calibration procedure which usually must be assumed in practical calibration applications, i.e., the function of generalised cost and the definition of cost in the calibration statistics. Since these factors are controlled in the proposed framework, any differences between the estimated distributions (by maximum likelihood and least squares), will result from the calibration statistics only. This enables us to evaluate the performance of the two estimators against the sample.

The next chapter will define specific hypotheses in the two areas of research, and will describe the hypothetical urban area to be modelled. This will enable us to use the framework for analysing the problems.

CHAPTER 5

AN ANALYSIS OF THE RELATIONSHIP OF THE TRAVEL SURVEY TO THE MEAN TRAVEL DISTRIBUTION, AND OF THE EFFECT OF BEHAVIORAL CONSIDERATIONS ON MODEL PERFORMANCE

INTRODUCTION

Chapter four defined the framework upon which the two calibration issues are to be tested. It now remains to clearly define the hypotheses concerning the relationship of the sample data to the mean travel distribution, and the goodness-of-fit of the model predictions using different methods of parameter estimation, to the trip survey.

Travel is to be distributed in a hypothetical urban system. The day-to-day variation of trips will be described by a specified probability density function. A four per cent random sample will then be taken from a distribution of travel on a "given" day. The analysis will consist of two operations. In the first, the correspondence of the random sample to the mean travel distribution will be examined. The results of this test will give us an indication of whether the single trip survey sufficiently represents the mean distribution for calibrating the spatial interaction model, since, in fact, this is the distribution the model is intended to predict.

Second, the spatial interaction model will be calibrated to the sample data by both the maximum likelihood and least-squares statistical estimators. The intention is to draw the sample data from travel distributions whose elements are dispersed around mean values, according to specified probability density functions. In the first example, the density functions of the matrix elements, the t_{ij}, will be Poisson. In the second, the density functions will be normal with common, constant variance. The correspondence of the model predictions (using the parameters estimated by the two estimation techniques) to the random sample, will then be examined.

The maximum likelihood statistics will make the same assumptions about the data in both examples. They will assume a Poisson density function for each t_{ij}. The least-squares statistics, by definition, will not change in the two examples. However, in the second case, the least-squares statistics will be identical to maximum likelihood statistics, which assume a normal density function for the t_{ij}, with common, constant variance. This is exactly how the data will occur.

We should, therefore, expect that if consideration of trip purpose and its characteristic probability density function is important in deriving optimum parameter values, which give the model a best fit to the data, then the maximum likelihood estimator should give the model a better fit to the data in the first example, and the least-squares estimator should give the model a better fit in the second example. If, however, the difference in goodness-of-fit between the models, predicted by the two estimators, is not significant, then we can conclude that trip purpose is not a necessary consideration in model calibration, and the specification of unique calibration statistics for specific trip purposes is not necessary to derive optimum parameter values.

These issues can formally be stated as hypotheses to be tested in the analysis. Given the conditions in the experiment:

1. A single trip survey is a reliable representation of the mean distribution of travel. This will be considered verified if there is a high degree of correspondence (measured by some specified criterion) between the sample matrix and the mean trip matrix.

2. The calibration technique whose assumptions are appropriate to the conditions of the data provides better parameter estimates. This will be considered verified if the parameter values generate a distribution giving the model a significantly better fit (measured by some specified criterion) to the sample.

If the second hypothesis is true, then trip purpose, distinguished by its characteristic probability density function describing the variance of the t_{ij} elements over each interchange, should be incorporated into calibration statistics.

The analysis has been developed to examine these hypotheses. However, before the analysis is performed, the criterion for measuring goodness-of-fit must be defined and its assumptions noted, to specify the limits to the conclusions which are to be drawn from this study.

THE RELATIONSHIP BETWEEN OPTIMUM FIT AND OPTIMUM PARAMETER VALUES

The statistical estimators, discussed in chapters two and three, determine optimum parameter values which may or may not give the model an optimum fit to the sample data. In estimating parameters by these techniques, we must be aware of the two alternative meanings of "best" parameter values. The parameters may be the best estimates in that they are optimum with respect to the statistical estimator. Alternatively, they may be the best estimates because they give the model an optimum fit to the sample data. Is there a unique solution involved, and if so, what conditions are necessary to statistically estimate parameter values which give the model an optimum goodness-of-fit to the sample data?

Initially, one must determine whether the optimality conditions of a statistical estimator are sufficient for optimising the model's goodness-of-fit. Wilson (1974, p. 320) states that after deriving parameter values by a statistical estimation technique, such as maximum likelihood or least-squares, we should measure the model's goodness-of-fit to the sample data with a correlation statistic, such as the coefficient of determination (R^2), or chi-squared (χ^2), both of which measure the correspondence of the model's output to the data. This, he suggests, gives us the overall indication we need to determine whether the parameters derived by optimising the statistic generate a distribution which corresponds optimally to the data. It also helps us choose between different forms of function, such as the travel impedance function.

It remains unclear why correlation statistics should be more reliable in measuring the model's goodness-of-fit than maximum likelihood or least-squares methods. In them, the measures of correspondence between the predicted variables and the sample data are still based upon implicit assumptions about the data (see Chapter 1). Batty (1970c, p. 112) applies the coefficient of determination to measure goodness-of-fit and assumes the statistic to be normally distributed. Yet there is no evidence which supports this assumption. Given the inconsistencies which can arise in the application of this statistic (page 9), it is presumptuous to make any assumptions about its distribution without making a thorough investigation of its properties. Since the statistic is unbiased, consistent, and efficient *only* if the data occur in the way postulated by the statistic, then one cannot infer that it gives any better indication of goodness-of-fit than statistical estimators, such as maximum likelihood or least-squares.

The assumptions made by the chi-squared statistic about the data are not as strict as those made by other statistics. Also, its distribution is such, that as the number of degrees of freedom on the statistic increases, the distribution of chi-squared approaches the normal. In spatial interaction modelling, there is likely to be a large number of degrees of freedom when using the statistic, due to the number of variables involved. In a system of n-origins and n-destinations, the chi-squared statistic is calculated over n^2 variables, i.e., the t_{ij}. The model which is compared to the data is calculated from n origin-specific factors, a_i, n destination-specific factors, b_j, and the parameter β. This reduces the number of degrees of freedom on the statistic to

(Wetherill, 1967, p. 202)

$$n^2 - 1 - (n + n + 1) = n^2 - 2n - 2 \tag{5.1}$$

For a system of fifty origins and fifty destinations, the size of the hypothetical system in the analysis, the statistic has 2400 degrees of freedom, which means the distribution of chi-squared is approximately normal.

This arbitrarily imposes a normal density function on the t_{ij}, which may be inappropriate in spatial interaction modelling. Therefore, its measure of fit may be no more valid than any other statistical measure. Wetherill (1967, p. 203) notes that there are often better tests which may be used when the data have some other than a normal distribution. Furthermore, other restrictions concerning expected frequencies (Wetherill, 1967, p. 203) make the statistic even less attractive to apply in the spatial interaction context.

How, then, can we measure whether optimum parameter values, derived from a statistical estimator, generate a distribution which yields an optimum fit to the data? Since the correlation statistics discussed above may not be reliable for this purpose, they cannot resolve this issue.

We still require a statistic to measure the correspondence between predictions by the maximum likelihood and least-squares estimators. This paper proposes to use the coefficient of determination to see if the maximum likelihood and least-squares estimators generate predictions which are *similar*, not to make any inferences as to the predictions' goodness-of-fit. By using this correlation statistic merely to get a "feel" for the similarity or dissimilarity of the two distributions, we are remaining consistent in the use of the statistic.

THE PHYSICAL SETTING FOR THE ANALYSIS

The idealized urban area to be modelled is shown in Figure 4. It is an abstraction of a "typical" urban system and is intended to be an instrument through which a specified set of trips can be distributed. From the distribution, a trip survey will be taken to calibrate the spatial interaction model by the methods of maximum likelihood and least-squares.

The area is divided into fifty zones, each of which is equal in size. Each zone is numbered for identification, and is manually allocated specified proportions of trip-origins and destinations. This operation doubly-constrains the distribution of trips in the system. The zones can be aggregated to form several distinct sub-areas (outlined by dark boundaries), which reflect different land use characteristics of a typical urban area. The outer sub-areas represent the suburban or residential sectors of the "city", and are characterised by many origins and few destinations in each zone. The inner sub-areas surrounding the central area have a greater number of destinations but still generate a significant volume of travel. The central area is the CBD, with many destinations and few trip-origins. The proportions of origins and destinations allocated to each zone are designed to reflect a gradual increase in destinations towards the central area from the

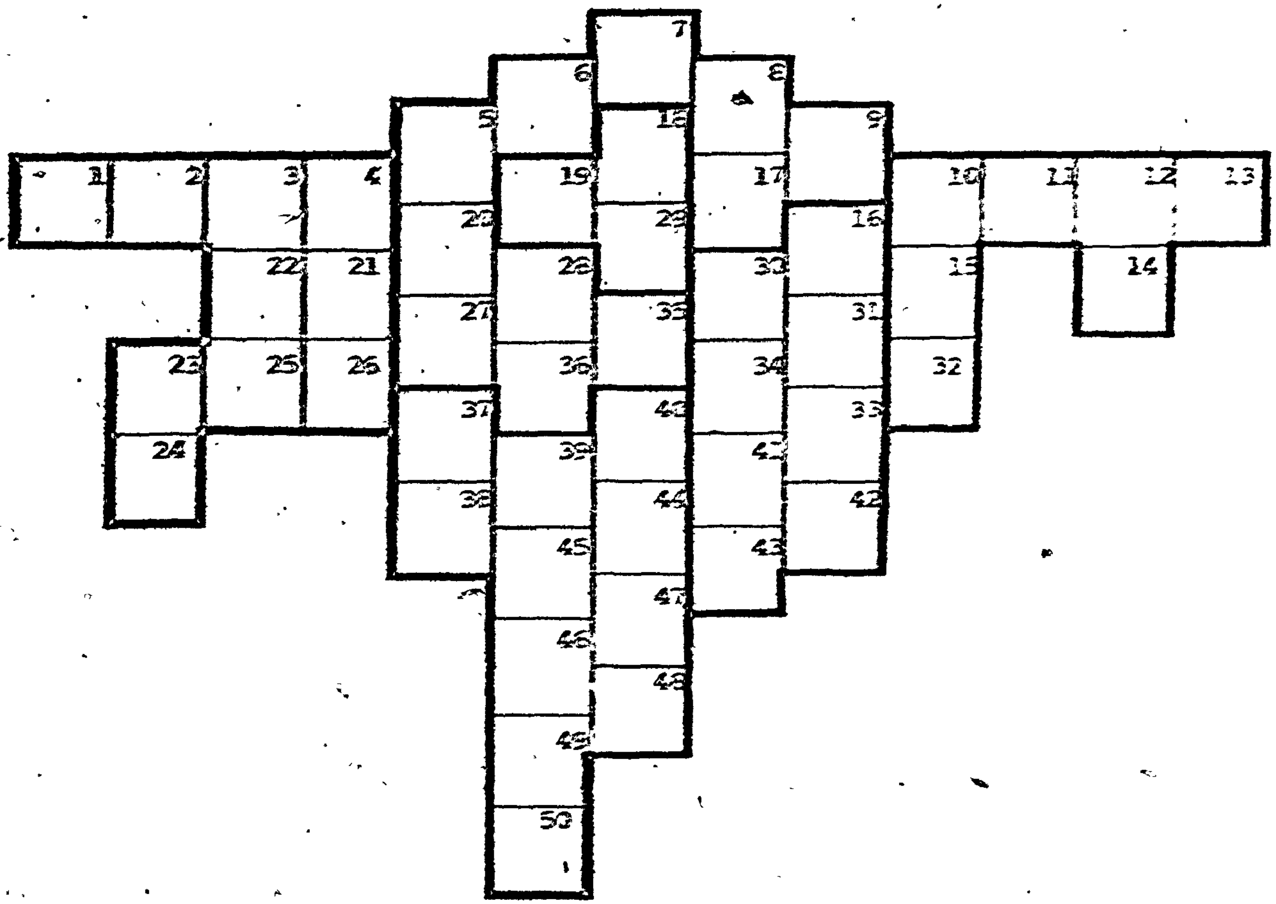

FIGURE 4: Configuration and Zoning of the Idealized Urban Area

outskirts (Table 1).

Although this is an idealized representation of the urban system, it enables us to generate a distribution of trips over the area and to avoid many problems which beset modellers in empirical studies. The configuration of the area is designed to be assymmetrical to prevent any trivial solutions from occurring in the generation of trips or the calibration of the model. By dividing the area into zones of equal size, and by strictly defining the number of origins and destinations in each, instead of using proxies for attractiveness, biases are prevented from entering the problem (Wilson, 1974, p. 69).

In the idealized system, we are assuming the zone size to be small enough to account for all of the potential inter-zonal interaction (Batty, Foot, *et al.*, 1973, pp. 353-354). Also, the example considers internal travel only. This avoids the related problems of dummy zones and closure, which usually must be taken into account in spatial interaction modelling (Batty, Foot, *et al.*, 1973, pp. 362-364).

The distribution to be generated and modelled is for a single broad classification of trip purpose. From the allocation of origins and destinations in the system, the distribution that results may well characterize journeys-to-work. Travel is considered in one direction only, from home to work. This is the general modelling procedure in urban transportation studies (Ben-Akiva, 1973, p. 34).

Under these idealized conditions, 100,000 trips will be generated in the system by a spatial interaction model with known parameter values. The generated distribution will be defined to be the *mean* distribution of daily trips that occur over a specified time period (e.g., a year)

TABLE 1

TRIP ORIGINS AND DESTINATIONS ALLOCATED TO EACH ZONE

ZONE	ORIGINS	DESTINATIONS	ZONE	ORIGINS	DESTINATIONS
1	2,000	56	26	3,500	102
2	1,299	226	27	1,900	907
3	3,200	68	28	1,099	1,248
4	3,599	226	29	1,499	13,619
5	850	2,496	30	2,000	2,723
6	1,000	2,837	31	1,699	1,021
7	700	5,674	32	2,899	90
8	800	3,404	33	1,099	3,858
9	2,100	1,361	34	2,200	1,134
10	3,800	124	35	2,099	453
11	3,499	90	36	2,499	113
12	1,600	11	37	2,700	873
13	2,100	11	38	3,000	771
14	1,000	22	39	3,399	726
15	3,100	1,021	40	3,300	158
16	2,500	2,383	41	1,799	1,248
17	899	1,929	42	1,000	4,653
18	500	20,429	43	1,700	4,426
19	1,000	11,349	44	2,499	1,430
20	1,050	2,269	45	3,900	590
21	2,100	124	46	2,000	408
22	4,000	136	47	1,399	1,634
23	2,299	68	48	1,000	964
24	1,600	45	49	1,200	249
25	1,400	79	50	600	136

in the urban area. The daily number of journeys on each interchange will be allowed to vary from this mean by, first, a Poisson density function and second, a normal density function. It is from these distributions that four per cent random samples will be drawn. The spatial interaction model will then be calibrated by the statistical estimators against these data, to determine the optimum parameter values. The algorithms which perform these operations will be described in the following sections.

A DESCRIPTION OF THE ALGORITHMS IN THE ANALYSIS

The research design is basically a controlled experiment consisting of two parts, each having a different function. The first part involves *system simulation.* Its function is to generate the variables with which to make *correspondence measures.* These measures are necessary to evaluate the competing hypotheses.

The sequence of operations in the analysis is shown in Figure 5. The remainder of this section briefly describes each of the operations.

System Simulation:

The first operation in the analysis is DISTRIBUTION. Its function is to generate travel throughout the hypothetical urban area. The trip matrix is generated by a spatial interaction model of the form:

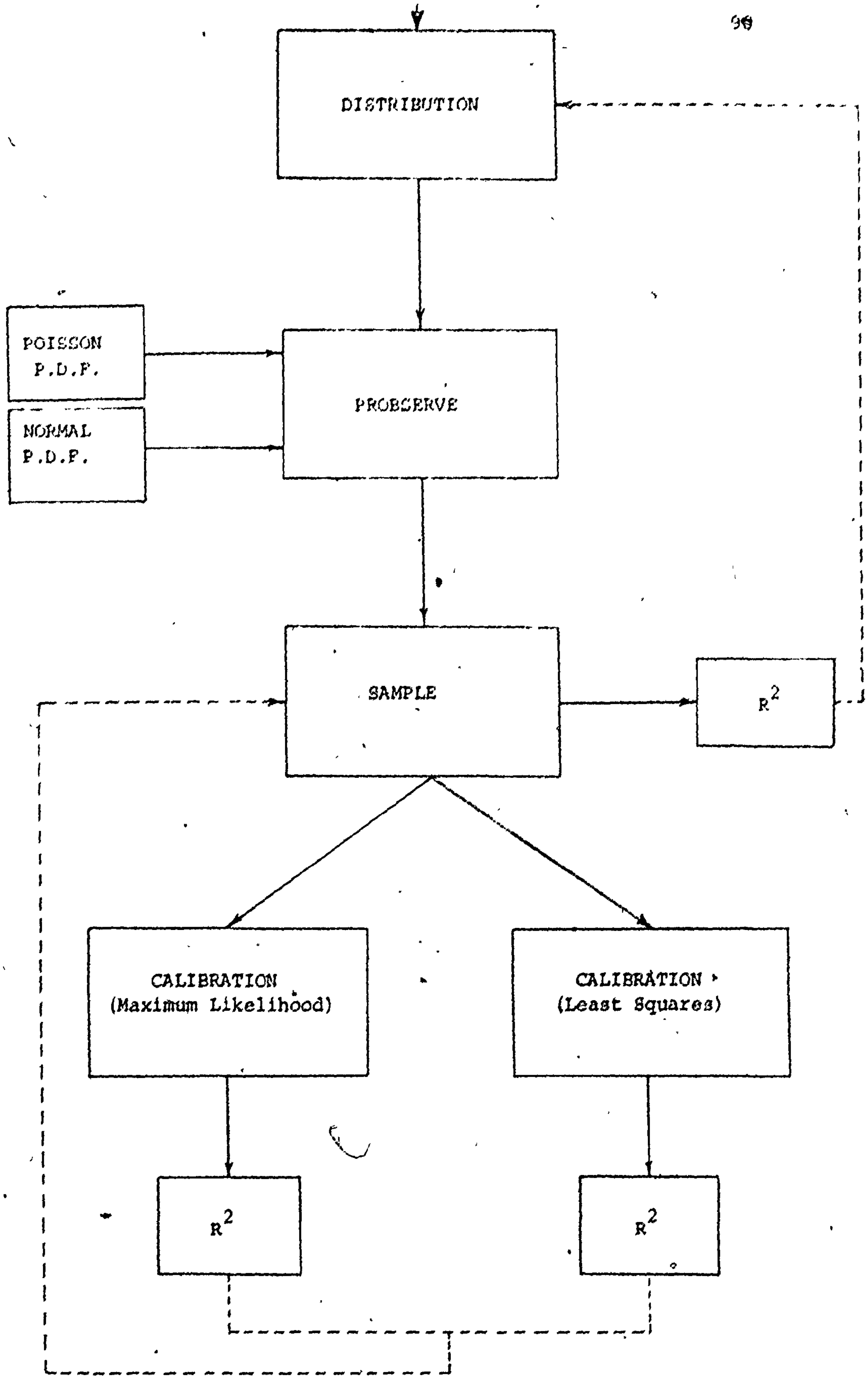

FIGURE 5: Sequence of Operations in the Analysis

$$t_{ij} = A_i\ O_i\ B_j\ D_j\ \exp\ (-\beta d_{ij}) \tag{5.2}$$

where O_i = the number of trip origins in zone i

D_j = the number of trip-destinations in zone j

d_{ij} = the distance between zones i and j

β = a parameter measuring the extent to which travel is considered

A_i, B_j = origin-specific and destination-specific balancing factors.

The value of β is arbitrarily specified ($\beta = 0.02$). Since β is regarded as a measure of the extent to which distance (in our case) is considered when travel decisions are made (Evans, 1973, p. 40), we expect this parameter value to affect the distribution of trips in such a way that the mean trip length becomes longer. This tends to promote more travel from the peripheral suburban areas over longer distances to the concentrated employment areas in and around the CBD. This pattern of travel would be expected over an efficient transportation network. The balancing factors are determined for this parameter value from the following equations (Wilson, 1970, p. 16).

$$A_i = \frac{1}{\sum_j B_j\ D_j\ \exp\ (-\beta d_{ij})} \qquad \text{for all } i \tag{5.3}$$

$$B_j = \frac{1}{\sum_i A_i\ O_i\ \exp\ (-\beta d_{ij})} \qquad \text{for all } j \tag{5.4}$$

The model distributes 100,000 trips throughout the fifty zone system on 2500 interchanges. The variables which make up the resulting 50 x 50 trip matrix represent the *mean* travel generated over these interchanges.

Since we expect the volume of travel to fluctuate on any given interchange, the function of PROBSERVE is to construct a new trip matrix which reflects this day-to-day variation. The operation assumes that a specified probability density function describes the variation in travel on each interchange, i.e., Poisson or normal, and uses random numbers to construct a trip matrix which would be likely to result if travel varied in this manner.

SAMPLE takes a four per cent "home-interview survey" which is used to calibrate the model. The origins in each zone are selected at random and their destinations are tabulated. The results are aggregated to produce a sample trip matrix, from which calibration statistics, specific to the maximum likelihood or least-squares estimator, are calculated.

These statistics are input into CALIBRATION to estimate the model parameter β, and balancing factors A_i and B_j for each statistical estimation technique. Trip matrices, generated by these estimated parameter values, are subsequently constructed and are input into the second stage of the analysis.

Correspondence Measures:

Only one measure of correspondence is used to test the hypotheses. In spite of its apparent weaknesses (pp. 17-18), correspondence between trip matrices is measured by the *coefficient of determination*. This statistic is chosen primarily because it is computationally simple and relatively easy to interpret. Although more reliable measures of fit are available, such as the *chi-squared statistic* or the *"expected information" statistic* (Morphet, 1975), these are not easily applicable, since many of the zonal interchanges in the sample trip matrix and the generated trip matrices are zero. Interchanges with t_{ij} elements equal to zero make these statistics undefined. To use the chi-squared or "expected information" statistic requires the removal of zero elements in the trip matrix, either by zonal aggregation or by excluding these elements from the analysis. Both of these methods, then, measure the correspondence between matrices on reduced information. For the purposes of this analysis, the coefficient of determination is the most satisfactory measure of correspondence of those taken into consideration.

The reliability of the trip survey in representing the mean distribution of trips is determined by measuring the correspondence between the sample trip matrices generated in SAMPLE and the mean trip matrix generated by DISTRIBUTION. The maximum likelihood and least-squares estimators are compared by measuring the correspondence between the trip matrices generated by the parameters derived in CALIBRATION, and the sample trip matrix output from SAMPLE.

These measures of correspondence should enable us to examine whether the hypotheses defined above are correct.

THE RELIABILITY OF A SINGLE TRIP SURVEY IN MODEL CALIBRATION

Two preliminary tests were made to show the correspondence of the trip matrices with elements varying by Poisson and normal density functions, defined as D_P and D_N respectively, to the mean trip matrix. One test measured the correspondence between the actual population-sized matrices. The other measured the correspondence between sample-sized matrices.

In the first test, it was found that the correspondence of both D_P and D_N to the population-sized mean distribution was very close. The R^2 value of D_P to the mean distribution was .9958. The R^2 value of D_N to the mean was .9998.

The second test measured the correspondence of four per cent samples drawn from D_P and D_N to a four per cent sample drawn from the mean distribution. In this case, the R^2 value of the Poisson sample to the mean sample was .9636. The R^2 value of the normal sample to the mean sample was somewhat higher, at .9878.

In order to determine whether these differences in the R^2 values are significant, it is necessary to examine the structure of the statistic and the characteristics of the trip interchange data it is measuring. Consider the R^2 statistic.

$$R^2 = 1 - \frac{\sum_i \sum_j (\hat{t}_{ij} - t_{ij})^2}{\sum_i \sum_j (t_{ij} - \frac{1}{N} \sum_i \sum_j t_{ij})^2} . \tag{5.5}$$

Since N is the number of variables, t_{ij}, the denominator of this expression is simply the sum of squared deviations of t_{ij} elements from the mean. The mean number of trips per interchange is:

$$\frac{100{,}000}{2{,}500} = 40$$

Consider the characteristics of the data. A large proportion of the t_{ij} elements is significantly less than the mean. This implies that the denominator of the statistic will in turn be large.

The numerator of R^2 is the sum of squared deviations of the predicted values, $\hat{t}_{ij}$, from the observed values. For small interzonal volumes, the magnitude of the numerator will be small, regardless of the density function of the t_{ij}.

What in fact is happening in these correspondence tests is that because of the high proportion of low volume interchanges, which deviate significantly from the mean, the value being subtracted from unity in equation (5.5) is extremely small. Therefore, significant differences in the distribution matrices are, in effect, "buried" in insignificant differences in the R^2 values.

The characteristics of the data, therefore, render interpretation of these values extremely difficult, if not impossible.

Although, there appears to be a better correspondence between D_N and the mean distribution than between D_P and the mean distribution, the significance of differences in correspondence cannot be accurately determined.

Table 2 displays the distribution of trips from a chosen zone, which is representative of the entire system, according to the three population-sized matrices. The table emphasizes the fact that while the Poisson distribution differs from both the mean and normal distributions, this difference is only reflected in the third decimal place of the R^2 statistic.

On the other hand, when sample-sized distributions are being compared, the denominator of the R^2 term is much smaller than when population-sized distributions are being compared. Although the total number of variables remains the same, the term $\sum_i \sum_j t_{ij}$ is only four per cent of the total number of trips distributed in the system. The mean number of trips per interchange, then, is only 1.6. The deviations of t_{ij} from the mean are thus much less and the right hand term in equation (5.5) becomes proportionately larger. The same differences in correspondence between two matrices will produce different R^2 values for different scales of investigation. We expect, therefore, the greater differences in R^2 values between the Poisson and normal samples to result partially from the sensitivity of the R^2 statistic to changes in scale. But the information loss resulting from taking a small sample cannot be precisely determined.

These preliminary tests emphasize the difficulties of using R^2 as a measure of correspondence. Although larger values of the statistic indicate better correspondence with the data, it is difficult to determine how much better this correspondence is. This problem can be partially overcome by graphically assessing as well as analytically assessing the results of the test, to assist in the interpretation of

TABLE 2

THE MEAN, POISSON AND NORMAL TRAVEL PATTERN

GENERATED FROM ZONE TWENTY-FIVE

DESTINATION ZONE	t_{ij} (MEAN)	t_{ij} (POISSON)	t_{ij} (NORMAL)	DESTINATION ZONE	t_{ij} (MEAN)	t_{ij} (POISSON)	t_{ij} (NORMAL)
1	0	0	0	26	1	1	1
2	3	3	3	27	13	11	14
3	1	0	1	28	17	17	18
4	3	3	3	29	191	165	194
5	35	40	36	30	37	44	36
6	39	48	40	31	13	10	13
7	78	73	77	32	1	1	1
8	46	45	45	33	52	45	50
9	18	20	18	34	15	8	15
10	1	0	1	35	6	7	6
11	1	3	1	36	1	2	1
12	0	0	0	37	12	13	13
13	0	1	0	38	11	12	11
14	0	1	0	39	10	7	10
15	13	8	12	40	2	2	2
16	32	31	31	41	17	13	17
17	26	32	26	42	64	60	62
18	284	288	282	43	62	69	62
19	160	168	163	44	20	19	20
20	32	28	34	45	8	11	8
21	1	2	2	46	5	5	5
22	2	3	2	47	23	24	23
23	1	3	1	48	13	10	13
24	0	1	0	49	3	6	3
25	1	1	1	50	1	2	1

the R^2 values.

To determine the reliability of a single random sample in representing the mean distribution of travel in the urban area, the Poisson and normal samples can be scaled up to the population size and compared to the actual mean distribution. R^2 values, showing the correspondence of the two random samples, drawn from D_P and D_N, scaled up to the population size to the mean distribution, are .8874 and .8969 respectively. There is a significant decrease in correspondence to the mean trip matrix. R^2 values of D_P and D_N to the mean were .9958 and .9998 respectively. After sampling and scaling, the correspondence has been reduced by approximately ten per cent.

Despite the reduction in correspondence, Batty (1970c) has suggested that R^2 values of 0.9 still indicate suitable "fits" to the mean distribution. However, it is useful to visualize what this R^2 value means in the spatial interaction context. Figure 6 shows how travel is distributed from one of the zones, zone twenty-five, to all destinations, as predicted by the scaled-up values of the two random samples, compared to the mean distribution of trips. (Note: The lines connecting the number of trips to each zone have no interpretive value. Their function is simply one of illustration, in this and in succeeding diagrams.)

The figure shows that the random samples are sensitive to major traffic flows out of the zone, but tend not to account for low volume interchanges. This is due to the "coarseness" of the sampling process, which is related to sample size. The probability of observing travel on a low-volume interchange during a home-interview survey is much

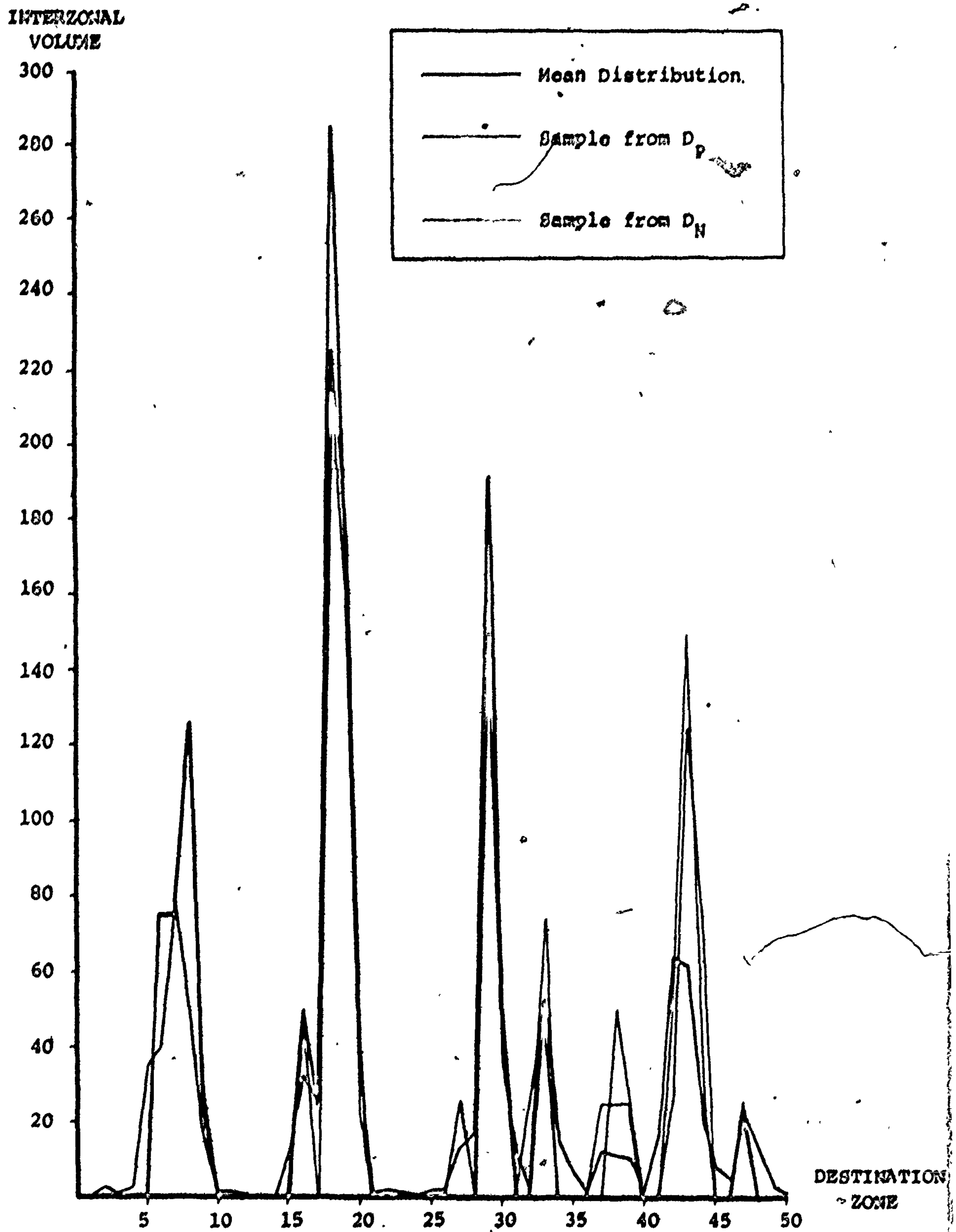

FIGURE 6: Scaled-Up Sample Predictions and the Mean Distribution of Trips Generated from Zone Twenty-Five

smaller than the probabilities of observation on interchanges of high volume. Clearly, as the sample size increases, it is more likely that some of these trips will be observed, if the survey is truly random (Chaterjee, *et al.*, 1974). However, a four per cent sample, consisting of only fifty-six "interviews" or sample points in this origin zone, is not large enough to observe this residual travel.

The other point to note is the consistent under-estimation by the samples of high-volume interchanges, and the over-estimation of medium volume interchanges (Figure 7). Although part of this inaccuracy may be due to the scaling-up of the samples (by a factor of twenty-five), there appear to be other unidentified factors which affect the sample predictions. The effect that this phenomenon has upon model predictions will be discussed later in this chapter.

The analysis has enabled us to draw several conclusions concerning the relationship of the samples to the mean trip distribution.

1. There exists a reasonable degree of correspondence between the two sample trip matrices and the mean trip matrix, although significantly reduced from the correspondence of the actual trip matrices to the mean trip matrix.

2. A small random sample tends to be a coarse representation of the actual distribution. Low volume interchanges are generally not observed in such a sample. However, the probability of observing these interchanges is a function of sample size, i.e., they are more likely to be observed in larger samples.

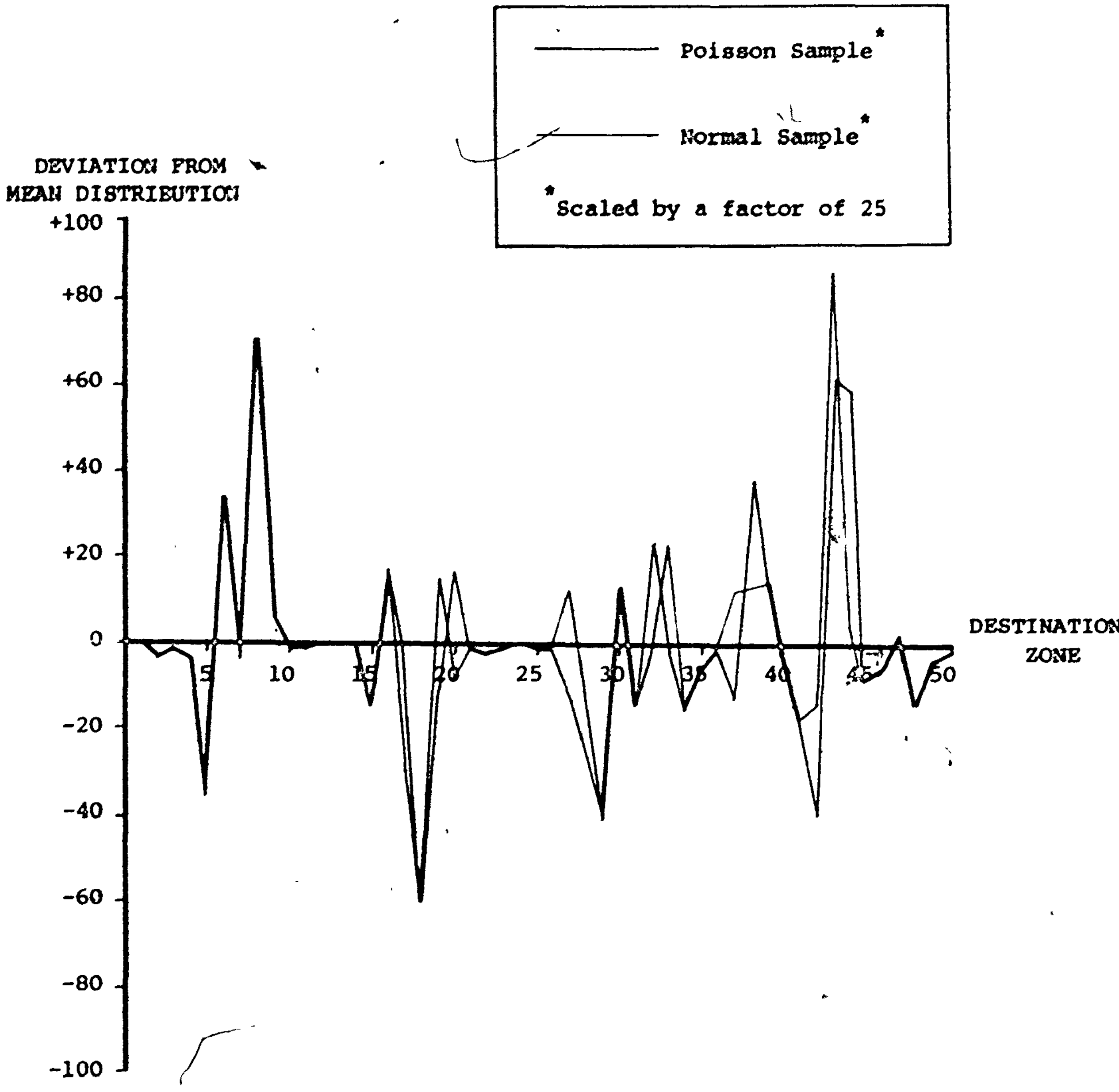

FIGURE 7: The Deviation of Scaled-Up Poisson and Normal Samples from the Mean Trip Distribution Generated from Zone Twenty-Five

3. High volume interchanges appear to be under-estimated and medium volume interchanges appear to be over-estimated in the sample. The effect of this will be examined in the following section.

A COMPARISON OF MODEL PREDICTIONS GENERATED BY THE MAXIMUM LIKELIHOOD AND LEAST-SQUARES ESTIMATORS

The analysis in this section involves measuring the correspondence between the predictions generated by parameters derived by two competing statistical estimators, under two different assumptions about travel over the interchanges in the system.

The correspondence, measured by R^2, between the model predictions and the sample data is given in Table 3.

TABLE 3

R^2 BETWEEN MODELS ESTIMATED BY MAXIMUM LIKELIHOOD OR LEAST-SQUARES, AND THE SAMPLE DATA USED TO CALIBRATE THE MODEL

	Model Estimated By M. L. Statistics R^2	Model Estimated By L. S. Statistics R^2
Sample Drawn from D_P	.8815	.8784
Sample Drawn from D_N	.8924	.8911

The measures of goodness-of-fit, in both cases, are made with respect to the sample distribution against which the model is calibrated. This is exactly the same procedure as practiced in model application.

The values in Table 3 indicate the differences in correspondence to the sample data to be only marginal. When the probability density function of t_{ij} is Poisson, the parameters generated by maximum likelihood statistics do not give the model a better fit to the sample data than do the parameters generated by least-squares, even though the maximum likelihood statistics assume the t_{ij} to occur exactly as postulated. Furthermore, when the density function of t_{ij} is normal, with common, constant variance, the parameters generated by maximum likelihood statistics, which assume t_{ij} to be Poisson, do not give the model a significantly poorer fit to the sample data, than the parameters derived from least-squares statistics, even though the statistical conditions, assumed in the maximum likelihood statistics are incorrect.

For one zone, the similarity in correspondence of the two distributions is qualitatively assessed in Figures 8 and 9. Comparison of the figures shows the predictions generated by maximum likelihood and least-squares to be almost identical, in fact, exactly identical when the density function of t_{25-j} is normal. One distribution can certainly not be preferred to the other, given this information.

Another feature shown in the figures is the over-estimation of high volume interchanges, and the under-estimation of medium volume interchanges by the models calibrated by the methods of maximum likelihood and least-squares. This agrees with the findings of Batty, and appears to be a characteristic feature of the gravity model. It tends to offset the characteristics of the sample trip matrix noted earlier, i.e., that

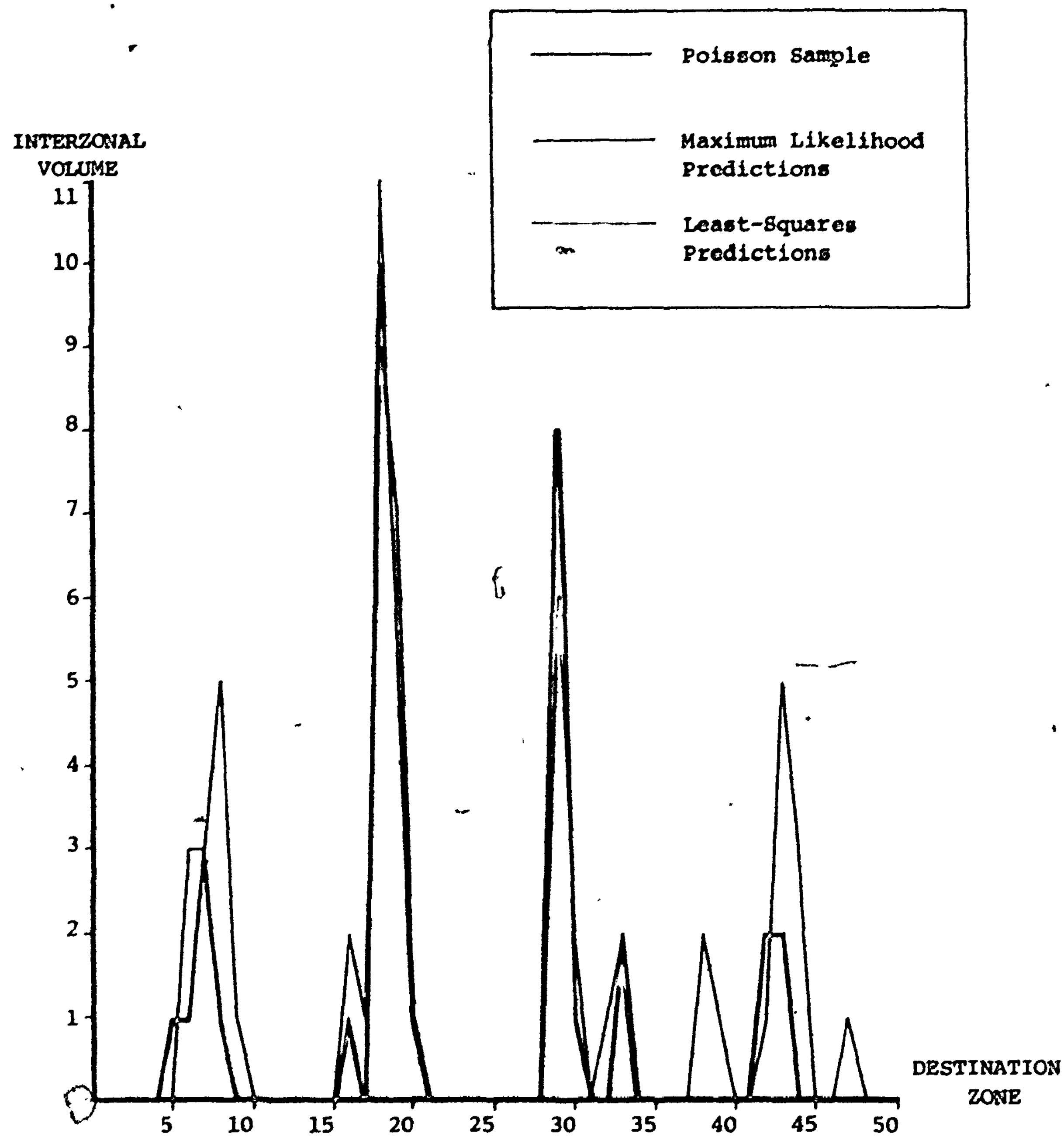

FIGURE 8: Comparison of Maximum Likelihood and Least-Squares Predictions with Poisson Sample for Trips Generated from Zone Twenty-Five

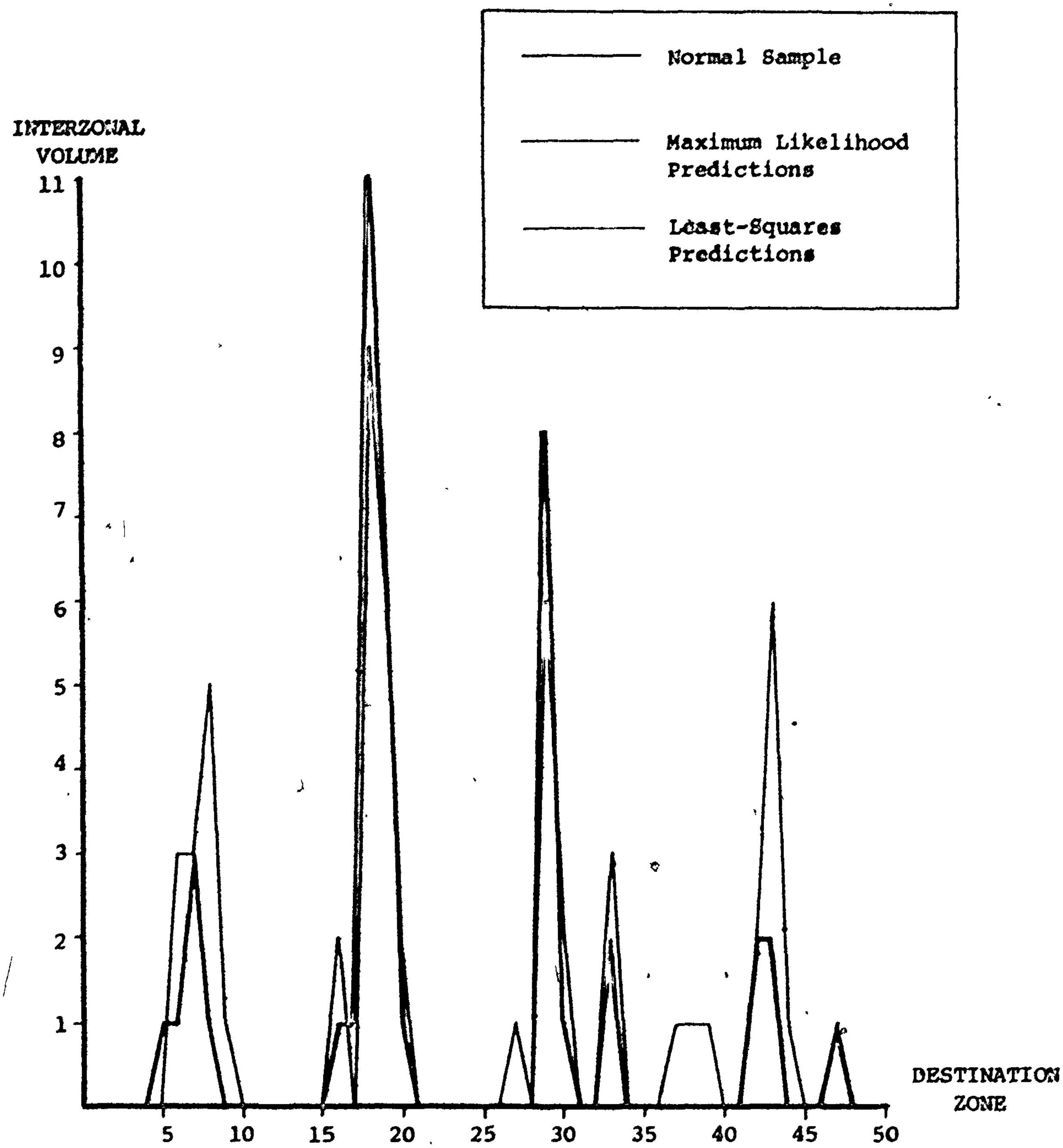

FIGURE 9: Comparison of Maximum Likelihood and Least-Squares Predictions with Normal Sample for Trips Generated from Zone Twenty-Five

high volume interchanges are under-estimated and low volume interchanges are over-estimated.

The effect that sample inaccuracies and the compensation tendencies of the gravity model have upon the model's correspondence to the mean trip distribution can be seen in Figure 10. In this figure, the model output has been scaled up by a factor of twenty-five for comparative purposes.

Figure 10 compares the mean volume of traffic generated from zone twenty-five to all destination zones, to the scaled-up traffic volume as predicted by the model calibrated by maximum likelihood against sample data in which each t_{ij} is normally distributed. Since the model predictions are essentially identical for both calibration methods, under both assumptions about t_{ij}, a single set of model predictions suffices.

Generally, the fit of the model to the mean trip distribution, from this origin-zone, is quite good. The over-compensation effect of the gravity model tends to negate the characteristics of the trip survey noted earlier, and model predictions reasonably approximate the mean travel volume originating from zone twenty-five. Those interchanges carrying only residual traffic are not accounted for in the model. This is due to the "coarseness" of the random sample, as outlined earlier.

The results of the analysis, therefore, contradict the second hypothesis defined in the chapter. The definition of trip purpose by characteristic probability density functions in the calibration statistics does not necessarily give the model a better fit to the trip survey or

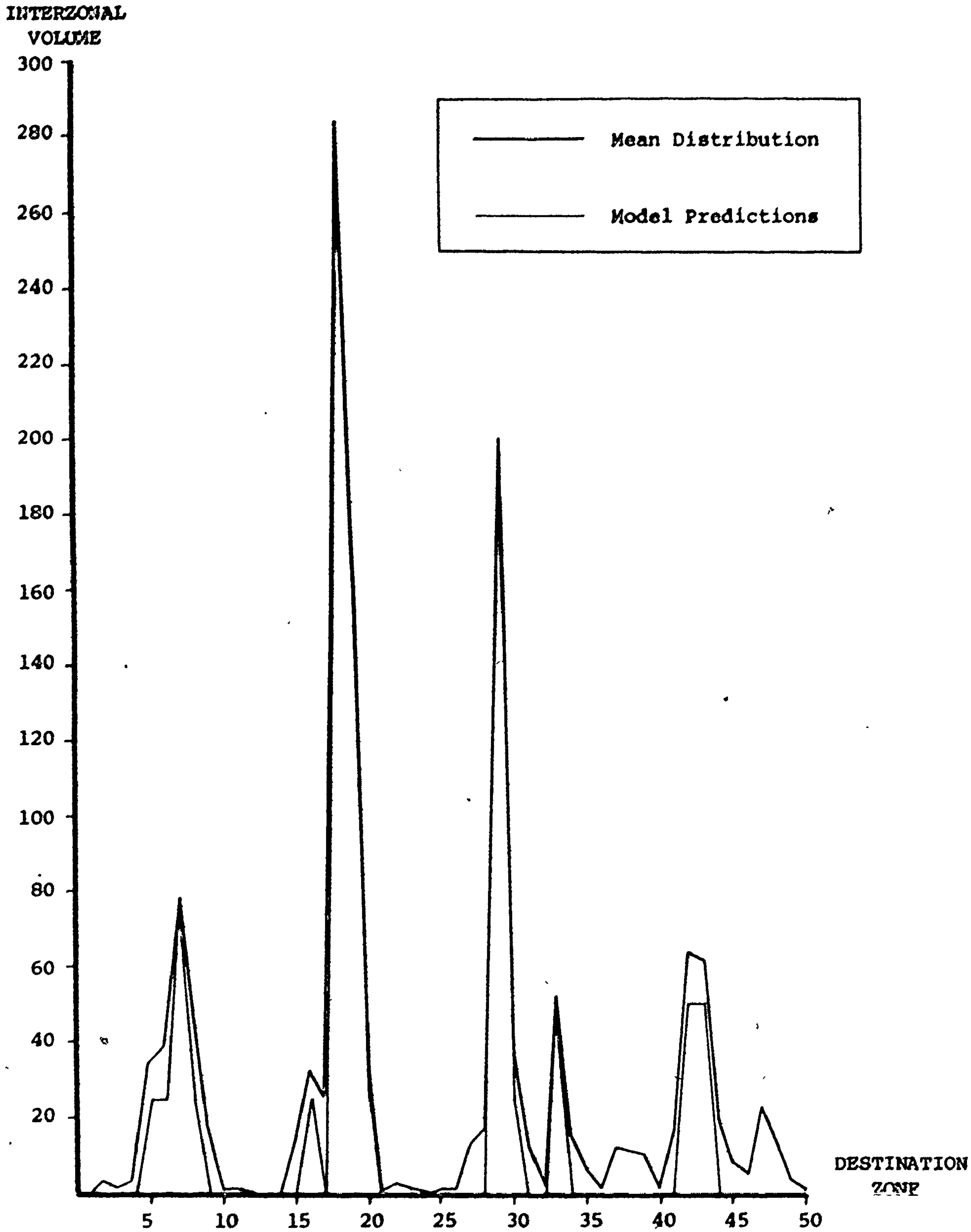

FIGURE 10: Scaled-Up Model Predictions as Calibrated by Maximum Likelihood and the Mean Trip Distribution Generated from Zone Twenty-Five

to the mean distribution of trips in the system. Although Kirby (1974) has shown that there are certain theoretical requirements which must be satisfied to derive best parameter values, in practice, these requirements do not appear to be necessary.

The remaining issue to be resolved is why these requirements do not have to be upheld in modelling practice.

FACTORS IN MODEL APPLICATION WHICH REDUCE THEORETICAL CALIBRATION REQUIREMENTS

Two factors can be identified which contribute to the contradiction of the second hypothesis. They are related to certain assumptions implied in the development of theoretical requirements for calibrating the model, which do not hold in practice. The first assumption concerns the effect that different density functions have upon the trip pattern in the system. The second assumption involves the sensitivity of the model itself.

If a specific probability density function is to be specified in the calibration statistics, we are assuming that the trip pattern generated by that density function is significantly different from the trip patterns generated by any other functions. Intuitively, this assumption appears reasonable. By assuming different density functions, we are trying to represent the characteristic variances in the day to day travel, specific to different trip purposes, over the set of interchanges in the system. We are expecting these variances to generate different aggregate travel patterns.

The analysis has taken two probability density functions, ϕ_1 and

ϕ_2, to be Poisson and normal respectively, with variance

$$\sigma^2(\phi_1) = t_{ij} \tag{5.6}$$

$$\sigma^2(\phi_2) = (0.2(t'_{ij}))^2 = 4 \tag{5.7}$$

where t'_{ij} is the median number of trip interchanges in the system (t'_{ij} = 11 trips). It has then generated a travel pattern for each of these assumptions about the variance of t_{ij}, and has taken a *single* random sample from each distribution. Instead of finding the travel patterns to be significantly different, the analysis has found the correspondence of each to be very similar.

The similarity in correspondence may be due to the fact that the majority of interchanges in the system carry small traffic volumes. Over these interchanges, the variance in volumes for both ϕ_1 and ϕ_2 will be essentially the same.

Other factors, such as the effect of sampling on the shape of the density function (Kirby, 1974, p. 99), may also influence the correspondence of the two trip patterns. However, this has not been researched in this study.

The characteristics of the travel pattern, therefore, may be such, that the differences between different "behavioral patterns" (identified by different probability density functions) may not be distinct. The effect of different statistical assumptions on the data in calibration statistics will thus be minimal.

The second, and perhaps more important factor concerns the sensitivity of the spatial interaction model to changes in the value of the parameter, β. Although little mention of this aspect is made in the literature, there is some evidence (Batty, 1970c, p. 111; Batty, 1971, p. 426; Batty and Mackie, 1972, p. 215) to suggest that the fit of the model to the sample remains relatively invariant, regardless of the correlation statistics used, over wide ranges of parameter values.

Recalling Table 3, the correspondence of the distributions generated by the model to the samples are very similar. Table 4 displays the parameter values generating the distributions which produce these measures of correspondence.

TABLE 4

MEASURES OF CORRESPONDENCE AND PARAMETER VALUES OF MODELS ESTIMATED BY MAXIMUM LIKELIHOOD AND LEAST-SQUARES

	Model Estimated By M. L. Statistics		Model Estimated By L. S. Statistics	
	β	R^2	β	R^2
Sample Drawn from D_P	0.0219	.8815	0.0799	.8784
Sample Drawn from D_N	-0.0138	.8924	0.0118	.8911

The table indicates that the best parameter values (meaning the parameter values closest to the actual value) are obtained when the assumptions implied in the calibration statistics are satisfied by the data. But it also indicates that similar measures of correspondence can be achieved from significantly different values of the parameter. This lack of sensitivity may be an indication of the robustness of the model, i.e., the capability of the balancing factors A_i and B_j to adjust the t_{ij} elements.

If the model is insensitive to changes in parameter values, then the specification of probability density functions in calibration statistics is no longer at issue if correlation statistics are used to measure the model's goodness-of-fit. This is because the model will most likely produce acceptable results, regardless of the calibration statistic defined.

It can be seen that if the two assumptions implied when developing the theoretical requirements for calibrating spatial interaction models -- concerning the different trip patterns generated by different probability density functions, and the sensitivity of the model itself -- are not satisfied in practical model calibration, the theoretical requirements become no longer necessary. The problem in model calibration becomes one of estimating parameter values as quickly and efficiently as possible.

SUMMARY

This chapter has presented the results of the analysis proposed in Chapter 4. It first formally outlined the two hypotheses to be tested, and defined the statistic to be used to measure the goodness-of-fit. In doing so, it specified the restricted conditions under which the hypotheses can be tested, to prevent the misinterpretation of results.

After describing the hypothetical area over which travel was distributed and tests made, the chapter outlined the operations in the analysis which produce the distributions to be examined.

The results of the analysis, which are limited in their content by the restricted test framework, as outlined in the introduction to the chapter, show that the random sample appears to suitably correspond to the mean travel distribution. This implies that reliable results can be generated by the spatial interaction model calibrated to this data.

The chapter also shows that statistical assumptions in the maximum likelihood calibration statistics, which are necessary to satisfy theoretical requirements for calibrating the model, do not significantly affect model performance. The chapter concludes by comparing the model output in relation to the sample, and then identifies factors which appear to eliminate the theoretical requirements for model calibration.

The final, and following chapter will summarize the findings of the paper and will suggest areas for further research.

CHAPTER 6

SUMMARY AND CONCLUSIONS

This paper has attempted to analyze two issues in the calibration of spatial interaction models. The first issue concerns the theoretical requirements for calibrating spatial interaction models as proposed by Kirby (1974). The second involves the reliability of the random sample in representing the mean travel distribution in the area to be modelled. The paper has been developed through four sections.

Chapter 1 has defined the problem of model calibration and has described the characteristics of the spatial interaction model which make it difficult to calibrate. It has then assessed the different approaches to model calibration which have evolved since the development of the Lowry model, and has stressed the shortcomings in each method. The chapter has argued that statistical estimation techniques possess properties which make these methods preferable to other calibration approaches.

Chapters 2 and 3 have examined the two principal methods of statistical estimation. These are the methods of maximum likelihood and least-squares. Chapter 2 has further argued that calibration is a problem of point estimation, and not hypothesis evaluation. It has therefore rejected Hyman's (1969) approach as a method of parameter estimation.

The paper has examined the two conflicting mathematical interpretations of the maximum likelihood estimator in calibrating the spatial interaction model. It has been shown that under certain conditions, the implicit assumptions in each can be reconciled, even though the calibration problem is approached from ~~two different~~ perspectives. Furthermore, the conditions under which these two approaches are complementary are likely to be observed when calibrating urban spatial interaction. Chapter 2 has gone on to define the statistical conditions which are necessary to satisfy the theoretical requirements of the maximum likelihood calibration statistics. It has stressed that the parameter estimates derived by the maximum likelihood estimator are unbiased only if the trip data correspond to these assumptions.

The chapter has then reviewed the work of Kirby (1974), who has attempted to apply behavioural hypotheses to the maximum likelihood calibration statistics. It has stated that the different calibration statistics which can be derived from the maximum likelihood estimator may represent different trip purposes which occur in the urban system. These trip purposes are characterized by different probability density functions over the inter-zonal interchanges and must be explicitly input into the maximum likelihood estimator in order to derive appropriate calibration statistics.

Chapter 3 has examined the least-squares estimator as a method of model calibration, as proposed by Cesario (1975). Through an examination of its properties, it has been found that unlike the method of maximum likelihood, the least-squares estimator makes no implicit assumptions about the distribution of trips over the zonal interchanges, and thus imposes no behavioral assumptions on the calibration technique. The paper has compared the two methods of statistical estimation by imposing conditions on the maximum likelihood estimator so that it yields the same calibration statistics as the least-squares estimator. It has shown that while the principle of least-squares makes no implicit assumptions about the sample data, the method of maximum likelihood must make very restrictive assumptions, in order to derive identical calibration statistics.

Thus, it has been shown that there appears to be a basic contradiction in the assertion of behavioral notions embedded in calibration statistics. Theoretically, the least-squares estimator can derive unbiased parameter estimates without making any assumptions about the probability density functions of the trip interchanges. Conversely, the maximum likelihood estimator can only yield unbiased parameter estimates by making implicit, and sometimes unrealistic assumptions about the nature of the traffic flow over the zonal interchanges. It follows that if indeed the trip distributions derived from the competing methods are similar, the behavioral properties of the maximum likelihood estimator are non-existent.

The fourth chapter has examined related problems in model calibration. Specifically, it has discussed the reliability of the sample observation, the trip survey, in representing the mean distribution of trips in the urban system. This is critical because it is the mean distribution which the spatial interaction model is assumed to generate. If travel over the set of interchanges varies from day to day, the random sample will retain these deviations from the mean. These biases will then be generated through the model. The paper has stressed that the magnitude of this bias must be examined in order to assess the quality of the model's output.

Second, the chapter has examined how the key variable in calibration statistics has been defined in the literature. This is the generalized cost variable, which has been defined as distance, time and various combinations of both. While it has concluded that the problem of variable definition can only be resolved through empirical examination, the chapter has proposed a hypothetical analytical framework, which controls the bias which can be introduced through variable misspecification. The paper has proposed to apply this framework to both the problem of determining the reliability of the sample observation, and to the problem of determining the existence of behavioral notions in maximum likelihood calibration statistics to see whether theoretical calibration requirements are necessary to derive unbiased model results.

The final section has presented and discussed the results of the analysis proposed in Chapter 4. After briefly discussing the concepts of best parameter estimates and optimum goodness-of-fit in order to interpret the outcomes of the analyses, the paper has formulated several conclusions.

The paper has found that in the constructed hypothetical framework, the trip survey, drawn at random, retains the essential characteristics of the mean trip distribution. It has found that the sampling process inherently loses some information about the distribution, especially concerning low volume interchanges, and has postulated that the information loss is directly attributable to sample size.

The analysis has also found that the satisfaction of the theoretical requirements for calibrating spatial interaction models does not have any appreciable effect on the goodness-of-fit of the generated distribution to the data. This serves to contradict the behavioural hypotheses about calibration statistics, as asserted by Kirby.

A further result has been observed in the analysis which suggests further study. An inherent property of the distribution generated by the gravity model was observed in the analysis. This is the tendency for the model to over-predict high volume interchanges and under-predict medium and low volume interchanges. This characteristic perfectly counter-acted an undesirable property of the random sample, that of under-estimating high-volume interchanges and over-estimating medium and low-volume interchanges. The result was a remarkably accurate macro-distribution of travel in the area.

A verification and explanation of these observations is clearly needed to reaffirm the usefulness of the spatial interaction model in the planning context.

The findings of this study can only be deemed tentative and this fact can only be appreciated through an evaluation of the analysis.

The problem of calibration has been approached by performing the same basic operations as would the analyst when applying the spatial interaction model to a real situation. A *single* random sample has been drawn from the population. Inferences about the relationship of the sample to the population, and about the fit of the generated distribution to the data have been based upon this lone sample only.

The principal difference, however, between the hypothetical framework and an empirical study has been that we have possessed additional information about the system and its variables that the analyst could not have collected. For example, we know exactly the function and variables which distributed the trips. We know how trips were distributed over the zonal interchanges, and the mean trip distribution. The empirical analogues are either unobservable or do not even exist in reality. We have used this information to test the modeller's assumptions (about the sample) and the theoretician's propositions (about statistical requirements) with regards to the issue of calibration.

Clearly, the design of the analysis could have been extended so that more conclusive results were obtained. Instead of a single random sample, several observations on the system could have been made to find an *average* correspondence to the mean distribution. Similarly, the spatial interaction model could have been calibrated to each observation to yield a range of parameter values. These values could have then been used to generate several distributions to find an average measure of goodness-of-fit to the data. However, further research must initially be directed to the sensitivity issue of the spatial interaction model, and at the statistical measure used in the analysis in order that the findings of this research endeavor be strengthened. These two points warrant additional comment.

Consider first the statistics used to measure goodness-of-fit in the spatial interaction context. The weaknesses of correlation statistics were brought out in Chapter 1. In spite of their weaknesses they are still recommended by modellers in order to measure model fit. But it appears that little can be inferred, especially from the R^2 statistic. Although its use does not require any information loss through aggregation or omission of variables, it has more peculiar properties than originally suspected, as seen in Chapter 5, which make it an especially poor statistic to use in spatial interaction modelling.

Clearly, what is needed is a statistic whose properties make it especially adaptable to the spatial interaction context. Its distribution and assumptions should be known, it should be sensitive, remain invariant under transformation and should be defined over all values of t_{ij}.

Secondly, the whole question of model sensitivity should be extensively researched. It seems to be clear that if model predictions remain invariant or maintain their goodness-of-fit throughout a range of parameter values, calibration methodologies become less of an issue. There is no point in interpreting statistical conditions in any context, including trip purpose, if a unique "best" optimum does not exist.

Perhaps the best approach to this would be to plot the objective functions of the statistical estimators over the range of parameter values in question. This would enable us to determine not only the general sensitivity of the statistic about the optimum but also the reliability of the statistic, by observing its behavior over the parameter range.

If in fact, as is suspected from this analysis, the statistics are generally insensitive to changes in parameter values around the optimum, research in the calibration field should be directed towards *efficiency* and *reliability* criteria rather than towards modifying statistics to incorporate non-existent or unobservable behavioral phenomena.

Answers to the questions posed throughout this paper are necessary to develop a sound theoretical and practical base for calibrating spatial interaction models. Although many points remain unresolved, this paper can be regarded as another step in addressing these issues.

BIBLIOGRAPHY

Alonso, L. (1972) *A Recreational Trip Making Model of the Spanish Penninsula.* Department of Geography, University of Reading: M.Sc. Dissertation.

Batty, M. (1970a) "Models and Projections of the Space Economy", *Town Planning Review*, 41, 121-147.

Batty, M. (1970b) "An Activity Allocation Model for the Nottinghamshire-Derbyshire Subregion", *Regional Studies*, 4, 307-332.

Batty, M. (1970c) "Some Problems of Calibrating the Lowry Model", *Environment and Planning*, 2, 95-114.

Batty, M. (1971) "Exploratory Calibration of a Retail Location Model Using Search by Golden Section", *Environment and Planning*, 3, 411-432.

Batty, M. (1972) "Recent Developments in Land Use Modelling: A Review of British Research", *Urban Studies*, 9, 151-177.

Batty, M. and S. Mackie (1972) "The Calibration of Gravity, Entropy and Related Models of Spatial Interaction", *Environment and Planning*, 4, 205-233.

Batty, M. and S. Mackie (1973) "Auto Calibration of Trip Distribution Models", *Traffic Engineering and Control*, 12, 2.

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

Batty, M., R. Bourke, D. Cormode, M. Anderson-Nicholls (1974) "Experiments in Urban Modelling for County Structure Planning: The Area 8 Pilot Model", *Environment and Planning A*, 6, 455-478.

Ben-Akiva, M. (1973) *Structure of Passenger Travel Demand Models.* Department of Civil Engineering, Massachusetts Institute of Technology: Ph.D. Dissertation.

Cesario, F.J. (1975) "Least-Squares Estimation of Trip Distribution Parameters", *Transportation Research*, 9, 13-18.

Charnes, A., W. Raike and C.O. Bettinger (1972) "An Extremal and Information-Theoretic Characterization of Some Inter-zonal Transfer Models", *Socio-Economic Planning Sciences*, 6, 531-537.

Chatterjee, A., G.G. Makowski and K.C. Sinha (1974) *Reliability Analysis of Origin-Destination Surveys and Determination of Optimal Sample Size.* National Technical Information Service. Springfield, Virginia.

Cripps, E.H. and D.H.S. Foot (1969a) "The Empirical Development of an Elementary Residential Location Model for Use in Sub-regional Planning", *Environment and Planning*, 1, 81-90.

Cripps, E.H. and D.H.S. Foot (1969b) "A Land Use Model for Sub-regional Planning", *Regional Studies*, 3, 243-268.

Draper, N.R. and H. Smith (1966) *Applied Regression Analysis.* New York: John Wiley.

Edwards, A.W.F. (1972) *Likelihood.* Cambridge: Cambridge University Press.

Evans, A.W. (1971) "The Calibration of Trip Distribution Models with Exponential or Similar Cost Functions", *Transportation Research*, 5, 15-38.

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

Freeman, H. (1963) *Introduction to Statistical Inference.* Reading: Addison-Wesley.

Freund, J.E. (1952) *Modern Elementary Statistics.* Englewood Cliffs: Prentice-Hall.

Hill, D.M., D. Brand and W.B. Hansen (1965) "Prototype Development of Statistical Land Use Model for Greater Boston Region", *Highway Research Record*, 114, 51.

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

Hyman, G.M. (1969) "The Calibration of Trip Distribution Models", *Environment and Planning*, 1, 105-112.

Isard, W. (1960) *Methods of Regional Analysis.* Cambridge: M.I.T. Press.

Kirby, H.R. (1974) "Theoretical Requirements for Calibrating Gravity Models", *Transportation Research*, 8, 97-104.

Larson, H.J. (1969) *Introduction to Probability Theory and Statistical Inference.* New York: John Wiley.

Lindley, D.V. (1965) *Introduction to Probability and Statistics from a Bayesian Viewpoint.* London: Cambridge University Press.

Lowry, I.S. (1963) "Location Parameters in the Pittsburgh Model", *Papers: The Regional Science Association*, 11, 145-165.

Lowry, I.S. (1965) "A Short Course in Model Design", *Journal of the American Institute of Planners*, 31, 158-165.

Mackie, S. (1972) *The Calibration of Spatial Interaction Models.* Department of Geography, University of Reading: Geographical Paper #14, January.

Masser, I. (1970) *Notes on an Application of the Lowry Model to Merseyside.* Department of Civic Design. University of Liverpool.

Morphet, R. (1975) "A Note on the Calculation and Calibration of Doubly Constrained Trip Distribution Models", *Transportation*, 4, 43-53.

Olsson, G. (1965) "Distance and Human Interaction: A Migration Study", *Geografiska Annaler*, 74B, 3-43.

Openshaw, S. (1973) "Insoluble Problems in Shopping Model Calibration When the Trip Pattern Is Not Known", *Regional Studies*, 7, 367-371.

O'Sullivan, P. (1971) "Forecasting Interregional Freight Flows in Great Britain", in M. Chisholm, A.E. Frey and P. Haggett (Eds.) *Regional Forecasting.* London: Butterworths, 443-450.

Reif, B. (1973) *Models in Urban and Regional Planning.* London: In Text Publishers.

Shannon, C. and W. Weaver (1949) *The Mathematical Theory of Communication.* Urbana: University of Illinois Press.

Sheppard, E.S. (1974) *Entropy in Geography: An Information Theoretic Approach to Bayesian Inference and Spatial Analysis.* Department of Geography, University of Toronto, Unpublished.

Siedmann, D.R. (1969) *The Construction of an Urban Growth Model.* Delaware Valley Regional Planning Commission: Report #1, Technical Supplement, Volume A, Philadelphia.

Tanner, J.C. (1961) *Factors Affecting the Amount of Traffic.* Road Research Laboratory: Technical Paper #51, London.

Tribus, M. (1969) *Rational Descriptions, Decisions and Designs.* New York: Pergammon.

Turner, G.C. (1970) *Severnside Shopping Model: A Discussion of the Calibration Procedure and Results.* Nathaniel Lichfield and Associates: Working Paper #9, London.

Turner, G.C. and R.J. Williams (1970) *A Note on the Activity Allocation Model for the Bristol Subregion.* Department of Town Planning, University College, London: Discussion Paper #9.

Webber, M.J. (1975) "The 'Meaning' of Entropy Maximizing Models", Paper presented at the Symposium on Mathematical Land Use Theory, McMaster University, Hamilton, April.

Wetherill, G.B. (1967) *Elementary Statistical Methods.* London: Methuen.

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

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

Wilson, A.G., A.F. Hawkins, G.J. Hill and D.J. Wagon (1969) "Calibration and Testing of the SELNEC Transport Model", *Regional Studies*, 3, 337-350.

Wonnacott, R.J. and T.H. Wonnacott (1970) *Econometrics.* New York: John Wiley.